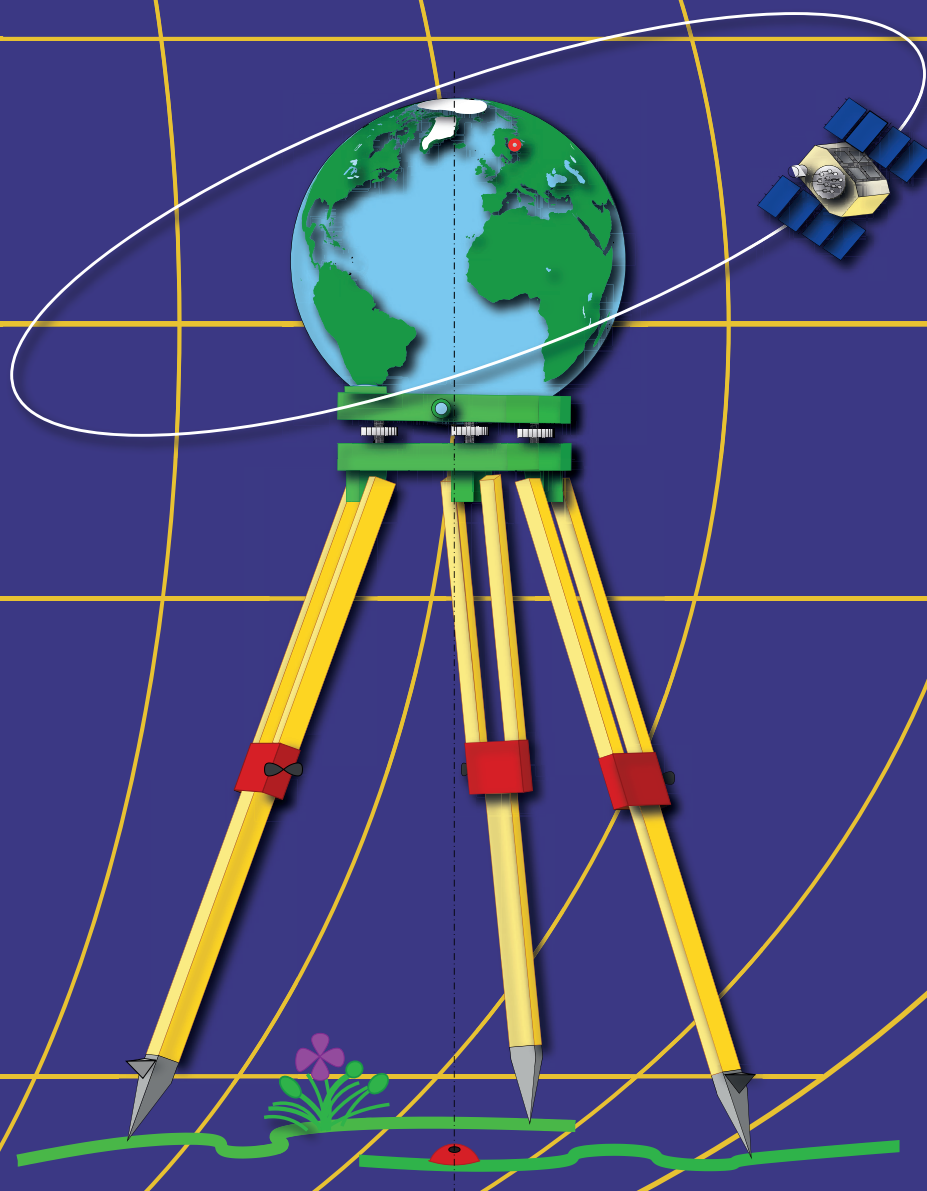


Martin Vermeer

GEODESY

The science underneath



Geodesy

The science underneath

Martin Vermeer

Aalto University publication series
SCIENCE + TECHNOLOGY 6/2019

© 2019 Martin Vermeer

ISBN (pdf) 978-952-60-8872-3
ISSN 1799-490X (pdf)
<http://urn.fi/URN:ISBN:978-952-60-8872-3>

Graphic design: Cover: Tarja Paalanen

Helsinki 2019

Finland

Author

Martin Vermeer

Name of the publication

Geodesy: The Science Underneath

Publisher School of Engineering**Unit** Department of Built Environment**Series** Aalto University publication series SCIENCE + TECHNOLOGY 6/2019**Field of research** Geodesy**Language** English**Abstract**

Geodesy is the science of precisely measuring and mapping the Earth's surface and locations of objects on it, the figure of the Earth and her gravity field, and changes in all these over time. Geodesy is an old science, going back to the days when land was taken into agricultural use and needed to be mapped. It is also a modern science, serving vital infrastructure needs of our developing global technological society.

This text aims to describe the foundations of both traditional geodesy, mapping the Earth within the constraints of the human living space, and modern geodesy, exploiting space technology for mapping and monitoring our planet as a whole, in a unified three-dimensional fashion. The approach is throughout at conveying an understanding of the concepts, of both the science and mathematics of measuring and mapping the Earth and the technologies used for doing so. The history of the science is not neglected, and the perspective of the presentation is unapologetically Finnish.

Keywords geodesy, land surveying, mapping, geodetic measurements, networks, co-ordinates, height, least-squares estimation, the built environment, figure of the Earth, GPS, space geodesy, gravity, geodynamics, geophysics

ISBN (pdf) 978-952-60-8872-3

ISSN (PDF) 1799-490X

Location of publisher Helsinki

Location of printing Helsinki

Year 2019

Pages 608

urn <http://urn.fi/URN:ISBN:978-952-60-8872-3>

Preface

ALTHOUGH FINLAND IS, and has been since independence, a superpower in the field of geodesy, there does not seem to be a modern geodesy textbook in the Finnish-language area. Finnish-language textbooks, as well as and popular books, do exist, but they are either badly outdated or treat only a certain sub-area of geodesy. Of these, we may mention the writings by Martti Tikka on measurement and instrument techniques and geodetic computation (Tikka, 1991, 1985), now largely obsolete, and Salmenperä (1998). The work on satellite positioning by Poutanen (1998) has now been updated (2017) and is very useful. The book by Kallio (1998) explains the least-squares statistical computation technique used in geodesy. All these sources have been helpful in writing this book.

Internationally, clearly more geodesy textbooks are on offer, and we have benefited from Torge (2001), Vaníček and Krakiwsky (1986), Kahmen and Faig (1988) in measurement and instrumental techniques, Heiskanen and Moritz (1967) in physical geodesy, and Hofmann-Wellenhof et al. (2001) in satellite geodesy.

The material in this book divides naturally into two parts: classical geodesy and modern geodesy. Each could be the textbook for its own course, which would each be worth three ECTS points.

The subjects discussed in the *classical geodesy* part (chapters 1–10) are the history of geodesy, the figure of the Earth and gravity, the reference ellipsoid, co-ordinates and heights, basics of geodetic measurements, units of measurement, uncertainty of measurement; Helmert transformations, the direct and inverse geodetic problems; levels and levelling, height systems, the geoid; theodolites and total stations, angle measurements; distance measurement using electromagnetic radiation and propagation of the measurement ray in the atmosphere; geodetic networks, measurement classes, network hierarchy; base and mapping measurements; area

and volume calculations.

In the part on *modern geodesy* (chapters 11–18), we concentrate instead on the development, during the past century, from geodesy closely tied to the Earth’s surface to genuinely three-dimensional geodesy, comprising space and satellite geodesy and truly three-dimensional positioning methods based on electromagnetism. We discuss three-dimensional reference systems, hyperbolic positioning systems and the global positioning system **GPS**; **GPS** satellites, orbits, signals, receivers; measurements of pseudo-range and carrier phase, measurement geometry, differencing of observations, integer-valued ambiguities and their fixing; processing **GPS** observations, relative and differential as well as real-time positioning. We also dive deeper into the statistical foundations of geodesy, including the least-squares method, residuals, statistical testing, outlier detection, reliability, and planning of measurement networks. Finally, we look at the borderlands between geodesy and geophysics, comprising the gravity field of the Earth and the gravimetric geoid; space geodesy, the rotational and orbital motions and deformations of the Earth; satellite orbits and the role of geodesy in geophysical research.

We have chosen in this text to concentrate on conceptual and fundamental matters. That also means describing the internal workings of instruments and processes which are in today’s systems handled automatically by smart software — even if to some, this may feel like we are teaching outdated skills.

Helsinki, 9th December, 2020,

Martin Vermeer

Second edition

A second, extensively corrected and improved edition was published on 25th September, 2022. No substantive content was added. The original text was [archived](#).



Acknowledgements

In drafting the various manuscript versions, lecturing transparencies and other materials from professors Teuvo Parm and Martti Martikainen were a great help. Assistant professor Jaakko Santala, teaching assistants Mauri Väisänen, Panu Salo and Henri Turto, post-doctoral researcher Octavian Andrei, professor Markku Poutanen, National Land Survey chief expert Marko Ollikainen, as well as students from many years of teaching are gratefully acknowledged for many useful remarks, corrections, and pieces of information.

Nicolàs de Hilster is gratefully acknowledged for contributing figure 9.8.

Several map images were drawn using the Generic Mapping Tools (Wessel et al., 2013).











































The English language was competently checked by the Finnish Translation Agency Aakkosto Oy. Tarja Paalanen designed the cover. Laura Mure and Matti Yrjölä helped with the practicalities of publishing.

This content is licenced under the *Creative Commons Attribution 4.0 International* (CC BY 4.0) licence, except as noted in the text or otherwise apparent.



Contents

Chapters

 	1. The history and societal status of geodesy	1
 	2. Geodetic measurements and co-ordinates	23
 	3. Map projections, datums and transformations	53
 	4. Height measurement and the levelling instrument	89
 	5. The theodolite	119
 	6. Angle measurement	157
 	7. Distance measurement	185
 	8. Base-network and detail-survey measurement	211
 	9. Construction surveying	235
 	10. Digital terrain models and volume calculation	251
 	11. The third dimension	265
 	12. The Global Positioning System (GPS)	283
 	13. Processing GPS observations	335
 	14. Adjustment calculus in geodesy	357
 	15. Statistical methods in geodesy	395
 	16. Gravity in geodesy	425
 	17. Space geodesy	457
 	18. Geodesy and geophysics	479
 	A. Properties of matrices	513
 	B. A short introduction to magnetohydrodynamics	521
 	C. The Kepler orbital elements for satellites	525

Preface **i**

List of Tables **xii**

List of Figures	xiii
Acronyms	xxiii
1. The history and societal status of geodesy	1
1.1 The figure of the Earth, early conceptions	1
1.2 Newton's laws and the figure of the Earth	5
1.3 The mathematical figure of the Earth or <i>geoid</i>	9
1.4 The geodesic	12
1.5 The flattening of the Earth and gravity	13
1.6 Reference surfaces and reference systems	14
1.7 The sub-fields of geodesy	14
1.8 Topographic surveying: from terrain to map	16
Self-test questions	20
2. Geodetic measurements and co-ordinates	23
2.1 Units of measurement	23
2.2 Measurement error and uncertainty	27
2.3 Stochastic quantities	30
2.4 Statistical distributions	32
2.5 Geodetic observables	42
2.6 About co-ordinates	46
2.7 Why we use plane co-ordinates	46
2.8 Co-ordinates of location in three dimensions	47
Self-test questions	50
Exercise 2–1: Co-ordinates and a street address	51
3. Map projections, datums and transformations	53
3.1 Map projections	53
3.2 The various co-ordinate solutions used in Finland	55
3.3 Map projections used in Finland	58
3.4 More about plane co-ordinates	65
3.5 The geodetic forward and inverse problems	68
3.6 The similarity co-ordinate transformation	71
3.7 Determining the transformation parameters	73
3.8 Datums and datum transformations	77
3.9 Map projections and height systems.	83
3.10 The time co-ordinate	84
Self-test questions	85
Exercise 3–1: Distances	86



4. Height measurement and the levelling instrument	89
4.1 Height, geopotential, and the geoid	89
4.2 Orthometric height	92
4.3 Height determination and levelling	93
4.4 The levelling instrument (“level”)	97
4.5 The measuring telescope	98
4.6 The tubular level	101
4.7 Checking and adjusting a levelling instrument	102
4.8 Self-levelling instrument	105
4.9 Digital levelling instrument	107
4.10 The levelling staff	108
4.11 Levelling methods	111
Self-test questions	116
Exercise 4–1: Heights	118
5. The theodolite	119
5.1 Horizontal angles and zenith angles	119
5.2 The axes of a theodolite	121
5.3 Construction of a theodolite	122
5.4 Theodolite handling in the field	124
5.5 Taking readings	136
5.6 Instrumental errors of a theodolite	142
5.7 Electronic theodolites	148
5.8 Case: Leica robotic tacheometer TCA2003	152
Self-test questions	155
6. Angle measurement	157
6.1 Horizontal angle measurement	157
6.2 Intersection and resection	158
6.3 Observation method of complete sets	161
6.4 Station adjustment of horizontal angles	162
6.5 Traverse measurement and computation	170
6.6 Open traverse	171
6.7 Closed traverse	174
6.8 Zenith angles and refraction	179
6.9 Heights of instrument and signal	183
Self-test questions	184



7. Distance measurement	185
7.1 Mechanical distance measurement	185
7.2 Electromagnetic radiation	188
7.3 Väisälä interferometry	191
7.4 Electronic distance measurement	194
7.5 Ray propagation in the atmosphere	201
7.6 “Curvature corrections”	204
7.7 Geometric reductions	206
Self-test questions	208
8. Base-network and detail-survey measurement	211
8.1 Objective and planning of base-network measurement . .	211
8.2 Guidance and standards	213
8.3 Network hierarchy and classification	214
8.4 The terrain, the ellipsoid and the map plane	218
8.5 Detail survey	222
8.6 Carrying out a detail survey	228
Self-test questions	234
9. Construction surveying	235
9.1 Zoning plans and setting out	235
9.2 Setting out and infrastructure	236
9.3 Straight lines, circular arcs, rounding of corners	239
9.4 Transition curve	243
9.5 Road and street surveying	244
9.6 Construction surveying	245
9.7 Other measurements	246
Self-test questions	249
10. Digital terrain models and volume calculation	251
10.1 Terrain-model measurement, construction, presentation .	253
10.2 Use of terrain models	255
10.3 Calculating surface areas	256
10.4 Volume calculations	260
Self-test questions	263
11. The third dimension	265
11.1 Geocentric co-ordinate reference systems	265
11.2 Topocentric co-ordinates	268
11.3 Three-dimensional transformations	270



11.4	Transformation in the case of small rotation angles	271
11.5	The transformation between two reference ellipsoids	272
11.6	Laplace azimuth measurements	275
11.7	Traditional “2D+1D” co-ordinates	277
11.8	Case: the transformation between ED50 and EUREF89	280
11.9	Case: the transformation between ITRF and ETRF	280
	Self-test questions	281
	Exercise 11–1: Greenwich: explain this	282
12.	The Global Positioning System (GPS)	283
12.1	Radio navigation and hyperbolic systems	284
12.2	The GPS satellite	287
12.3	The GPS segments	289
12.4	The GPS constellation	291
12.5	Codes in the GPS signal	291
12.6	GPS receivers	297
12.7	Observables of GPS	300
12.8	GPS measurement geometry	308
12.9	Measurement geometry and sensitivity of observations	309
12.10	Orbits of the GPS satellites	324
12.11	The International GNSS Service IGS	331
	Self-test questions	331
	Exercise 12–1: Calculation of DOP quantities	333
13.	Processing GPS observations	335
13.1	Forming difference observations	335
13.2	Relative (static) GPS	340
13.3	Fixing ambiguities	342
13.4	Real-time positioning	344
13.5	SBAS systems	351
13.6	Real-time support services in Finland	352
13.7	TheGDGPS system	353
	Self-test questions	354
	Exercise 13–1: Geodetic GPS positioning	354
14.	Adjustment calculus in geodesy	357
14.1	Why adjustment?	357
14.2	The average	359
14.3	Linear regression	361
14.4	Theory of least-squares adjustment	362



14.5	Examples of the least-squares method	366
14.6	Linearisation of geodetic models	370
14.7	Propagation of variances	376
14.8	Example: the forward geodetic problem	378
14.9	Observables and observation equations in practice	382
14.10	Tacheometer measurement	388
14.11	Helmert transformation in the plane	389
	Self-test questions	392
	Exercise 14–1: Helmert transformation parameter estimation	392
15.	Statistical methods in geodesy	395
15.1	The method of least-squares	395
15.2	The residuals from the adjustment	397
15.3	Testing, hypotheses for testing	400
15.4	Overall validation	402
15.5	Locating gross errors	405
15.6	Calculation example: linear regression	406
15.7	Adding a gross error	408
15.8	Significance level of the test	409
15.9	Reliability	412
15.10	Deformation analysis	417
	Self-test questions	424
16.	Gravity in geodesy	425
16.1	Measuring gravity	425
16.2	Gravity and geopotential	428
16.3	Gravity anomalies	433
16.4	The gravimetric geoid	435
16.5	The gravity field and heights	439
16.6	Bouguer anomalies	447
16.7	Astronomical position determination	448
16.8	Measuring the gravity gradient	450
	Self-test questions	454
	Exercise 16–1: Gravimetric geoid computation	454
17.	Space geodesy	457
17.1	Earth rotation, orbital motion, sidereal time	457
17.2	Heavenly and Earthly co-ordinates	460
17.3	Väisälä's stellar triangulation	462
17.4	Variations in the Earth's rotation	465



17.5	Precession and nutation of the Earth	469
17.6	Space weather	470
17.7	Satellite orbital motion	473
17.8	Choosing a satellite orbit	475
17.9	Satellite orbital precession, Sun-synchronous orbit	476
	Self-test questions	478
18.	Geodesy and geophysics	479
18.1	Geodynamics	479
18.2	Plate tectonics	483
18.3	Glacial isostatic adjustment (GIA)	488
18.4	Local geodynamics	492
18.5	Deformation monitoring	493
18.6	Studying the Earth's gravity field from orbit	494
18.7	Atmospheric research and GNSS	499
18.8	Long-term variations in Earth rotation axis and orbit	503
18.9	Land-ice research and climate change	505
18.10	Geodetic oceanography	506
	Self-test questions	510
A.	Properties of matrices	513
A.1	Adding and subtracting matrices	513
A.2	Matrices and vectors	514
A.3	The unit matrix	514
A.4	Matrix multiplication	515
A.5	The transpose	516
A.6	The inverse matrix	516
A.7	The determinant of a matrix	518
A.8	Vectorial products	519
B.	A short introduction to magnetohydrodynamics	521
B.1	Plasma	521
B.2	Maxwell's equations	521
B.3	"Frozen-in" magnetic field	522
B.4	History of the field	523
	Self-test questions	523
C.	The Kepler orbital elements for satellites	525
C.1	Angular elements describing the <i>orientation</i> in space of the orbit	525



C.2	Elements describing the orbit's <i>size and shape</i>	525
C.3	Elements describing the satellite's place in its orbit, its "time table"	526
	Self-test questions	526
	Bibliography	527
	Index	549

List of Tables

1.1	Topographic surveying	21
2.1	Measured quantities, units and their symbols	24
2.2	Prefixes indicating order of magnitude in the SI system . . .	25
2.3	Non-SI units accepted for use with the SI	25
2.4	Die-cast statistics	31
2.5	On correlation	40
3.1	Alternative vertical datums <i>A</i> and <i>B</i>	80
4.1	Classification of levelling instruments	98
4.2	Classification of levelling staffs	110
6.1	Calculation template for station adjustment	169
6.2	Traverse calculation template	178
7.1	Calculating the constant and frequency error of a range- finder by linear regression	200
7.2	Examples of distance reductions	207
8.1	Methods for base-network measurement	214
8.2	Goodness of approximation by the reference ellipsoid	220
8.3	Classification of topographic data	233
11.1	Transformation parameters between EUREF89 and ED50 .	280
11.2	Transformation parameters between ITRF2005 and ETRF- 2005	281
12.1	Codes included in the GPS signal	292
12.2	How does dendrochronology work	294
12.3	Start of a RINEX file	300



12.4	Properties of carrier waves	304
12.5	Deriving the influence formula by linearisation	312
12.6	Variants of the DOP quantity	314
12.7	Precise ephemeris in the original SP3 format	329
12.8	DOP calculation script	332
13.1	Effect of forming difference observations on the magnitude of various errors	338
13.2	Summary of GPS observables and difference quantities . . .	340
13.3	Relation between orbit error, length of vector, and position- ing error	341
14.1	Measurement results for linear regression	369
14.2	Point set given in two different co-ordinate frames	392
14.3	Calculation script for Helmert transformation	394
15.1	The planning and measurement process	401
15.2	Rejection bounds for significance levels in a two-sided test based on the standard normal distribution	406
15.3	Example of linear regression	406
15.4	Values of the chi-squared distribution	407
15.5	Example of linear regression calculation	408
15.6	Example of linear regression with a simulated gross error .	409
15.7	Rejection bound and significance level, normal distribution .	411
15.8	Size of gross error and corresponding power of test	411
15.9	Horizontal deformation analysis, co-ordinates	422
16.1	Normal potential and normal gravity according to GRS80 .	431
16.2	Properties of various height types	446
17.1	Kepler's third law for Earth satellites	476

List of Figures

1.1	A lunar eclipse	2
1.2	The grade measurement of Eratosthenes	2
1.3	The Snellius grade measurement	4
1.4	Astronomically determining the difference of plumb-line di- rections	5
1.5	Different mass distribution models for the Earth	6
1.6	Parameters of an ellipsoid of revolution	7



1.7	The Lapland grade measurement	9
1.8	The northernmost point of the Struve chain	10
1.9	Deviations of the plumb line and shape of the geoid	11
1.10	The geodesic in the plane, on the sphere and on the ellipsoid	12
1.11	Spatial planning	19
2.1	A public standard metre in Paris	24
2.2	Examples of different error types	30
2.3	A stochastic quantity on a continuous 2-D value set	33
2.4	Probability density distribution as limit of histograms	33
2.5	Properties of the normal distribution	34
2.6	Probability values for the normal distribution	35
2.7	Some examples of correlation	37
2.8	A two-dimensional probability density distribution	39
2.9	Triangulation by means of a plane table and alidade	43
2.10	Forming a stereo model in photogrammetry	44
2.11	The Greenwich meridian for tourists	48
2.12	Rectangular and geodetic co-ordinates	49
2.13	Geodetic co-ordinates	50
3.1	Depicting the curved surface of the Earth to the map plane using different projections	55
3.2	Systematic shift between road network and aerial-photo- graph base	56
3.3	Imaging the Earth's surface as a narrow zone onto a plane .	58
3.4	Zone division of the KKKJ Gauss–Krüger projection	59
3.5	Geometry of one zone of the Finnish KKKJ system	60
3.6	Scale distortion of Gauss–Krüger and UTM projections	63
3.7	Triangulated affine transformation of the Finnish NLS	64
3.8	Geodetic plane co-ordinates and the quadrants of the plane	65
3.9	A local co-ordinate frame	66
3.10	Temporary co-ordinates	67
3.11	Forward geodetic problem in the plane	68
3.12	Half-angle formula for the arc tangent	70
3.13	Friedrich Robert Helmert	72
3.14	Similarity or Helmert transformation in the plane	72
3.15	Stages of the Helmert transformation in the plane	74
3.16	Fundamental benchmark PP2000 of the N2000 height da- tum at Metsähovi research station	79



3.17	Alternative vertical datums	80
3.18	Two different datums of a horizontal network	82
4.1	Different height types map geopotential numbers in different ways to metric heights	91
4.2	Orthometric heights are metric distances from the geoid	92
4.3	Important reference surfaces and height concepts	93
4.4	The Finnish geoid model FIN2000	94
4.5	The geometry of levelling	96
4.6	Levelling instrument	98
4.7	Telescope	99
4.8	Measuring telescope	100
4.9	Parallax of a measuring telescope	101
4.10	Tubular spirit level	102
4.11	The geometry of a field check	103
4.12	Adjusting the horizon of a levelling instrument	104
4.13	An early self-levelling levelling instrument	106
4.14	A modern self-levelling levelling instrument	106
4.15	Principle of operation of the pendulum compensator	107
4.16	Levelling staffs and foliage	108
4.17	Graduation alternatives for the staff scale	109
4.18	Various temporary levelling-staff supports	111
4.19	Traverse levelling	112
4.20	Area levelling	113
4.21	Self-calculating levelling staff	113
4.22	Principle of operation of a laser level	115
4.23	A profile and cross-sections	115
4.24	Metsähovi research station from the air	117
5.1	An old-fashioned theodolite	120
5.2	Horizontal angle and zenith angle	121
5.3	The axes and circles of a theodolite	122
5.4	Theodolite construction	123
5.5	Forced-centring device or plate	125
5.6	Various monument types	126
5.7	Theodolite axes	127
5.8	Precise levelling of a theodolite using the alidade level	128
5.9	String plummet and rod plummet	129
5.10	Optical plummet	131



5.11	A benchmark seen through an optical plummet	131
5.12	Optical plummet and bull's-eye level used together	132
5.13	Problem situation	132
5.14	Principle of forced centring	133
5.15	Measuring a network using forced centring	134
5.16	Checking an optical plummet	135
5.17	Good targets for horizontal angles	136
5.18	Targeting	137
5.19	Various types of reading microscopes	138
5.20	Optical micrometer and its reading	139
5.21	Reading the graduation circle. One circle location	140
5.22	Reading the graduation circle. Two opposite circle locations	141
5.23	Turning the sight axis by shifting the crosshairs	142
5.24	Trunnion-axis tilt	145
5.25	Observing a zenith angle	146
5.26	Index error	147
5.27	The Gray code	149
5.28	An absolute and an incremental encoding circle	150
5.29	Electronic readout of the horizontal circle	151
5.30	A spinning circle converts measurement of angles into one of time differences	152
5.31	Leica TCA2003 robot tacheometer control panel	153
5.32	Leica TCA2003 robot tacheometer	153
6.1	Use situations for horizontal angle measurement	158
6.2	Intersection and resection	159
6.3	Observation method of complete sets	162
6.4	An open and a closed traverse	171
6.5	An open traverse	172
6.6	The geometry of a closed traverse	175
6.7	Effects of refraction and Earth curvature on zenith-angle measurement	180
6.8	Trigonometric levelling traverse	182
6.9	Heights of instrument and signal	183
7.1	Sag correction of measuring tape	186
7.2	Slope reduction of slant ranges	188
7.3	The phase of a wave motion	188
7.4	The electromagnetic radiation spectrum	190



7.5	Polarisation of electromagnetic radiation	192
7.6	Väisälä's interference method	193
7.7	Dendrochronology	193
7.8	Fizeau's method for measuring the speed of light	195
7.9	One method of electronic phase measurement	196
7.10	Ambiguities, or integer unknowns, are resolved by using several different wavelengths	197
7.11	Corner-cube prism	198
7.12	A prism pack for measurement over long distances	199
7.13	Targeting a signal with a corner-cube prism	199
7.14	The second velocity correction	205
7.15	The terrain correction of distance measurement	205
7.16	Reduction of distance measurement to a reference level	206
8.1	Significance of network hierarchy, mistakes often made	215
8.2	Finnish continuously operating GNSS network FinnRef	216
8.3	Finnish EUREF-FIN first-stage densification network	217
8.4	A triangulation network and a traverse in space	219
8.5	Use of reference ellipsoid and map projection plane when mapping the Earth	220
8.6	Transferring the geometry for adjustment of a small network to the map projection plane	222
8.7	Tools of the right-angle survey method	223
8.8	Right-angle survey	224
8.9	Tie-in survey	225
8.10	Radial survey	226
8.11	Free-stationing survey	228
8.12	Workflow diagram of detail survey	230
8.13	The encoding process for topographic data	232
8.14	Attribute data of objects in multiple layers	233
9.1	Setting out onto the terrain, process description.	237
9.2	Zoning-plan calculation — an example	238
9.3	Setting out using the radial survey method	238
9.4	Straight setting-out method	239
9.5	Rounding of corners with a circular arc	240
9.6	Rounding of corners with a compound curve	241
9.7	Principle of the clothoid	243
9.8	Machine guidance, case Easter Scheldt	248



10.1	Terrain model ETOPO2 version 2 on the Finnish territory	253
10.2	Presentation of terrain models: triangulated network or point-grid	254
10.3	Use of setting-out measures in calculating surface areas	257
10.4	Calculating surface area	258
10.5	A polar planimeter from 1908	259
10.6	Graphical proof of the planimeter equation	259
10.7	Simpson's integration rule in volume calculation	261
10.8	Alternatives for quadrature	262
10.9	Volume calculation from digital terrain models	263
11.1	Inertial and terrestrial co-ordinate reference systems	266
11.2	A right-handed co-ordinate frame	267
11.3	Geocentric and geodetic latitude and transversal radius of curvature	268
11.4	Topocentric and geocentric co-ordinate frames	269
11.5	Differential connection between rectangular and geodetic co-ordinates	274
11.6	Effect of a datum transformation on geodetic quantities	275
11.7	Deviation of local plumb line from ellipsoidal normal	276
11.8	The Laplace phenomenon: the effect of the plumb-line deviation on the azimuth	278
11.9	Greenwich geometry: longitude is a <i>direction</i> , not a <i>place</i>	282
12.1	The Decca system	285
12.2	A Decca receiver	285
12.3	The NNSS Transit system	287
12.4	Positioning satellites	288
12.5	The three segments of the GPS	290
12.6	The original GPS constellation	292
12.7	Principle of phase modulation	293
12.8	Correlation method for determining the travel time of the GPS signal	294
12.9	The various frequencies and effective wavelengths of the GPS signal	296
12.10	Control panel of the Ashtech Z-12	298
12.11	<i>Choke-ring</i> GNSS antenna for precise geodetic work	299
12.12	The electric centre of an antenna is not a self-evident thing	300
12.13	Pseudo-range observation	303



12.14	Measurement of the phase of the GPS signal's carrier wave	305
12.15	Propagation of a wave packet in a dispersive medium	306
12.16	Geometry of GPS positioning	308
12.17	Geometry between a GPS satellite and an observation site .	311
12.18	DOP and error ellipsoids and mean error of unit weight . . .	317
12.19	DOP ellipsoid, principal axes along co-ordinate axes	317
12.20	The circle singularity or "dangerous circle" for GPS	322
12.21	Calculation example of DOP quantities	323
12.22	The six orbital planes of GPS satellites in the Helsinki sky .	325
12.23	Satellite orbital motion described by position and velocity vectors	326
12.24	The tracking stations of the IGS	330
13.1	The "common-mode" error assumption	335
13.2	Forming difference observations of various types	337
13.3	Double difference, short distance between GPS receivers . .	341
13.4	One-dimensional ambiguity resolution, ranging	342
13.5	Various ambiguity-resolution methods	343
13.6	Principle of operation of the DGPS method	346
13.7	Principle of operation of the RTK method	348
13.8	Satellite-based augmentation systems (SBAS)	352
14.1	Triangulation network	358
14.2	Metaphor: a large weight means a small correction	360
14.3	The idea of linear regression	361
14.4	Linear regression, definitions of quantities	362
14.5	Calculation example of linear regression	369
14.6	One-dimensional mapping and linearisation	372
14.7	A two-dimensional mapping	374
14.8	Quantities related to the error ellipse	378
14.9	Geometry of azimuth measurement	385
14.10	Geometry of zenith-angle measurement	387
15.1	Least-squares adjustment as an orthogonal projection	399
15.2	Chi-squared distribution with four degrees of freedom	403
15.3	Example of linear regression, error in observation 3	410
15.4	Statistical testing based on normal distribution	410
15.5	Harmonisation of the significance levels of the overall vali- dation and per-observation tests	413
15.6	An example of reliability	413



15.7	Another example of reliability	414
15.8	Height deformation monitoring network	419
15.9	Horizontal deformation monitoring network	421
16.1	Absolute or ballistic gravimeter	427
16.2	Relative or spring gravimeter	427
16.3	The terrain height depicted by height contours, and height gradients	428
16.4	Geopotential table	429
16.5	Normal gravity field of the Earth	431
16.6	Equipotential surfaces and lines of force of geopotential and normal potential	432
16.7	Equipotential surfaces of true and normal gravity field . . .	433
16.8	True and normal gravity vectors	434
16.9	Relationship between gravity and geoid height variations .	436
16.10	Geometry of the Stokes integral equation	437
16.11	Global geoid model EGM2008	439
16.12	The gravity vector is the <i>gradient</i> of the geopotential	440
16.13	The path integral of work	441
16.14	Heights and equipotential surfaces	442
16.15	Free-air and Bouguer anomalies for Southern Finland . . .	448
16.16	Root of a mountain range and its effect on the plumb line . .	449
16.17	A levelling instrument converted to astrolabe	450
16.18	Gravitational-gradient or tidal force field	451
17.1	Earth orbit around the Sun, and apparent path of the Sun across the celestial sphere	458
17.2	The vernal equinox and its movement, <i>precession</i>	459
17.3	Hour angle and declination on the celestial sphere	460
17.4	Yrjö Väisälä's stellar triangulation (a)	462
17.5	Yrjö Väisälä's stellar triangulation (b)	463
17.6	Satellite geodesy from NASA photographic archives	465
17.7	Polar motion for the period 1970–2000	466
17.8	Polar motion causes variations in station latitudes	467
17.9	The Earth's precession	470
17.10	The corona of the Sun	471
17.11	Sunspots and their magnetic field lines	472
17.12	Space weather, the magnetosphere	473
17.13	Ellipse, definition and how to draw one	474



17.14 Kepler's orbital elements	475
17.15 Sun-synchronous orbit	477
18.1 A LAGEOS satellite	481
18.2 Principle of operation of very long baseline interferometry	482
18.3 Metsähovi radio telescope used for VLBI measurements	483
18.4 Wegener's continental drift theory, Mid-Atlantic Ridge	484
18.5 Internal structure of the Earth	485
18.6 Palaeomagnetism and sea-floor spreading	486
18.7 Global plate tectonics	487
18.8 Mechanisms of plate tectonics	488
18.9 Post-glacial land uplift in Fennoscandia	490
18.10 Horizontal and vertical motions in Fennoscandia as determined by the BIFROST project	491
18.11 InSAR image	495
18.12 Determining the Earth's gravitational field by tracking the orbit of a low-flying satellite	496
18.13 Basic idea of the GRACE satellite pair	497
18.14 Determining the Earth's gravitational field with the gravitational gradiometer on-board the GOCE satellite	498
18.15 Sea-surface topography map produced by GOCE	499
18.16 Saturation partial pressure of water vapour	501
18.17 Use of GNSS for studying the troposphere	501
18.18 GNSS radio occultation	503
18.19 Milanković cycles over the past 800 000 years	504
18.20 Measurement geometry of satellite radar altimetry	506
18.21 Link between sea-surface topography and ocean currents	507
18.22 Tide gauges of the Baltic Sea, Seasat ground tracks, and mean sea level	511
18.23 <i>Schematic of the Earth's orbital changes (IPCC)</i>	512



Acronyms

ABCDEFGHIJKLMNOPQRSTUVWXYZ

A

- ANNA 1B** (“Army, Navy, NASA, Air Force”), 1962-60A, geodetic satellite 464
- APPS** Automatic Precise Positioning Service (JPL) 355
- ARP** antenna reference point 298
- ATR** automatic target recognition 150, 154, 155
- AUSPOS** Australian online GPS processing service 355, 356

B

- BGI** *Bureau Gravimétrique International*, International Gravimetric Bureau 448
- BIFROST** Baseline Inferences for Fennoscandian Rebound, Sea level, and Tectonics, a Nordic geodynamics research project 491
- BIPM** *Bureau International des Poids et Mesures*, International Bureau of Weights and Measures 23

C

- C/A code** Coarse / Acquisition, Civilian Access GPS code 291–293, 295–297, 300, 303, 347
- CAD** computer-aided design 21, 53, 255
- Caltech** California Institute of Technology 495
- CCD** charge-coupled device, image sensor type 107, 149, 150, 154
- CDMA** code division multiple access 293, 344
- CERN** *Organisation européenne pour la recherche nucléaire*, European Organization for Nuclear Research 114
- CHAMP** 2000-39B, Challenging Minisatellite Payload, German satellite 494, 502
- CIO** Conventional International Origin, reference pole for polar motion 468

D

- Decca** marine navigation system 45, 284–287, 331, 348
- DEM** digital elevation model 251, 256
- DGPS** differential GPS 345–347, 349, 350, 353, 354

- DHM** digital height model 251
- DNA** deoxyribonucleic acid, helix-shaped macromolecule that carries and replicates the genetic information of almost all Earth's organisms 189
- DOP** dilution of precision, a measure for the geometric strength of satellite positioning. xxv, xxvii, xxviii, 312–318, 323, 332, 333
- DORIS** Doppler Orbitography and Radiopositioning Integrated by Satellite, a French satellite positioning system 214, 480, 483
- DTM** digital terrain model 251, 256, 447
- DVD** Digital Versatile Disc 230
- DWT** discrete wavelet transform 255
- E**
- ECEF** Earth-centred, Earth-fixed 266
- ED50** European Datum 1950 11, 272, 277, 279–281
- EEST** Eastern European Summer Time 48
- EET** Eastern European Time 48
- EGM2008** Earth Gravity Model 2008 118, 439, 448
- EGM96** Earth Gravity Model 1996 118
- EGNOS** European Geostationary Navigation Overlay System, an SBAS for the European area 351
- eLoran** navigation system, planned (“Enhanced Loran”) 284
- EOP** Earth orientation parameters 85, 331, 468
- ERT** electrical resistivity tomography 248
- ET** ephemeris time, efemeridiaika 468
- ETRF** European Terrestrial Reference Frame 77, 281
- ETRS** European Terrestrial Reference System. Coincides with ITRS for the epoch 1989.0. Also called ETRS89 xxiv, 57, 58, 77, 86, 279, 281
- ETRS-GK** Gauss–Krüger map projection system for Finland 62, 235, 392
- ETRS-TM35FIN** UTM map projection for Finland, zone 35 61, 62, 87
- EUREF IAG** Reference Frame Subcommittee for Europe 57, 86, 281, 354
- EUREF89** First European realisation of ETRS89 57, 280
- EUREF-FIN** Finnish national realisation of ETRS89 48, 57, 58, 61, 85, 213, 214, 217, 233, 235, 279–281
- F**
- FAT** file allocation table (file system) 155
- FDMA** frequency division multiple access 344
- FGI** Finnish Geodetic Institute, 1918–2015, Finnish Geospatial Research Institute, 2015– 79, 216, 217, 354, 483, 510
- FIN2000** Finnish geoid model 94
- FIN2005N00** Finnish geoid model 279
- FRS** Fellow of the Royal Society (of London) 188, 194, 395, 420, 486, 504, 522



FRSE Fellow of the Royal Society of Edinburgh 188, 194, 420, 486, 522

G

GAST Greenwich Apparent Sidereal Time 459, 461, 478

GCM general circulation model 256

GCP ground control point 254

GDOP geometric DOP 312, 314, 316

GIA glacial isostatic adjustment 85, 488, 490, 509

GIS geographic information system 21

GLONASS Global Navigation Satellite System (Russian) 283, 284, 318, 330, 341, 344, 350, 352, 479

GMT Greenwich Mean Time 48

GNSS Global Navigation Satellite Systems, generic name xxv, xxvii, 21, 45, 47, 85, 86, 118, 126, 133, 158, 170, 184, 198, 204, 212, 214, 216–218, 221, 228, 234, 236, 247, 280, 282, 284, 302, 328, 330, 331, 341, 342, 344, 353–355, 433, 465, 468, 479, 480, 482, 491–493, 497, 499–503, 506, 507, 509–511

GOCE Gravity Field and Steady-State Ocean Circulation Explorer 95, 497–499, 508, 511

GPR ground-penetrating radar 248

GPS Global Positioning System ii, xxiii, xxv, xxvii, xxviii, 14, 45, 46, 48, 93, 118, 163, 192, 193, 197, 203, 217, 234, 281, 283, 284, 286, 287, 289–300, 302, 303, 305–309, 311–314, 317–319, 321, 322, 324–328, 331–333, 335, 338, 340–345, 347, 349–351, 354, 382, 469, 473, 476, 479, 482, 488, 494, 496, 502, 503, 525

GPS/MET 1995-17C, **GPS** radio occultation satellite 502

GPU geopotential unit, $10\text{m}^2/\text{s}^2$ 441, 442

GRACE 2002-12A, 2002-12B, Gravity Recovery and Climate Experiment. This was a pair of satellites 496, 497, 505, 511

GRS80 Geodetic Reference System 1980 49, 61, 84, 86, 94, 118, 219, 279, 431, 439, 476

GSI (Leica) Geo Serial Interface 154, 155

H

HDOP horizontal DOP 314–316, 318, 321, 324, 333

I

IAG International Association of Geodesy xxiv, 11, 57, 202, 331, 466, 480

IB inverted barometer 507

IERS International Earth Rotation and Reference Systems Service 466, 468

IGS International GNSS Service 328, 330, 331, 355, 503

i.i.d. independent and identically distributed 364



- InSAR** interferometric synthetic-aperture radar 494, 495
- INSPIRE** Infrastructure for Spatial Information in the European Community 252
- IONEX** Ionosphere Map Exchange Format 503
- IP** Internet Protocol 481
- IPCC** Intergovernmental Panel on Climate Change 512
- ITRF** International Terrestrial Reference Frame, a realisation of the **ITRS** 58, 86, 281
- ITRS** International Terrestrial Reference System xxiv, xxvi, 272
- J**
- Jason** 1–3, Joint Altimetry Satellite Oceanography Network, series of radar altimetric satellites 509, 512
- JHS** Julkisen hallinnon suosituksset, Recommendations for Public Administration 213
- JPEG 2000** image format 255
- JPL** Jet Propulsion Laboratory, **NASA** xxiii, 331, 353, 355, 495
- JUHTA** Julkisen hallinnon tietohallinnon neuvottelukunta, Advisory Committee on Information Management in Public Administration 213
- K**
- KKJ** *Kartastokoordinaattijärjestelmä*, Finnish National Map Grid Co-ordinate System (obsolete) xxviii, 46, 56–62, 67, 85, 233, 277, 279, 280, 350, 392
- KM10** Finnish national terrain model, resolution 10m 252
- KM2** Finnish national terrain model, resolution 2m 252
- L**
- LAGEOS** 1–2, Laser Geodynamics Satellite 481
- LAST** Local Apparent Sidereal Time 459, 460
- LCD** liquid crystal display 154
- LED** light-emitting diode 44, 196
- LHC** Large Hadron Collider 114
- LoD** length of day 85, 331, 466
- Loran-C** marine navigation system 284
- M**
- MHD** magnetohydrodynamics 472, 523
- MIF** member of the *Institut de France*. The Institute comprises five learned academies including the Academy of Sciences 194
- MRI** magnetic resonance imaging 469
- MSAS** Multi-functional Satellite Augmentation System (Japan) 351



N

- N2000** Finnish height system, epoch 2000.0 78, 79, 93, 96, 111, 213, 235, 279
- N60** Finnish height system, epoch 1960.0 78, 93, 96, 111, 277, 279
- NAP** Normaal Amsterdams Peil, Amsterdam Ordnance Datum, a Western European height datum 78, 93
- NASA** National Aeronautics and Space Administration, US xxiii, xxvi, 353, 464, 465, 495
- NGS** National Geodetic Survey, US 86
- NLS** National Land Survey of Finland 86, 87
- NNSS** Navy Navigation Satellite System, "Transit", "Doppler" 284, 286, 287
- NOAA** National Oceanic and Atmospheric Administration, US 252, 499
- NUVEL** global plate-motion model 487
- NWP** numerical weather prediction 256, 466

O

- Omega** marine navigation system 284

P

- P code** Precise / Protected GPS code 291–293, 295–297, 300, 303, 347
- PAGEOS** 1966-56A, Passive Geodetic Earth Orbiting Satellite 464
- PC** personal computer 155, 300
- PCMCIA** memory card bus standard 155
- PDOP** position DOP 314–316, 318, 333
- PPP** Precise Point Positioning, a precise geodetic positioning technique using a single GNSS receiver 319, 353

R

- RINEX** Receiver-Independent Exchange Format 300, 302, 333, 353–355
- RS232** Recommended Standard 232, serial interface 155
- RTCM** or **RTCM-SC104**, "Radio Technical Commission for Maritime Services Special Committee 104", a popular differential GNSS standard xxvii, 352
- RTK** real-time kinematic positioning xxviii, 21, 170, 214, 218, 222, 228, 234, 236, 319, 347–350, 352–354

S

- SA** selective availability (GPS) 345
- SAR** synthetic-aperture radar 252, 494, 505
- SBAS** satellite-based augmentation systems xxiv, xxviii, 351, 352, 354
- Seasat** 1978-64A, radar altimetric satellite 511
- SI** *Système International d'Unités*, International System of Units 23–26, 185, 426, 441, 442, 446
- SMS** Short Message Service (mobile telephony) 21



SoL Safety of Life [291](#), [351](#)

SOPAC Scripps Orbit and Permanent Array Center, San Diego, USA [354](#)

SP3 Standard Product 3, precise ephemeris data format [328–330](#)

SRTM Shuttle Radar Topography Mission [252](#), [264](#)

SST satellite-to-satellite tracking [497](#)

T

TAI *Temps Atomique International*, International Atomic Time [468](#), [469](#)

TDOP time [DOP](#) [314](#), [316](#)

TEC total electron content [503](#)

TIN triangulated irregular network [255](#)

TOPEX/Poseidon 1992-52A, radar altimetric satellite [509](#), [510](#), [512](#)

U

UDP User Datagram Protocol [481](#)

UNAVCO A non-profit university-governed consortium that facilitates geoscience research and education using geodesy. Formerly University NAVSTAR Consortium [118](#), [488](#)

USB Universal Serial Bus [155](#), [230](#)

UTC Universal Time, Co-ordinated, global civil time scale [468](#), [469](#)

UTM Universal Transverse Mercator (map projection) [xxiv](#), [58](#), [61–63](#), [154](#), [208](#), [221](#), [279](#)

V

VDOP vertical [DOP](#) [314](#), [315](#), [318](#), [321](#), [324](#), [333](#)

VGOS [VLBI](#) Global Observing System [483](#)

VHS Video Home System [481](#)

VLBI very long baseline interferometry [xxviii](#), [193](#), [214](#), [465](#), [468](#), [480](#), [482](#), [483](#)

VRS-RTK virtual reference station [RTK](#) [352](#)

VVJ Vanha valtion järjestelmä (obsolete) [57](#), [61](#), [279](#)

W

WAAS Wide Area Augmentation System, an [SBAS](#) for the North American area [351](#)

WADGPS wide-area differential [GPS](#) [351](#)

WGS84 World Geodetic System 1984 [58](#)

Y

YKJ *Yhtenäiskoordinaatisto*, [KKJ](#)'s Uniform Co-ordinate System [61](#), [62](#)





The history and societal status of geodesy

1

[...] Nous avons été sur le fleuve, fort incommodés de grosses Mouches à tête verte, qui tirent le sang par-tout où elles picquent ; nous nous trouvâmes sur Niwa [Nivavaara], persécutés de plusieurs autres espèces encore plus cruelles.

Maupertuis (1738), PDF page 44, page 16



1.1 The figure of the Earth, early conceptions

In traditional societies, undoubtedly the most common conception of the figure of the Earth was that the Earth is a flat disc extending to the horizon, with the sky curving like a dome over her. On the inner surface of the dome, the celestial bodies describe their complicated orbits. Children also have generally the same conception. Only with formal education does this “naïve world model” give way. Psychologically, from the viewpoint of childhood development, this is by no means an easy process, surely as difficult as it was for society as a whole, back in time as a stage in the historical development of science.

However, the antique Hellenes were already aware of the spherical shape of the Earth. Free of preconceptions, they had observed how, during a lunar eclipse, the Earth cast her shadow on the surface of the Moon. They also observed that a lunar eclipse that was high in the sky at one end of the Mediterranean happened near the horizon at the other end. Assuming that this was one and the same event, this could only mean that the Earth’s surface must be curved at least in the east-west direction. And the colder climates found further north were a sign to early observers that the Earth is also curved in the south-north direction.

Eratosthenes, the “father of geography”, lived 276–195 BCE.¹ He was ¹

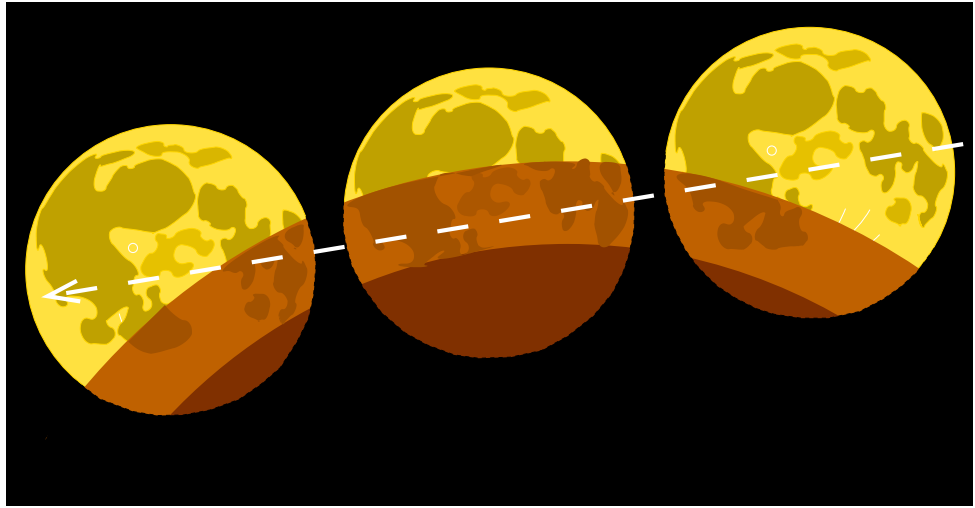


FIGURE 1.1. A lunar eclipse. The shape of the shadow, always circular, shows that the Earth must be a sphere.

the first to measure the size, or radius, of this spherical Earth. The measurement was the same in principle as the later grade measurements: measure the length of an arc on the surface of the Earth by geodetic means, and the difference in direction between the plumb lines at the ends of the arc by astronomical means. By combining the length ℓ of the arc and the difference between the plumb-line directions γ one obtains

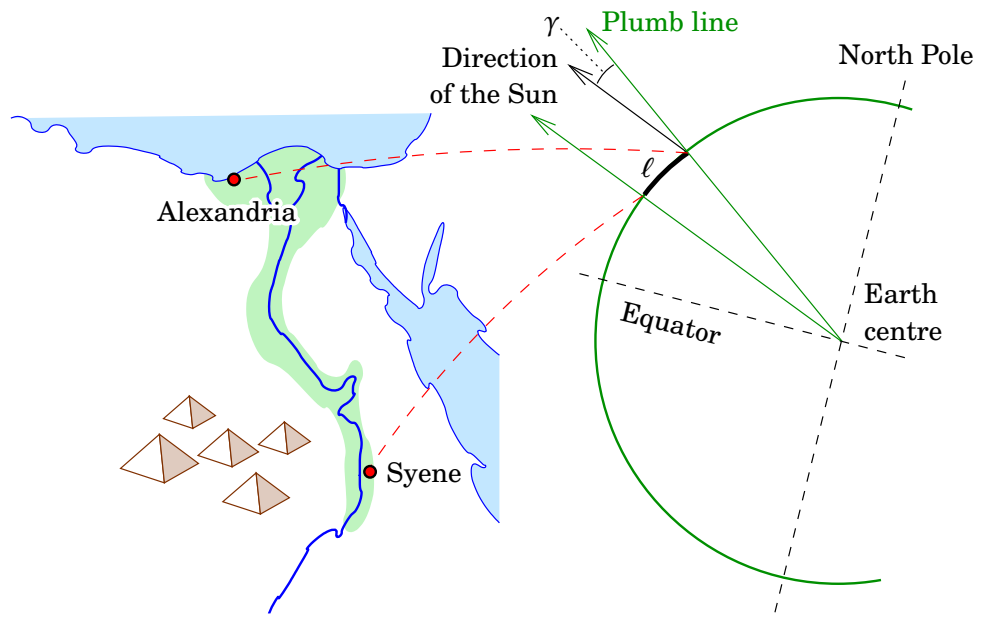


FIGURE 1.2. The grade measurement of Eratosthenes.

¹“Before the common (or Christian) era”.

for the radius of the Earth

$$R = \frac{\ell}{\gamma}. \quad (1.1)$$

See figure 1.2.

The information on the directions of the plumb lines was obtained from the midsummer Sun, which in Syene (today's Aswan) did not throw any shadows at all.² In Alexandria, on the other hand, the Sun was not in the zenith but, based on the lengths of shadows, some fiftieth part of a circle further to the south. Eratosthenes obtained a value for the radius of the Earth³ of 6317–7420 km — pretty close to the current best value of 6371 km.

More information can be found in Torge (2001), pages 5–6.

The principle of *triangulation*, so important in geodesy, that, in a network consisting of triangles, the geometry may be uniquely determined if, in addition to the angles of the triangle, only *one distance* is measured, was presumably discovered by Gemma Frisius⁴ in 1533 (Crane, 2002, pages 56–57).

The use of the method for grade measurement also happened for the first time in the Netherlands, using the numerous church towers dotting the prosperous but flat country. Snellius⁵ was among the first to use triangulation to determine the length of an arc ℓ . By measuring one length in the network, and otherwise only angles, he managed to determine the distance between two cities, Bergen op Zoom and Alkmaar, although the cities are separated by the broad river branches of the Rhine delta. See figure 1.3.

The secret of triangulation is that with the aid of angle measurements one can build, either computationally or graphically, a *scale model* of the whole measurement network, where all proportions and shapes are correct. To determine the true scale, it suffices to measure just *one distance* in the model also in reality. In the case of Snellius, this was the distance pq , in the meadow by Leiden, a *baseline* of only 326 “roeden”.⁶

²The story that he used the circumstance that the Sun illuminated the bottom of a well is apparently a misunderstanding (Dreyer, 1914).

³In fact he obtained his results in a unit called the *stadium*, the length of which varies. The length used by Eratosthenes is controversial.

⁴Gemma Frisius (1508–1555) was a Dutch polymath.

⁵Willebrord Snell van Rooyen (1580–1626) was a Dutch astronomer and mathematician.

⁶The version of the unit used by Snellius was the “*Rijnlandse roede*”, 3.766 m.



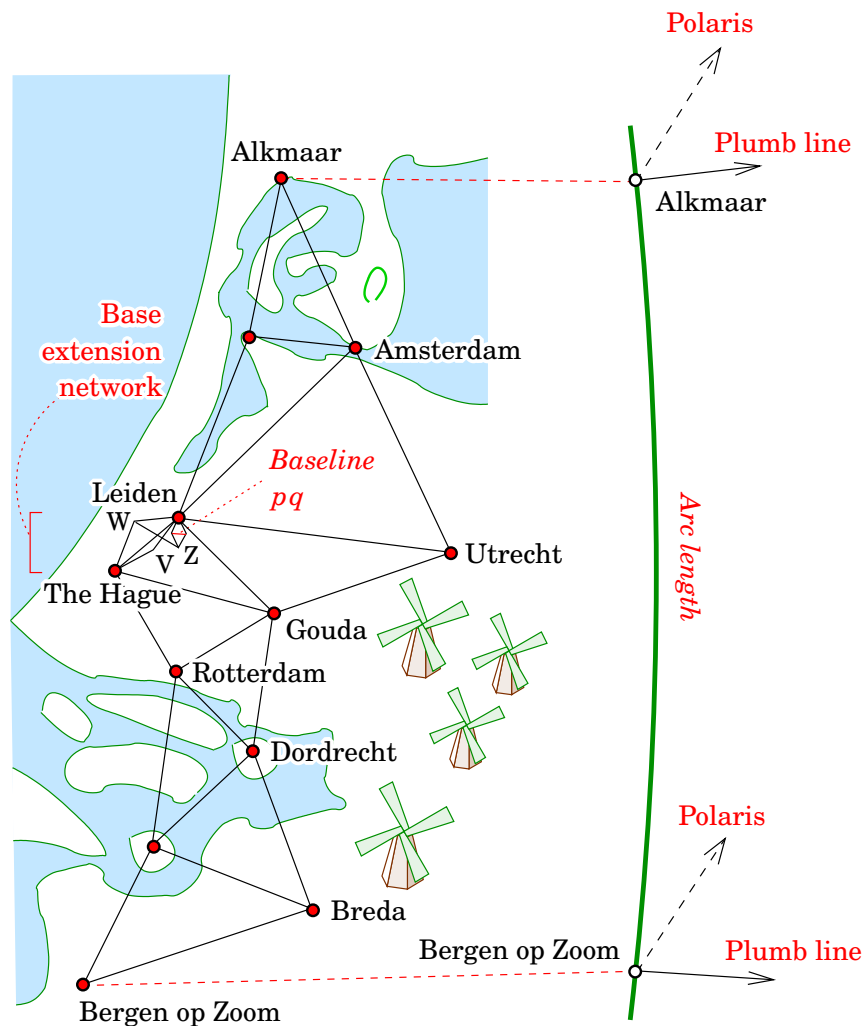


FIGURE 1.3. The Snellius grade measurement. The length of baseline pq is 326.45 *roeden* (1229m). This length was derived through a local base extension network from the only measured length, the original baseline of length 87.05 *roeden* (328m) (personal communication L. Aardoom).



V = Voorschoten, W = Wassenaar, Z = Zoeterwoude.

Using *astronomical position determination*, one may determine the difference between the directions of the plumb line in two locations, see figure 1.4. When travelling along the meridian in the north-south direction, the *absolute direction* of the local plumb line, the direction with respect to the stars, changes. The local *plane of the horizon*, always perpendicular to the plumb line, the local direction of gravity, also turns by the same amount in the same direction.

The direction in space of the rotation axis of the Earth is very stable due to the gyroscope phenomenon. It points to a place in the sky near



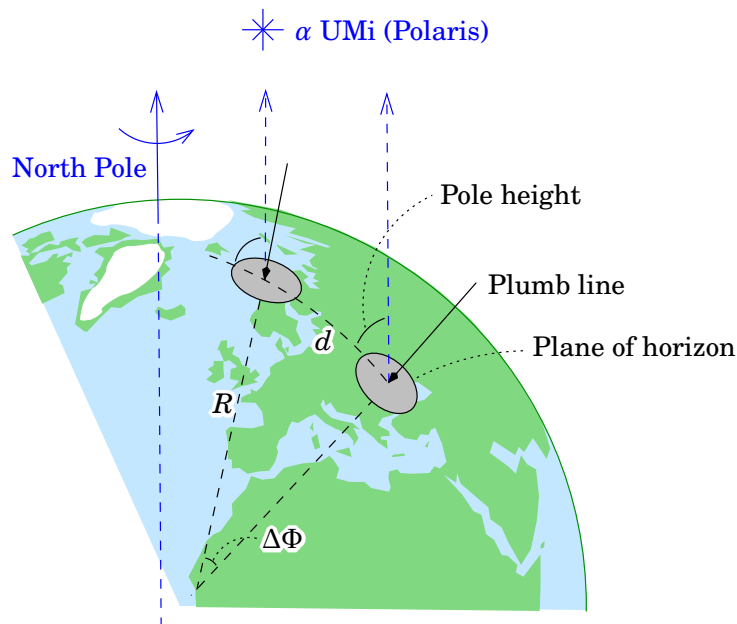


FIGURE 1.4. Astronomically determining the difference in the north-south direction $\Delta\Phi$ of plumb-line directions. From the direction difference and the metric distance d , one can determine the radius of curvature of the Earth $R = d/\Delta\Phi$, assumed constant.

the star α Ursae Minoris, or Polaris. This star gives us the direction of the north. The *latitude* Φ of a location is obtained by determining astronomically the elevation angle of this *celestial pole* above the horizon. This is easiest to do using Polaris, although a precise determination is a little more involved.

By thus measuring astronomically the difference in plumb-line directions between Alkmaar and Bergen of Zoom, and combining this with the metric distance obtained by triangulation, Snellius managed to determine the radius of curvature of the Earth. The method is referred to as *grade measurement*.

1.2 Newton's laws and the figure of the Earth

The understanding of the figure of the Earth made a great leap forward when Newton⁷ published in 1687 his main work, the *Principia*⁷ (*Philosophiæ Naturalis Principia Mathematica*, “Mathematical principles of natural philosophy [physics]”). In this opus he created the foundations of the whole of classical mechanics, including celestial mechanics.

⁷Sir Isaac Newton (1642–1727) was an English physicist and mathematician, the father of classical mechanics.

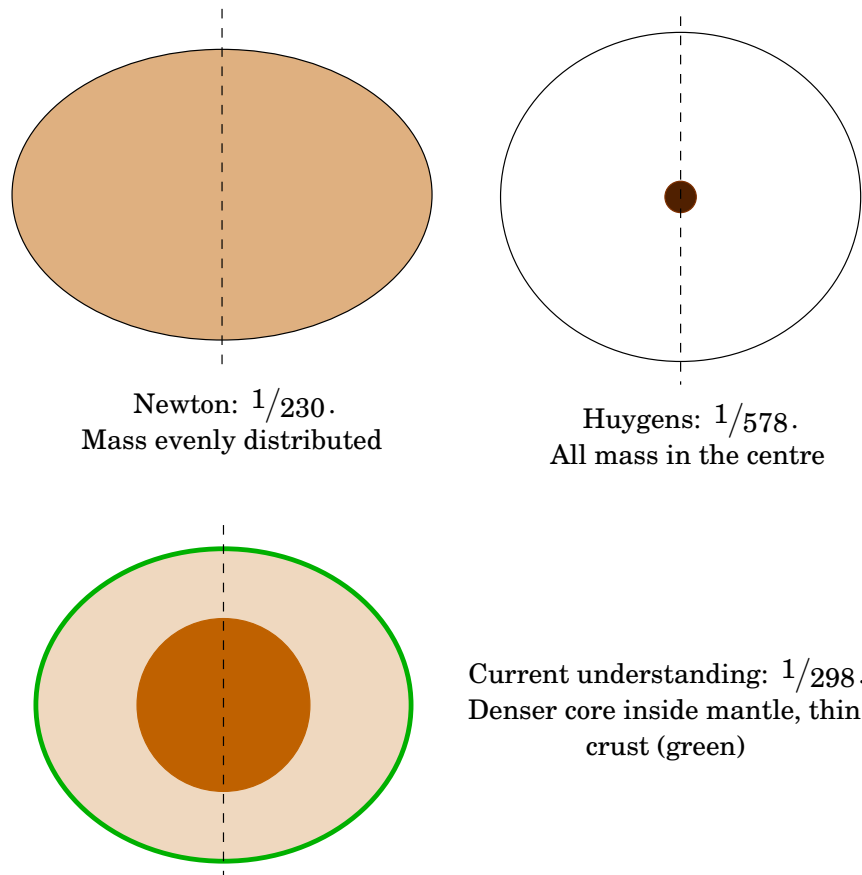


FIGURE 1.5. Different mass distribution models for the Earth, and their theoretical flattening values.

The universal law of gravitation Between two masses m_1 and m_2 acts an *attraction* F of strength

$$F = G \frac{m_1 m_2}{r^2},$$

in which r is the distance separating the masses. The constant G is Newton's universal gravitational constant, the value of which is $6.674 \cdot 10^{-11} \text{ m}^3/\text{kg s}^2$.

This attraction acts between *all* pairs of masses. So, not only does the Earth's attraction act on the Moon and the Sun's attraction on the Earth, but the Moon's attraction also affects the Earth, etc. In geophysics again, we know that the attraction works between all *parts* of the Earth: the sea, atmosphere, mountains all affect the gravitational field surrounding the Earth. And, because our Earth consists of materials that — however more or less reluctantly — deform under the influence of external force, gravitation also shapes the physical figure of the Earth.

In the *Principia*, Newton calculated, using his famous laws, that a



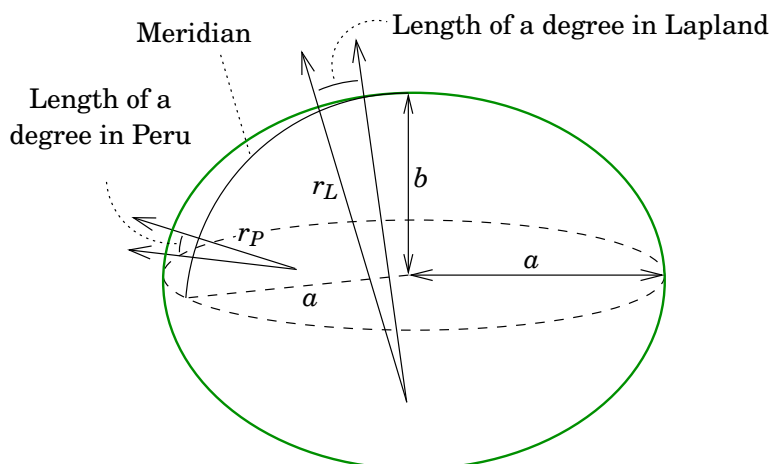


FIGURE 1.6. Parameters of an ellipsoid of revolution.

homogeneous, liquid Earth, in equilibrium and rotating once in 24 hours, with gravitation acting between its elements of liquid, would be *flattened* at the poles by centrifugal force (figure 1.6). The definition of the flattening (oblateness) is

$$f = \frac{a - b}{a}, \quad (1.2)$$

in which a and b are the semi-major and semi-minor axes of the Earth ellipsoid; in other words, the equatorial and polar radii.

iso- ja
pikkuakselin
puolikkaat

The theoretical flattening calculated by Newton was $f = 1/230$.

The assumption that the Earth is of homogeneous density is not correct. Christiaan Huygens calculated in 1690, by assuming that all the Earth's mass is concentrated in her centre — or at least, that the Earth's attraction emanates from her centre — that the flattening would only be $f = 1/578$. As we know today, the truth lies between these two extremes: the density of the Earth's crust is about 2.7 g/cm^3 , that of the underlying mantle is $3.0\text{--}5.4 \text{ g/cm}^3$, and the density of the iron core of the Earth is $10\text{--}13 \text{ g/cm}^3$. The average density of the whole Earth is about 5.4 g/cm^3 . So, while the density increases a great deal toward the centre of the Earth, a large part of the Earth's mass is nevertheless far from her centre.

In Newton's days there were influential scientists, like the astronomer Cassini,⁸ who believed that the Earth was elongated like a rugby ball, $b > a$, and not flattened. An empirical answer to the question was needed!

The flattening issue remained unsolved until half a century later, when

⁸Jean Dominique (Giovanni Domenico) Cassini (1625–1712) was an Italian-French astronomer — explorer of the Saturn system — mathematician, map maker, and engineer.



the French Academy of Sciences organised two expeditions, one to Finnish Lapland — then part of the Swedish empire — (1736–1737), the other to Peru, South America (1735–1744). The goal of the expeditions was to measure, by geodetic and astronomical means, the *length of a meridian arc of one degree* on two different latitudes, one close to the equator in Peru, the other close to the North Pole in Lapland in the Torne river valley. This was thus a similar *grade measurement* to the one Snellius had carried out over a century earlier. . . . but this measurement took place far away from the home country, in strange lands in different climate zones, one of them even beyond the ocean.

The idea of the measurement is illustrated in figure 1.6. Using astronomical measurements, a baseline is established in the north-south direction, at the end points of which the directions of the plumb line differ from each other *by one degree*. Over land, the distance between the points ⁹ is measured in metres.⁹ If Newton was right, the length of a degree close to the North Pole would be greater than that of one close to the equator, in other words, the *radius of curvature* of the Earth would, at the poles, be longer than at the equator:

$$r_L > r_P.$$

The joint result of both expeditions was an empirical flattening of $f = 1/210$. For comparison, the current best value for the flattening of the Earth is $f \approx 1/298.257$.

Much has been written about the adventures of the expedition led by Pierre L. M. de Maupertuis in the Torne river valley 1736–1737, for example [Rovaniemi, The Degree Measurement Expedition](#), and in the French original ([Maupertuis 1738](#)).

¹⁰ Of the later grade measurements we may mention Struve's¹⁰ Russian-Nordic grade measurement (the “Struve chain”) 1816–1855, which extended from Norway’s Atlantic coast all the way to the Black Sea ([Wikipedia, Struve Geodetic Arc](#)). Some points of the chain have also been preserved on Finnish territory.

⁹In reality, the French Academy of Sciences measurements used the *toise* as the unit of length, as the metre had not been invented yet.

¹⁰Friedrich Georg Wilhelm von Struve (1793–1864) was a Russian astronomer and geodesist.





FIGURE 1.7. The grade measurement project of the French Academy of Sciences: the Lapland grade measurement network.



1.3 The mathematical figure of the Earth or *geoid*

The changes from place to place in the direction of the plumb line along an arc on the Earth’s surface can thus be used to find out about the true figure of the Earth. In the previous section we described how the grade-measurement project of the French Academy of Sciences exploited this phenomenon for determining the figure of the Earth, *assuming* that the Earth has the figure of an ellipsoid of revolution.

With the aid of more precise geodetic measurements it was noticed that this assumption does not *precisely* apply. Already in the context of the Peru grade measurement, Pierre Bouguer¹¹ noticed that the direction of the plumb line on both sides of the Andes had a tendency to deflect

¹¹Pierre Bouguer (1698–1758) was a French polymath, mostly a geophysicist and shipbuilder.





FIGURE 1.8. The northernmost point of the Struve chain in Fuglenes, Norway, Franz (2005).



towards the mountain range, and he interpreted this correctly as an expression of Newtonian *gravitation* or attraction. George Everest¹² in India noticed the same phenomenon near the Himalayas. As geodetic measurements, especially astronomical determinations of the direction of the plumb line, progressed, the understanding spread that the figure of the Earth is irregular.

People started to speak about the “mathematical figure of the Earth” or *geoid* (J. B. Listing,¹³ 1873), the continuation of mean sea level under the continental masses, a surface that is everywhere perpendicular to plumb lines, and along which a fluid at rest — like sea water — would settle. See figure 1.9.

In 1862, under the leadership of the Prussian J. J. Baeyer,¹⁴ the “*Mittel-europäische Gradmessung*” (“Central European Grade Measurement”)

¹²Sir George Everest (1790–1866) was a geodesist and geographer born in Wales, director-general of the Survey of India. In 1865 the highest peak in the Himalayas was named Mount Everest in his honour but against his protestations.

¹³Johann Benedict Listing (1808–1882) was a German mathematician, the inventor of topology.

¹⁴Johann Jacob Baeyer (1794–1885) was a Prussian military officer, geodesist, and diplomat of geodesy.



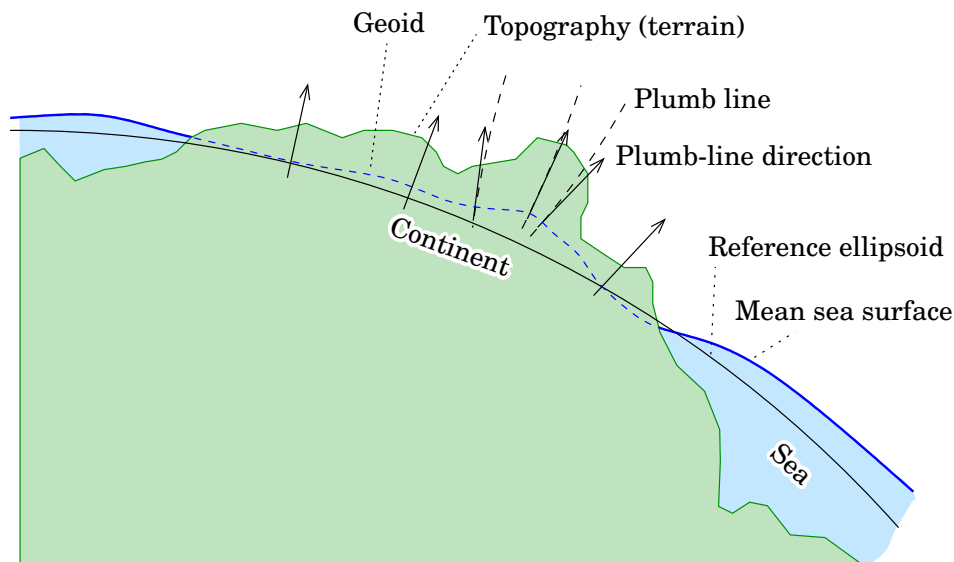


FIGURE 1.9. Deviations of the plumb line and the shape of the geoid.

was established, which later developed into the global organisation [IAG](#), *International Association of Geodesy*. Its task was determining the figure of the Earth or geoid, especially on the European territory, and uniting the geodetic networks of Europe into a single network. This objective was not properly achieved until 1950, when the first joint European network adjustment [ED50](#), “European Datum 1950”, was completed, even though

verkkotasoitus

only on the Western European territory. Elsewhere, for example in North America, continental-scale triangulation networks were being measured, to determine the figure and flattening of the Earth as well as the locations of points on the Earth’s surface in support of map-making. Determining the general figure of the Earth is however difficult from the Earth’s surface using classical geodetic techniques, because extended networks on the Earth’s surface are not geometrically strong, and their unification across oceans is impossible.

kolmiomittaus-
verkko

Satellites have fundamentally changed this picture: satellite techniques have provided precise data on, for example, the flattening of the Earth by exploiting the rapid precession of the satellite orbital plane it causes. Several weeks after the launch of Sputnik, much better values were already becoming available for the flattening, and the American Vanguard 1 satellite showed the Earth to be “pear shaped” — although only very, very slightly.

ratatason
prekessio



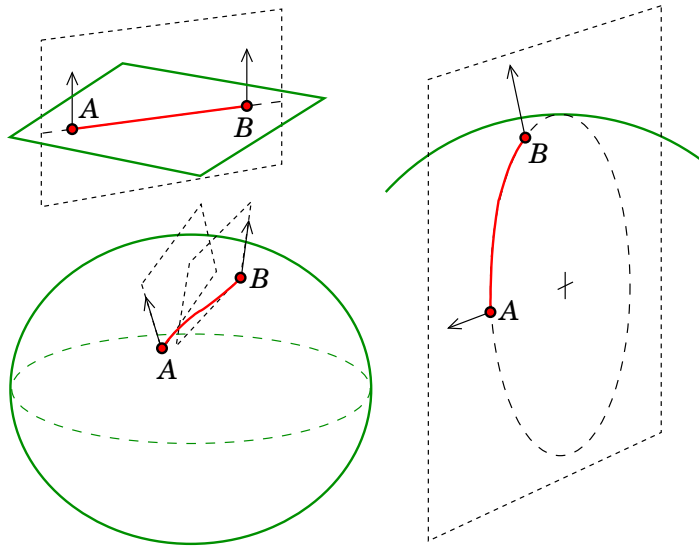


FIGURE 1.10. The geodesic in the plane, on the sphere and on the ellipsoid of revolution. The arrows depict the local normal to the surface. The plane of the normal section is not uniquely defined for the ellipsoid of revolution.



1.4 The geodesic

pyrähdy-
ellipsoidi
isoympyrä
15

In the plane, the shortest path between two points is the *straight line*. On the curved surface of a sphere, or of an ellipsoid of revolution, the shortest path is a curve. In the case of the sphere it is an *arc of a great circle*;¹⁵ in the case of the ellipsoid it is a surface curve that is ever so slightly S-shaped.

Figure 1.10 shows this general concept of “the shortest path within a surface” or *geodesic*.

geodeettinen
viiva

normaali-
leikkaus

The figure shows another curiosity: both in the plane and on the sphere, the plumb lines, or *normals*, to the surface at the end points *A* and *B* lie in the same plane, together with the connecting line or curve itself. This plane is called the *normal section*. In the case of an ellipsoid of revolution, this is however not the case: in the general case *there is no normal section containing both surface normals*. This effect is indeed extremely small and it can be ignored in all but the most precise calculations.

Traditional geodetic measurement networks on the Earth’s surface, especially triangulation networks, can be considered to consist of *mea-*

¹⁵Between any two points there are two great-circle arcs, one the shortest and the other the longest between the points. Only in the case of antipodes is there an infinity of great-circle arcs, all 180° long.

surement lines that are geodesics on some reference surface, usually a reference ellipsoid.

In practice, the geodetic instruments and signals are never precisely on the reference surface, but are some distance above or, more rarely, below it. Then, a *reduction* of the raw observations to this reference surface must be made. This applies to both angle and distance observations. Generally, the corrections needed are small.



1.5 The flattening of the Earth and gravity

As we explained above, the Hellenes were already well aware of the sphericity of the Earth and even her approximate size. During the 17th and 18th centuries, the idea developed of the flattening of the Earth, that is, the *ellipsoid of revolution* as descriptive of the figure of the Earth. Astronomers observed the flattening of Jupiter, describing it correctly as a dynamic phenomenon caused by the planet's fast rotation.

On a flattened Earth, of course also *gravity* changes with latitude. This was in fact observed with a pendulum clock, the pendulum of which had to be shortened when travelling to Cayenne in French Guyana (Jean Richer,¹⁶ 1672), so the clock would run on time again. Gravity is weaker near the equator than in France. Upon returning to France, Richer had again to lengthen his pendulum, so it would swing on time.

Newton and Huygens calculated theoretically a value for the Earth's flattening f , equation 1.2. A. C. Clairaut¹⁷ again derived his famous equation giving the relationship between flattening f and “gravity flattening”

$$\beta = \frac{\gamma_b - \gamma_a}{\gamma_a}.$$

Here, γ_a and γ_b are the accelerations of gravity on the equator and on the poles, respectively. Clairaut's theorem's approximate but elegant form is (Heiskanen and Moritz, 1967, equation 2-99)

$$f + \beta = \frac{5}{2} \frac{\omega^2 a}{\gamma_a},$$

in which ω is the rotation rate or angular velocity of the Earth. The expression $\omega^2 a / \gamma_a$ represents the centrifugal force on the equator caused by the Earth's rotation as a fraction of gravity there.

¹⁶Jean Richer (1630–1696) was a French astronomer.

¹⁷Alexis Clairaut (1713–1765) was a French mathematician and astronomer.





1.6 Reference surfaces and reference systems

As a reference surface for heights, the *geoid* is used, that level surface, or *equipotential surface*, of the gravity field of the Earth, which is on average on the same level as the mean sea level. This is thus the surface, the minimum state of potential energy, which sea water would reach if there were no currents in the ocean, no salinity or temperature differences, no variations in air pressure above it, and so on. In reality, all these disturbing factors exist and the global mean sea level deviates by over one metre from this equipotential surface, both above and below it. Of this deviation, part varies in time, like the tides, part is permanent, the *sea-surface topography*.

suolaisuus

meritopografia

The *levelled heights* of countries may be understood as heights above this geoid surface. In practice, the heights are tied to sea-level observations by tide gauges or *mareographs* operating on the coast, and the reference level is transferred inland by levelling, thus creating a *height system*.

More about the geoid and heights can be found in chapter 4.

A space geodetic measurement method, like [GPS](#), provides a way to determine the height of a point from the reference ellipsoid, because the ellipsoid of revolution is a simple mathematical surface in space. The reference ellipsoid is also otherwise a good approximation of the true figure of the Earth, which has traditionally been used as a reference surface for the adjustment of national or continental triangulation networks.

tasoitus

In very small areas and for many purposes, the “flat Earth approximation”, the assumption that the curvature of the Earth may be neglected, continues to serve us. The locations of points may be described by two plane co-ordinates and the height by one height co-ordinate, the vertical distance in metres from the reference surface. Depending on the application, a “small area” may be a building site, a city, or all of Finland — or, in special situations, taking special care, even an area the size of Europe, for example [Strang van Hees \(1990\)](#).



1.7 The sub-fields of geodesy

Geodesy is defined as “the science of measuring and mapping the Earth’s surface” (Helmert, 1880; see [Torge, 2001](#) page 1). This definition continues to hold today. Geodesy also includes the mapping of the sea floor, as well as the determination of the gravity field of the Earth, or *geopotential*.

merenpohja



More and more, the study of *changes* in the shape of the Earth and of the physical mechanisms causing these, *geodynamics*, has also become a part of the geodetic research field.

Thus, geodesy, in particular physical geodesy, belongs to the Earth sciences. However, geodesy also clearly belongs to the engineering sciences. In Finland, geodesy is currently (2018) being taught both at the University of Helsinki (two docents, external lecturers) and at Aalto University (one professor, docents, external lecturers).

According to Torge, geodesy may be divided into three sub-fields:

Global geodesy, also “measuring the Earth” (*geomensuration*, German *Erdmessung*). More precisely (Torge, 2001, page 2):

“The problem of geodesy is to determine the figure and external gravity field of the earth and of other celestial bodies as a function of time, from observations on and exterior to the surfaces of these bodies.”

Figure of the Earth:

- The *physical figure* of the Earth: the solid surface of the Earth, the interface between solid and gaseous or liquid matter — atmosphere, ocean — with all its mountains and depths.
- The *mathematical figure* of the Earth or *geoid*: the equipotential surface of the Earth’s gravity field that on average coincides with the mean sea surface, and that may be considered the continuation of the mean sea surface under the land masses. See section 1.3.

Geodetic surveying, surveying science. To this belongs the measurement of national — and today, international — geodetic and gravimetric base networks as a basis for mapping. In this work, the Earth’s curvature and gravity field must be taken into account.

runkoverkko

Ordinary surveying (“plane surveying”). To this belong, for example, topographic surveying and engineering surveying measurements. These measurements are of such areal extent and accuracy level, that in all calculations the curvature of the Earth may be neglected or taken into account by simple correction formulas.

maastomittaus

These measurements are made, besides by geodetic means, often also photogrammetrically by means of aerial photography. The objective of the work is always the production of geometrically precise and correct mapping material for use by society.



Computations may be done using plane co-ordinates x and y . Maps for small areas may be drawn without the use of a map projection, and the separate height co-ordinate H may be assumed to be purely metric (“metres above sea level”) without unpleasant consequences.



1.8 Topographic surveying: from terrain to map

The booklet “*Maastomittaus ja kartoitus*” (“*Topographic Surveying and Mapping*”) (Heiskanen and Härmälä, 1963) states:

maastomittaus

“Topographic surveying and the mapping that goes with it most often serve to provide of a larger or smaller area a depiction that is as correct and precise as possible, in the form of a *map*.”

hypoteekki

Society needs maps and location-based information for many purposes. In today’s society, the ownership of real estate, its buying and selling, and especially its use as collateral (mortgage) for loans aimed at maintaining and developing the property’s value, are foundational for a modern society with a high level of investment. To this end exists the *cadastral system*, which registers as reliably as possible the state of properties and the rights tied to them. There are millions of real-estate properties and parcels and their collective monetary value is astronomical.¹⁸

paikkatieto

18

Another use of maps and geospatial information that is important to society, is the *planning and construction of infrastructure*.¹⁹ roads, railways, bridges, tunnels, canals, airports, harbours, power plants, water works, telephone and data networks and so on. This is a vital public service from the viewpoint of economic productivity.

kaavoitus

kaava

In every developed society, construction is limited in some sense. You may not build what and how you wish, even on land that you own. *Zoning* regulates in a co-ordinated way the purposes for which land may and may not be used. These regulations are contained in *zoning plans*, the

¹⁸According to a report by RAKLI ry (2014) “All of Finland’s building stock together has a value, including parcels, of some 480 billion euros”. This means almost a hundred thousand euros for every Finnish man, woman and child.

¹⁹There are many definitions of infrastructure. The most perceptive one is undoubtedly that it is “those things the importance of which does not reach our awareness until they stop working”.



approval process of which is statutorily prescribed, public and consisting of several stages. The reason for this is that zoning affects the value of real-estate property, so the legal status of property owners requires that the democratic approval process of zoning plans contains adequate instruments of appeal. Maps and other surveying-based information sources are essential for this process.

In Finland, the planning of land use and the associated local infrastructure construction happens for the most part in the public administration, most often in municipalities. We speak of *spatial planning*.

yhdyskuntasuunnittelu

Topographic surveying is, in terms of volume, the overwhelmingly largest field of application of land-surveying.



1.8.1 Spatial planning

Spatial planning is a continuous activity, to which belongs zoning, planning of land use and the built environment, and managing building activity. In spatial planning a major role falls to the municipality. City planning and regional planning are forms of spatial planning.

The technical planning and building of a community require reliable information on the environment which is being planned and built. Figure 1.11 gives an idea of all the places where geospatial information plays a role in the continuous process of spatial planning and construction.

Topographic surveying is present in the whole process:

- Topographic information must be measured onto the map in a certain co-ordinate reference frame. maastotieto
- When the zoning plan is ready, it has to be set out on the terrain.
- The properties have to be measured and mapped.
- The technical structures must be placed onto the terrain.

Topographic surveying has the following *technical tasks*:

- Creating a *base network* for mapping: getting the measurements into a certain, known co-ordinate frame.
- Mapping terrain details while using base-network points: *detail survey*. kartoitusmittaus
- *Setting out*: the transfer of a plan onto the terrain, to be realised in its correct location. Setting or staking out is, in a way, *the inverse problem of mapping*. maastoonmerkintä





1.8.2 Carrying out the tasks of topographic surveying

Successfully carrying out the tasks of topographic surveying requires that the surveyor understands the following things:

- geodetic instruments
- measurement technique, measurement planning, measurement conditions
- geodetic computation: co-ordinates and transformations, derived quantities, accuracy
- zoning-plan calculations, setting out
- geographic information systems
- maps, printing, reporting.

Professor Matti Martikainen drafted an attractive tableau, here presented as tableau 1.1, on the role of the measurement plan and the place of topographic surveying in the whole of the measurement and mapping process. Our version is slightly modernised. The various parts of the tableau belong to the sub-fields of geodesy, photogrammetry, and cartography.



1.8.3 The end product of topographic surveying

The visible end product of topographic surveying is a *map*. A map must be, before all else, *correct*, but also clearly drawn, give all relevant information, and it is nice if it is attractive as well (Heiskanen and Härmälä, 1963).

A suitable scale is chosen for the map, which defines the level of accuracy of the information presented. The scale is chosen to suit the purpose of use of the map. The objects to be presented need to be *generalised* in a suitable way: too-small details should be taken out, however, essential details should be made clearer.²⁰

Many maps are digital. Then, the significance of the scale is not quite as clear.

Furthermore there is much location-related information in numerical form (*geospatial information*, consisting of location and attribute data)

sijainti- ja ominaisuustieto

²⁰For example, on a road map, the widths of the drawn roads have no relationship whatever with the widths of the roads in the terrain! The drawn width expresses the *importance* of the road to traffic. This is how generalisation works.



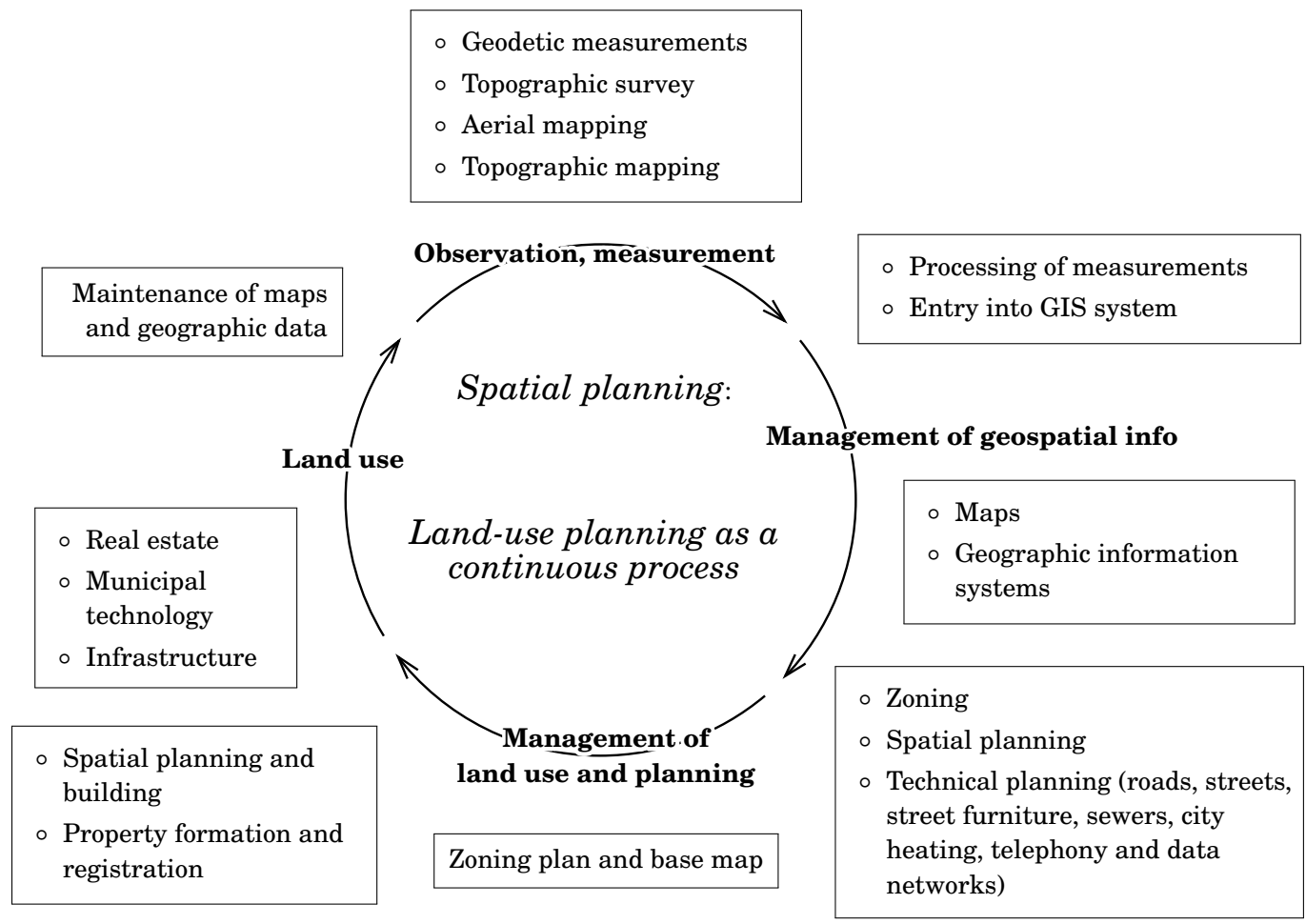


FIGURE 1.11. The roles of the map and topographic surveying in spatial planning and civil engineering.

as well as “metadata”: data describing other data, for example map information. The *legend* on a paper map is an example of metadata.

The following items of information are or can be part of the end product:

- korkeuskäyrä
 - Plane co-ordinates. These state the location inside a municipality, a country, the world. On the map, co-ordinate curves, a co-ordinate grid.
 - Height information, for example height contours, height values of points, possibly profiles.
 - The forms of the physical Earth’s surface, presented in various ways.
- ominaisuustieto
 - Attribute data. The measured objects are shown on the map in accordance with an agreed presentational style. A suitable identifier or symbol is given to every piece of information.

The information shown on the map may also be divided into natural and cultural data. See table 8.3.



Self-test questions

1. You have a car with an FM radio, a clock, a passport, a few weeks of free time, and money for food and lodgings and fuel. How do you establish to your own satisfaction that the Earth is round, not flat? Leah (2017).
2. How does Newton’s universal law of gravitation explain that the larger celestial bodies are approximately spherical? Why are they not precisely spherical?
3. Describe the hypotheses of Newton and Huygens on the interior mass distribution of the Earth and its effect on the Earth’s flattening.
4. Describe the idea of grade measurement as well as the objective of the grade-measurement project of the French Academy of Sciences.
5. What is, in equation 1.1, the unit in which the angle γ is expressed?
6. What is the “mathematical figure of the Earth”, and how is it related to the direction of the plumb line?
7. What is the geoid, and what is the reference ellipsoid?
8. Clairaut’s theorem gives the relationship between the rotation of the Earth, her flattening, and her “gravity flattening”. Intuitively,



TABLEAU 1.1. Topographic surveying as part of the whole measurement and mapping process.

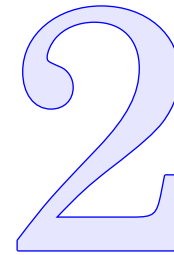
Measurement plan					
Base network measurement (GNSS), computation					
<i>Collection of topographic data</i>	Mapping surveying (often GNSS-RTK)	Existing maps	Digital imagery	Laser scanning	Photographs
<i>Processing of topographic data</i>	Geodetic computation	Digitisation	Image processing	Processing	Stereo photogrammetry
<i>Data integration</i>	Geographic information system (GIS)				
<i>Products</i>	<i>Graphical</i>	<i>Numerical</i>	<i>Textual</i>	<i>Metadata</i>	
<i>Data presentation to the end user</i>	Topographic maps	Point data bases	Reports	Map legend	
	Thematic maps	Elevation models	SMS alerts		
	Special maps	Digital CAD models	Navigator voice instructions		
	Customer printouts 3D visualisation				

give two reasons why gravity on the equator should be weaker than at the poles, and one reason why it should be stronger.

9. What is a geodesic?
10. What is the task of topographic surveying according to Heiskanen and Härmälä?
11. Name the three main fields of application of topographic surveying in society.
12. What are the three technical tasks of topographic surveying?
13. What is generalisation of map information? Give an example.
14. Think of reasons why society would want to disallow building a bicycle factory in the middle of the residential area of Eira, Helsinki.



Geodetic measurements and co-ordinates



METPΩ XPΩ

”Use measure”, Pittacus of Mytilene (640–568 BCE), appears in the seal of the [BIPM](#)



2.1 Units of measurement



2.1.1 Definitions

When we talk in geodesy, like more generally in physics, about measurement units to be used, we distinguish between *units* and *quantities*. For example, length is a *quantity*, the *unit* of which may be, for example, the metre [m].¹ So, for example:

1

The length of the distance AB is 15 metres, or 15 m.

Quantity	Value	Unit	Symbol
Length	15	Metre	m

In the literature, the term *dimension* is also used, for example the dimension of volume is length³, the dimension of acceleration is length × time⁻². This way is used to express how the definition of a certain quantity depends on the definitions of other quantities. For example, if one wants to precisely measure accelerations, one has to precisely measure both lengths (distances) and time intervals. This belongs to the field of *metrology*, the science of measurement.

¹The official symbol of a unit according to the [SI](#) system is always written in upright letters, *not in italics!* Italics is used for mathematical symbols. So $E = mc^2$, but $J = \text{kg m}^2/\text{s}^2$.



FIGURE 2.1. A public standard metre in Paris. [Wikimedia Commons, Standard metre in Paris.](#)

In Finland, as in most countries of the world, the **SI** system or International System of Units (**SI** = *Système International d'Unités*) is used.

The system consists of *base units* and *derived units*. There are seven base units, **BIPM, SI base units**, table 2.1.

2.1.2 Prefixes

One may add to the **SI** units (but not to the additional units,² see subsection 2.1.3!) a prefix indicating the order of magnitude according to table 2.2. The table is not complete.

TABLE 2.1. Measured quantities, units and their symbols. The list of derived quantities is incomplete.

	Quantity	Unit	Symbol	How derived
Base units	Length	Metre	m	
	Mass	Kilogram	kg	
	Time	Second	s	
	Electric current	Ampere	A	
	Temperature	Kelvin	K	
	Luminous intensity	Candela	cd	
	Amount of substance	Mole	mol	
Derived units	Plane angle	Radian	rad	
	Solid angle	Steradian	sr	
	Frequency	Hertz	Hz	s^{-1}
	Force	Newton	N	$kgms^{-2}$
	Pressure	Pascal	Pa	Nm^{-2}
	Energy	Joule	J	Nm
	Power	Watt	W	Js^{-1}
	Electric tension	Volt	V	WA^{-1}
	Electrical resistance	Ohm	Ω	VA^{-1}





TABLE 2.2. Prefixes indicating decimal order of magnitude in the SI system.

Value	Prefix	Symbol	Value	Prefix	Symbol
+1	deca	da	+6	mega	M
+2	hecto	h	+9	giga	G
+3	kilo	k	+12	tera	T
−1	deci	d	−6	micro	μ
−2	centi	c	−9	nano	n
−3	milli	m	−12	pico	p

Example: 1 MHz = 10^{+6} Hz.



2.1.3 Non-SI units accepted for use with SI

In daily life, and in many scientific disciplines, a generous number of non-SI units are in widespread use and are not going away soon. For example, a calendar, or a clock, would not be very practical in kilo- and megaseconds. The international metrological community, realising this, has created a category for these units, “Non-SI units accepted for use with the SI” ([Wikipedia, Non-SI units mentioned in the SI](#)).

Table 2.3 gives some often-used additional units “accepted for use with the SI”. lisäyksikkö



TABLE 2.3. Non-SI units accepted for use with the SI.

Quantity	Base unit	Accepted unit	Symbol
Time	s	Hour, minute, second	h m s
		Day	d
		Year	a
Plane angle	rad	Degree, minute, second	$^{\circ} / ' / ''$
		Gon	gon, ^g
Temperature	K	Degree Celsius	$^{\circ}\text{C}$
Volume	m^3	Litre	l
Mass	kg	Tonne	t

²This is nevertheless widely done, for example kcal means kilocalorie, a traditional unit of energy content of chemical substances and food. In computing again, the prefix k, or sometimes K, expresses the non-standard binary order of magnitude $1024\times$ ([Wikipedia, Kilobyte](#)). And the monetary unit k\$ is also used!



Celsius temperature is obtained from temperature in kelvin by subtracting 273.15 K from it:

$$T_1 = 0^\circ\text{C} = 273.15\text{K}, \quad T_2 = 0\text{K} = -273.15^\circ\text{C}.$$

Temperature *differences* are the same on the Celsius and Kelvin scales: $\Delta T = 1^\circ\text{C} = 1\text{K}$.



2.1.4 Length and time

Of the base units, both length and time are based on atomic phenomena.

The metre is the distance travelled by light in a vacuum in $1/299\,792\,458$ seconds. In practice, the metre is realised by an *iodine-stabilised helium-neon laser*, the wavelength of which is very precisely (2.5 parts in 10^{11}) known (Penzes, undated).

The second, again, is “the duration of 9 192 631 770 oscillation periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the ^{133}Cs atom” [official SI definition]: it is based on the use of a *caesium clock* (Hardis, 2018).



2.1.5 Units of angle

The *radian* and *steradian* units are dimensionless numbers (“bare numbers”) because they are ratios. For example, an angle in radians is the ratio between the length of a circular arc and its radius, and thus dimensionless. There exist however other units of angle, like the degree and the gon. Therefore it makes sense, in a way, to also treat all of these as units.

The situation is somewhat similar for logarithmic scales: Richter (total energy of earthquakes), the magnitude scale of stars, the decibel (dB) scale, the $^\circ\text{DIN}$ scale for the light sensitivity of photographic emulsions, and the pH, the degree of acidity of a solution.

Degrees In ordinary life we use as unit of angle the *degree*, symbol $^\circ$.

Geographical latitude and longitude as read from a map are also commonly given in degrees. A *right angle* is 90° , and a *straight angle* 180° . In addition to degrees, we have as traditional units the *minute* ($'$) and the *second* ($''$). These behave like their namesakes in time measurement: one degree is 60 minutes and one minute 60 seconds.

oikokulma



Calculation example Convert degrees, minutes, seconds to degrees and decimals:

$$\begin{aligned} 56^\circ 47' 33'' &= 56^\circ + \left(\frac{47}{60}\right)^\circ + \left(\frac{33}{60 \times 60}\right)^\circ = \\ &= 56^\circ + 0.783333\dots + 0.0091666\dots = 56.7924999\dots \end{aligned}$$

Convert back in the other direction (note the rounding error!):

$$\begin{aligned} 56.7925 &= 56^\circ + (60 \times 0.7925)' = \\ &= 56^\circ + 47.55 = \\ &= 56^\circ 47' + 0.55 = \\ &= 56^\circ 47' + (60 \times 0.55)'' = 56^\circ 47' 33''. \end{aligned}$$

Gon In geodesy and geodetic instrumentation, the *gon* is often used as a measurement unit.³ Sometimes the name “new degree” is used. A new minute, or centigon, is 0.01 gon, a new second, or decimilligon, 0.0001 gon. One notation is $1.2345 \text{ gon} = 1^{\text{g}} 23^{\text{c}} 45^{\text{cc}}$. In units of gon, a right angle is 100^{g} .

Radians A full circle contains 2π radians, 360 degrees (360°) and 400 gon (400^{g}). A right angle thus has $2\pi/4 = \pi/2$ radians.

Here are some useful equations to convert an angle given in radians to gon and degrees. See [Kahmen and Faig \(1988\)](#).

A full circle is

$$\begin{aligned} 2\pi \text{ rad} &= 400^{\text{g}} = 360^\circ, \\ 1 \text{ rad} &= \left(\frac{360}{2\pi}\right)^\circ \approx 57.2957795 = \left(\frac{400}{2\pi}\right)^{\text{g}} \approx 63.6619772, \\ \alpha \text{ rad} &= \left(\frac{360}{2\pi} \alpha\right)^\circ = \left(\frac{400}{2\pi} \alpha\right)^{\text{g}}. \end{aligned}$$

And

$$\alpha^\circ = \left(\frac{400}{360} \alpha\right)^{\text{g}} = 1.1111\dots \times \alpha^{\text{g}}, \quad \alpha^{\text{g}} = \left(\frac{360}{400} \alpha\right)^\circ = 0.9 \times \alpha^\circ.$$

2.2 Measurement error and uncertainty

No measurement is perfectly accurate. Land surveying, being a human activity, is also prone to error. Just like in computer programming, one

³Another name used is *grad*. The unit was taken into use by the French in connection with the Revolution and introduction of the metric system. Today, it is only used in the land-surveying field.

should not even try to measure completely errorlessly. A more realistic goal is, in addition to the usual carefulness, to develop methods by which

- Errors of a certain size can be noticed and removed from the observational material (statistical testing).
- The impact of not removed, or not even noticed, errors on the end result can be evaluated and minimised (adjustment calculus).

tasoituslasku

With these methods, measurement errors may be taken into account and as correct as possible measurement results be produced, the quality of which, their accuracy, is known, and the magnitude of any errors possibly still hiding out in them is also known or at least guessable.

tasoitus The processing of the measurements, the *adjustment*, will yield the following results:

- The “best” value for the unknown quantities, based on the measurements:
 - the most likely value
 - the statistical expectancy
 - a value with respect to which the remaining deviations are as little damaging as possible.

odotusarvo

4

- A judgement on the “goodness” of the original measurements, their *precision*.⁴ We speak of standard deviation or *mean error*. Nowadays the official term is *standard uncertainty*.
- Similarly, a judgement on the “goodness” or precision of the *end results*, the computed values (estimates) of the unknown quantities of interest.

karkea virhe

- A judgement on the possible occurrence of gross errors and their maximal magnitude after statistical testing has been done: *reliability*.
- A judgement on the possible presence of systematic errors and their magnitude.

⁵ *Measurement error* is the difference between the value to be measured⁵ and the measured value. The *measured value* is the result of an, often

⁴In English the terms “precision” and “accuracy” are used. *Precision* refers only to statistical spread, the variability of measurement values, whereas *accuracy* refers to the deviance of the observation values from the “correct values”. The latter concept thus also includes systematic errors.

⁵The “true value”, although one may philosophically ask if it even exists.



complex,⁶ measurement process.

6

2.2.1 Types of errors

We may divide the errors arising from this measurement and reduction process into the following categories:

Random errors represent the inherent, natural imprecision of the measurement process. Often it is assumed that these random errors are *normally distributed*, meaning the distribution of errors is a pretty Gaussian bell curve or normal distribution — on which more later on.

If this is so — and one should not assume this without further testing! — one has available the *least-squares adjustment method*, which minimises the joint impact of random errors in the end result.

pienimmän
neliösumman
tasoitus

Gross errors are caused by human mistakes or measurement device malfunctions. They happen only now and then, and they cannot be described by a statistical distribution. An example of a gross error is writing a digit wrong into an observation notebook.

karkea virhe

One strives to eliminate gross errors by statistical testing. Statistical testing is a broad scientific subject. As a rule of thumb, one may say that if a function of measurements differs from its known expectancy (for example zero) by more than three times its own, known standard deviation (mean error) σ , there is reason to suspect that one or more measurements are in error. This criterion is called the *three-sigma rule*.

Systematic errors are a sign that there are deficiencies in the theories used to describe the measurement process, observation geometry and physics of the observation situation. The occurrence of significant systematic errors should lead to a study and re-assessment of the theoretical assumptions. As an example, we know that the sum of the angles of a triangle must be 180° : the triangle condition. If the sum of the measured angles of a plane triangle differs significantly and consistently from this value, one must, for example, suspect

- lateral refraction in the atmosphere

⁶Sometimes the measurement process includes a pretty complex reduction chain or modelling effort.

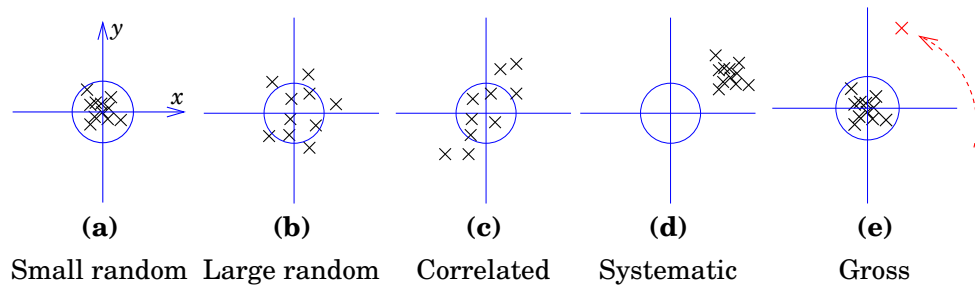


FIGURE 2.2. Examples of different error types. Random precise, random imprecise, correlated, systematic, gross.



- the triangle is so large, that on the curved surface of the Earth we no longer have a plane triangle and the theorem no longer applies.

Uncertainty, knowledge concerning the possible size of errors, can also be divided into different categories (GUM, 2008):

Type A uncertainty is standard uncertainty or standard error calculated for example from repeated measurements. This type of uncertainty can be handled by means of statistical theory: the concept corresponds to that of *random errors*.

Type B uncertainty is uncertainty that can *not* be determined by statistical means from the measurements, but needs to be assessed in some other way. For example, *systematic errors* can be determined by calibration, obtained from information by the device manufacturer, etc.

There is no uncertainty concept corresponding to gross errors. They should be avoided or eliminated using statistical testing.



2.3 Stochastic quantities



2.3.1 Discrete stochastic quantities

A measurement process is a *random* or *stochastic quantity*. A stochastic quantity is obtained by doing something the outcome of which is random. For example, casting a die creates a stochastic quantity \underline{n} , the realisations (“throws”) of which are n_1, n_2, n_3, \dots .

The possible values of a die cast are the integer numbers $\{1, 2, 3, 4, 5, 6\}$. It is the same with a coin flip, if we take as the values heads being a 0





TABLE 2.4. Die-cast statistics.

Number of throws	Value						Mean \pm standard deviation
	1	2	3	4	5	6	
60	14	8	9	12	9	8	10 ± 2.4
%	13.33	23.33	13.33	15.00	20.00	15.00	16.67 ± 4.08
600	103	91	106	114	87	99	100 ± 9.9
%	16.50	17.17	15.17	17.67	19.00	14.50	16.67 ± 1.65
6000	973	1007	1003	962	1015	1040	1000 ± 28.5
%	17.33	16.22	16.78	16.72	16.03	16.92	16.67 ± 0.47
60 000	10 138	9936	10 057	10 029	9925	9915	$10 000 \pm 89.5$
%	16.52	16.90	16.56	16.76	16.72	16.54	16.67 ± 0.15

and tails being a 1. We say that the value set or *codomain* is $\{0, 1\}$, a [arvojoukko](#) discrete value set.

If we perform a die cast again and again — meaning that we collect *realisations* n_i , $i = 1, 2, 3, \dots$ — we may always after a certain total number of casts tabulate the results. We obtain the example table 2.4, which expresses how many casts of the total number were ones, how many twos, etc.

According to experience, the greater the number of die casts, the smaller tends to be the deviation of the end result from the ideal outcome, the outcome in which the frequency of occurrence of every value in percents would be $16.666\dots\%$, or $\frac{1}{6}$. This empirical result is called the *law of large numbers*.



2.3.2 Expectancy

Based on this, we may assign to the outcomes of die casts a theoretical *probability value*, which expresses how often, “in the long run”, a certain value will happen. For a balanced (“fair”) die, the probability values are

$$p(1) = p(2) = \dots = p(6) = \frac{1}{6}.$$

In the discrete case, the *expectancy* — the value around which the casts [odotusarvo](#) group themselves, their “centre of mass” — is computed by the equation^{7 7}

$$E\{\underline{n}\} \stackrel{\text{def}}{=} \sum_{i=1}^N i \cdot p(i) \quad (2.1)$$



which yields $\frac{21}{6} = 3.5$. Here, N is the number of alternatives. The value 3.5 is not even a possible cast value!

In the case of a fair coin

$$p(0) = p(1) = 0.5,$$

and the expectancy is, by the same equation, 0.5. This value is also not a possible throw.



2.4 Statistical distributions



2.4.1 Real-valued quantities and density distributions

Geodetic measurements are generally stochastic quantities, the value set or *codomain* of which is a sub-set of the real numbers \mathbb{R} : a *continuous* set. For example, the result of a distance measurement is a distance in metres, and it is a real⁸ number. It is the same for angle measurements. The value set may be bounded, $\alpha \in [0, 360^\circ)$, but it is in any case continuous.

In the case of a continuous value set, we speak of a *probability density distribution*, or a *distribution* in short. If we make a large number of measurements of the same object, we may draw a *histogram*. Divide up the value set into “baskets”, and count and show the numbers of measurement results that fall into each basket.

As an example of a continuous statistical distribution may serve the places of the goals of a football match scored in the goal gate, figure 2.3.

The more measurements, the more bars we may draw, and the narrower we may make them. In the limit for an infinite number of measurements we obtain a continuous curve, the probability density distribution, which expresses the probability with which a measurement result will fall within some value interval $[x_1, x_2]$. See figure 2.4 for an example.

If on a certain interval $[x_1, x_2]$ the integral

$$\int_{x_1}^{x_2} p(x) dx$$

⁷More generally,

$$E\{\underline{n}\} \stackrel{\text{def}}{=} \sum_{i=1}^N v_i \cdot p(i),$$

in which v_i is the value alternative with serial number i . For a die, $v_i = i$.

⁸Actually a rational number $\in \mathbb{Q}$.



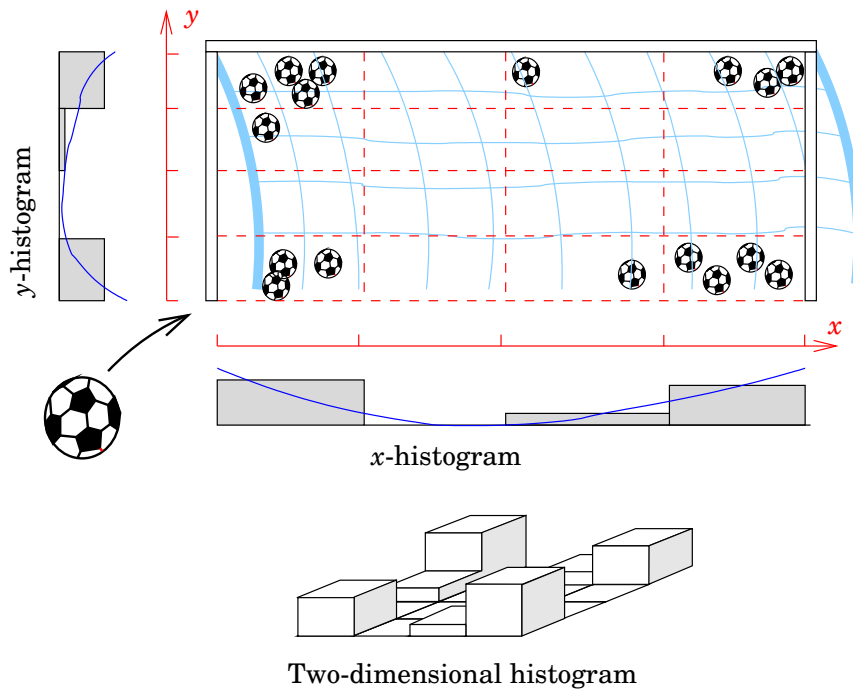


FIGURE 2.3. An example of a stochastic quantity on a continuous (two-dimensional) value set. Also histograms are drawn, both according to arguments x and y and according to both together. The blue curve is a possible probability density distribution function.

is zero, we say that it is *impossible* for a value x in this interval to occur as a realisation of the stochastic quantity. If the integral is 1, we say that the occurrence of some $x \in [x_1, x_2]$ is *certain*. In all other cases, the value of the integral gives the *probability* that the realisation of the stochastic quantity would hit inside this interval. The value of the probability is

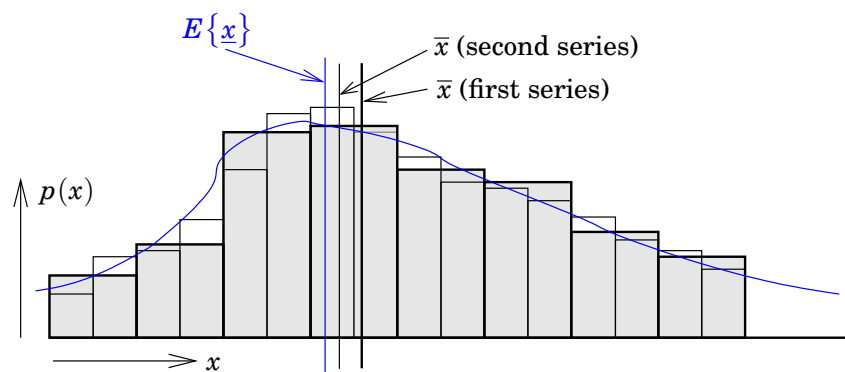


FIGURE 2.4. The probability density distribution $p(x)$ of a real-valued stochastic quantity as the limit of histograms. The expectancy $E\{x\}$ and the averages of the two measurement series, the second containing twice the number of measurements of the first, are also drawn.



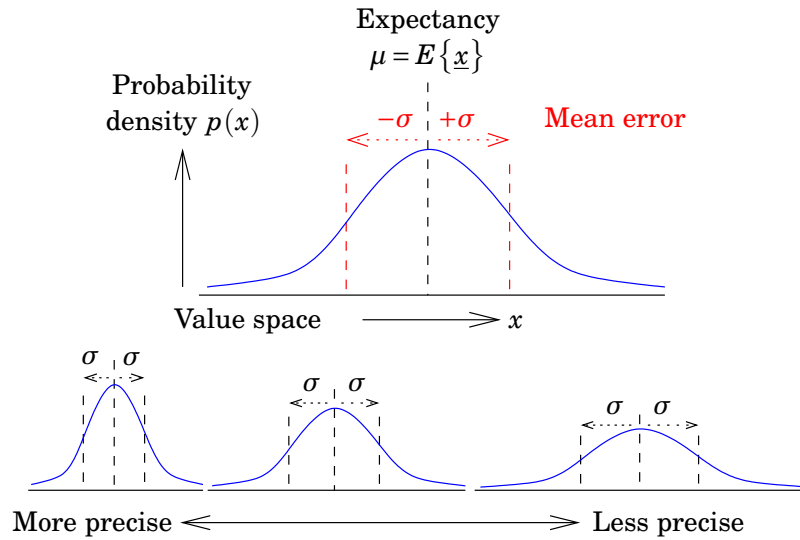


FIGURE 2.5. Important properties of the normal distribution.

⁹ thus always between zero and one.⁹

Always

$$\int_{-\infty}^{+\infty} p(x) dx = 1,$$

because the joint probability that the measurement value will be *any* real number, is 1, in other words, 100%: it is *certain*.

For a continuous stochastic quantity one can also compute an expectancy. The expectancy integral looks similar to the discrete counterpart 2.1:

$$E\{\underline{x}\} = \int_{-\infty}^{+\infty} x \cdot p(x) dx.$$

The expectancy is the x co-ordinate of the centre of mass of the area under the density distribution curve $p(x)$.



2.4.2 The Gaussian bell curve

In geodesy we use almost always the *normal* or *Gaussian distribution* (“bell curve”).

In figure 2.5 we see the *expectancy* $\mu = E\{\underline{x}\}$. The expectancy may be understood as the “true value” of the observed quantity \underline{x} , which is not itself measurable, but around which the measurement values group under the influence of their random errors. In the figure we also see the *mean error*¹⁰ σ , which describes the tendency of an individual measurement

⁹Mathematically, probability is a *measure*, like, for example, surface area or volume.

¹⁰... or standard deviation, or *standard uncertainty*.



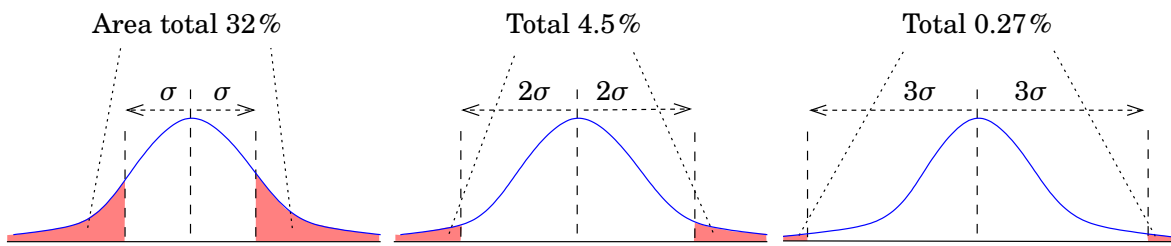


FIGURE 2.6. Probability values for the normal distribution.

to differ from the expectancy. In the case of the normal distribution, deviations equal in magnitude to the mean error $\mu \pm \sigma$ are located in the curve's *inflection points*.

taivutus piste

A concept by the name of *variance* is also used, the square of the mean error:

$$\text{Var}\{\underline{x}\} = \sigma^2,$$

the formal definition of which is¹¹

11

$$\text{Var}\{\underline{x}\} \stackrel{\text{def}}{=} \mathbf{E}\left\{(\underline{x} - \mathbf{E}\{\underline{x}\})^2\right\}. \quad (2.2)$$

Here $\mathbf{E}\{\cdot\}$ is the expectancy operator as defined above.

Like expectancy, variance is also the kind of “true value” that theoretically exists, but that we never actually really measure. When we have available a finite number of observations or *sample*, we may calculate the *sample mean* and *sample variance*, that, for the number of observations increasing, move ever closer and closer to the expectancy and variance. This phenomenon is also called the *law of large numbers*.

otos

The mathematical expression $p(x)$ for the normal distribution — which we do not present here — gives through integration the following probability values:

- the probability that x deviates from $\mathbf{E}\{\underline{x}\}$ more than an amount σ (in either direction): 32%
- the probability that x deviates from $\mathbf{E}\{\underline{x}\}$ more than an amount 2σ : 4.5%
- the probability that x deviates from $\mathbf{E}\{\underline{x}\}$ more than an amount 3σ : 0.27%.

¹¹So, the expectancy of the square of the deviation of a stochastic quantity from its own expectancy. This is a kind of cost function: if the cost of the “error” $\underline{x} - \mathbf{E}\{\underline{x}\}$ is proportional to its square, then $\text{Var}\{\underline{x}\}$ is the expected cost of the error.



In practice, a normally distributed stochastic quantity really never deviates from its expectancy by over three times its mean error.

This rule of thumb, the *three-sigma rule*, is made use of in statistical testing.

There exist a large number of different statistical distributions in addition to the normal distribution. These are contained in software packages used for statistical computation, such as MATLAB and R, and in the numerical libraries accompanying general programming languages.¹²



2.4.3 Covariance and correlation

When there are two stochastic quantities \underline{x} and \underline{y} , one can, besides their separate behaviours, also study how they behave *together*.

Let us give the name *covariance* to the expression

$$\text{Cov}\{\underline{x}, \underline{y}\} \stackrel{\text{def}}{=} E \left\{ (\underline{x} - E\{\underline{x}\}) (\underline{y} - E\{\underline{y}\}) \right\}. \quad (2.3)$$

This definition is analogous to the one for variance, equation 2.2, but describes the “behaving in the same way” property of the quantities \underline{x} and \underline{y} ; the similarity in their random behaviours.

Often it makes sense to *scale* this covariance relative to the variances of the quantities \underline{x} and \underline{y} in the following way:

$$\text{Corr}\{\underline{x}, \underline{y}\} \stackrel{\text{def}}{=} \frac{\text{Cov}\{\underline{x}, \underline{y}\}}{\sqrt{\text{Var}\{\underline{x}\} \text{Var}\{\underline{y}\}}}.$$

This is how *correlation* is defined between the quantities \underline{x} and \underline{y} . Correlation always lies in the interval $[-1, 1]$, or $[-100\%, 100\%]$.

The correlation (and covariance) between statistically independent quantities \underline{x} and \underline{y} equals 0. The non-vanishing of correlation may be a sign of a cause-and-effect relationship — but one cannot say that \underline{x} is the cause and \underline{y} the effect! \underline{y} may be the cause of \underline{x} , or \underline{x} and \underline{y} may have a common cause \underline{z} .

If the correlation is 1, or 100%, we speak of *perfect correlation*. In this case there exists an exact functional relationship between \underline{x} and \underline{y} . If the realisation value of one quantity is given, the corresponding realisation value of the other can be calculated precisely.

¹²A good overview is [Easton and McColl](#).



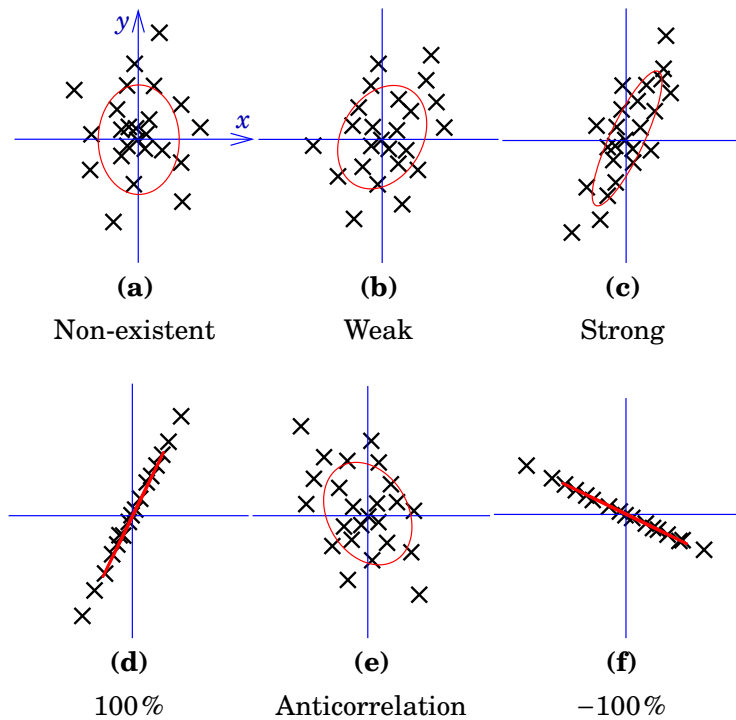


FIGURE 2.7. Some examples of correlation.

If the correlation is negative, we speak of *anticorrelation*. If the correlation is -1 , or -100% , there exists a perfect correlation between the quantities \underline{x} and $-\underline{y}$, and we speak of *perfect anticorrelation*.

Figure 2.7 shows some examples of correlation.

2.4.4 Propagation law of errors

An important property of mean errors is the *propagation law of errors*.

virheiden
kasautumislaki

If a stochastic quantity \underline{z} is a linear combination of two other stochastic quantities \underline{x} and \underline{y} ,

$$\underline{z} = a\underline{x} + b\underline{y},$$

we may also write

$$E\{\underline{z}\} = aE\{\underline{x}\} + bE\{\underline{y}\},$$

and thus

$$(\underline{z} - E\{\underline{z}\}) = a(\underline{x} - E\{\underline{x}\}) + b(\underline{y} - E\{\underline{y}\}).$$

This is probably intuitively clear. Now according to the definition of variance, equation 2.2, it follows that

$$\text{Var}\{\underline{z}\} = E\left\{(\underline{z} - E\{\underline{z}\})^2\right\} =$$



$$= a^2 E \left\{ (\underline{x} - E\{\underline{x}\})^2 \right\} + b^2 E \left\{ (\underline{y} - E\{\underline{y}\})^2 \right\} + 2ab E \left\{ (\underline{x} - E\{\underline{x}\})(\underline{y} - E\{\underline{y}\}) \right\}.$$

The definitions 2.2 and 2.3 now yield

$$\text{Var}\{\underline{z}\} = a^2 \text{Var}\{\underline{x}\} + b^2 \text{Var}\{\underline{y}\} + 2ab \text{Cov}\{\underline{x}, \underline{y}\}. \quad (2.4)$$

varianssien
kasautumislaki

This equation is called the *law of propagation of variances*.

In the case where $\text{Cov}\{\underline{x}, \underline{y}\} = 0$ — meaning that the quantities \underline{x} and \underline{y} are *uncorrelated* — we obtain

$$\text{Var}\{\underline{z}\} = a^2 \text{Var}\{\underline{x}\} + b^2 \text{Var}\{\underline{y}\},$$

or, in a more compact notation,

$$\sigma_z^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2.$$

This can be generalised to more than two quantities. For example, if

$$\underline{w} = a\underline{x} + b\underline{y} + c\underline{z},$$

we obtain

$$\sigma_w^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + c^2 \sigma_z^2.$$

A *special case* is the situation in which z is the sum or difference of x and y , so $a = \pm 1$, $b = \pm 1$, and x and y are statistically independent of each other. Then we get a simple, Pythagoras-like, much-used equation:

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2.$$

Often, this special case is referred to as the law of error propagation.

Example We have measured the distance between points A and B in three parts,

$$\underline{s}_{AB} = \underline{s}_1 + \underline{s}_2 + \underline{s}_3$$

(so $a = b = c = 1$), in which

$$\underline{s}_1 = 10 \text{ m} \pm 0.05 \text{ m}, \quad \underline{s}_2 = 20 \text{ m} \pm 0.2 \text{ m}, \quad \underline{s}_3 = 15 \text{ m} \pm 0.1 \text{ m}.$$

We observe that

$$\sigma_1 = 0.05 \text{ m}, \quad \sigma_2 = 0.2 \text{ m}, \quad \sigma_3 = 0.1 \text{ m},$$



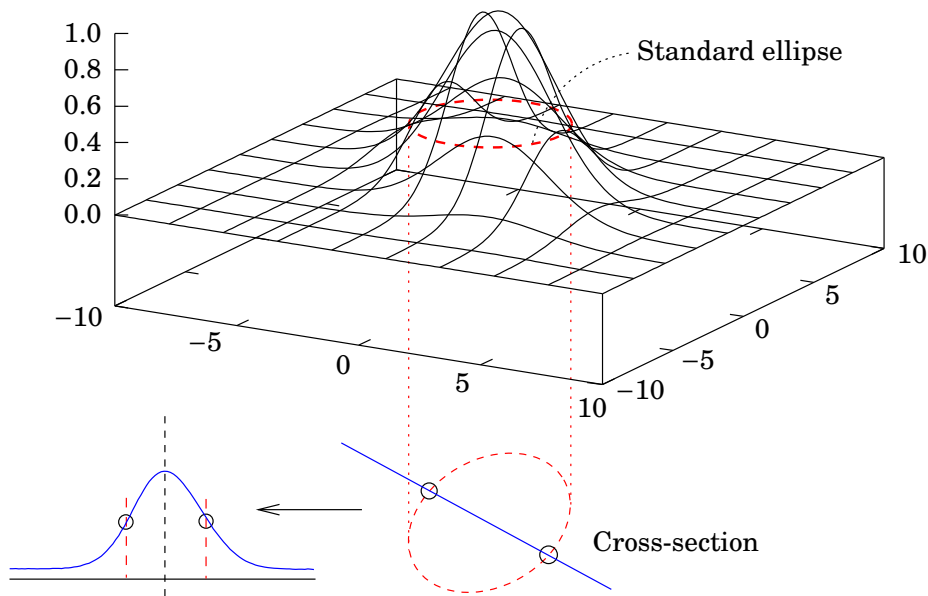


FIGURE 2.8. A two-dimensional probability density distribution.

and

$$\begin{aligned} \sigma_{AB}^2 &= \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = (0.05\text{ m})^2 + (0.2\text{ m})^2 + (0.1\text{ m})^2 = \\ &= 0.0525\text{ m}^2, \end{aligned}$$

so $\sigma_{AB} = \sqrt{\sigma_{AB}^2} \approx 0.23\text{ m}$, and the result is

$$\underline{s}_{AB} = 45\text{ m} \pm 0.23\text{ m}.$$

(The three measurements are assumed to be statistically independent of each other.)

2.4.5 Multi-dimensional distributions

Often, like in the above football example (figure 2.3), we speak of stochastic quantities consisting of several components. An example is given by the co-ordinates (x, y) of a point in a plane. In the same way as described above, we may draw two-dimensional histograms and speak about the probability density distribution $p(x, y)$.

Figure 2.8 shows a two-dimensional normal density distribution, and its *standard ellipse* or *error ellipse*. It is the two-dimensional counterpart to the mean error. A vertical cross-section again always produces a one-dimensional Gaussian bell curve, the mean-error points $\mu \pm \sigma$ of which are precisely the intersections of the cutting plane and the error ellipse.





TABLEAU 2.5. On correlation.

Correlation is a curious thing. It is always good to remember that statistical dependence does *not* necessarily mean that there is a direct cause-and-effect relationship, only that the observations have something in common. The connection may be indirect, like in the famous case in which the ice-cream sales over summer correlated with drowning statistics. The connection may also be uninteresting, like the correlation between two phenomena both trending in time, for example, between the Ethiopian gross domestic product and Danish wind-power output.

Whereas the mean-error points of the one-dimensional distribution are at $x = \mu \pm \sigma$, the equation for the error ellipse is more complicated. The expectancy (“true value”) of the stochastic co-ordinate pair $(\underline{x}, \underline{y})$ is itself a co-ordinate pair, (μ_x, μ_y) , the centre of the error ellipse.

In the general case where the axes of the error ellipse are not parallel to the co-ordinate axes (and of different lengths!), the pair of stochastic quantities \underline{x} , \underline{y} can no longer be treated as independent variables: they *correlate* with each other. This means that the most probable value of x will depend on the actual value of y , and vice versa. In adjustment calculus, this statistical dependence must be taken into account in order to arrive at the best possible — *optimal* — estimates both of μ_x and of μ_y .

In three-dimensional space there exist probability density distributions of three arguments $p(x, y, z)$, the visual representations of which are triaxial (and three-dimensional) *error ellipsoids*.



2.4.6 The variance matrix

The variances of multi-dimensional stochastic quantities themselves also become multi-dimensional: they become *variance matrices*. Let¹³

$$\underline{x} \stackrel{\text{def}}{=} \begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix},$$

then the variance (or variance-covariance) matrix is

$$\text{Var}\{\underline{x}\} \stackrel{\text{def}}{=} \begin{bmatrix} \text{Var}\{\underline{x}\} & \text{Cov}\{\underline{x}, \underline{y}\} \\ \text{Cov}\{\underline{x}, \underline{y}\} & \text{Var}\{\underline{y}\} \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} = \Sigma_{xx},$$

¹³... and of course also the expectancy is a vector: $E\{\underline{x}\} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$!



in which again apply the definitions

$$\begin{aligned} \text{Var}\{\underline{x}\} &\stackrel{\text{def}}{=} E\left\{\left(\underline{x} - E\{\underline{x}\}\right)^2\right\}, \\ \text{Var}\{\underline{y}\} &\stackrel{\text{def}}{=} E\left\{\left(\underline{y} - E\{\underline{y}\}\right)^2\right\}, \\ \text{Cov}\{\underline{x}, \underline{y}\} &\stackrel{\text{def}}{=} E\left\{\left(\underline{x} - E\{\underline{x}\}\right)\left(\underline{y} - E\{\underline{y}\}\right)\right\}, \end{aligned}$$

and in which we see the often-used notation¹⁴

14

$$\text{Var}\{\underline{x}\} = \sigma_x^2, \quad \text{Var}\{\underline{y}\} = \sigma_y^2, \quad \text{Cov}\{\underline{x}, \underline{y}\} = \sigma_{xy}.$$

In appendix A a short explanation is given of the basics of matrix calculus.

In fact, the *error ellipse* is a graphical representation of this 2×2 sized variance matrix. If $\sigma_{xy} = 0$, we say that \underline{x} and \underline{y} do not correlate with each other, or, carelessly,¹⁵ that they are statistically independent of each other.

15

In addition, the propagation law of variances, equation 2.4,

$$\text{Var}\{\underline{z}\} = a^2 \text{Var}\{\underline{x}\} + b^2 \text{Var}\{\underline{y}\} + 2ab \text{Cov}\{\underline{x}, \underline{y}\},$$

may be written into a more general form: if we form the row vector

$$\underline{a} \stackrel{\text{def}}{=} \begin{bmatrix} a & b \end{bmatrix},$$

and the corresponding column vector¹⁶

16

$$\underline{a}^\top = \begin{bmatrix} a & b \end{bmatrix}^\top = \begin{bmatrix} a \\ b \end{bmatrix},$$

we obtain

$$\text{Var}\{\underline{z}\} = \begin{bmatrix} a & b \end{bmatrix} \cdot \begin{bmatrix} \text{Var}\{\underline{x}\} & \text{Cov}\{\underline{x}, \underline{y}\} \\ \text{Cov}\{\underline{x}, \underline{y}\} & \text{Var}\{\underline{y}\} \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \underline{a} \cdot \text{Var}\{\underline{x}\} \cdot \underline{a}^\top.$$

Understanding this requires knowing how to multiply matrices (row \times column), see appendix A. Writing it out gives us back the original equation 2.4.

This is presented in the literature symbolically in somewhat varying notations:

$$\sigma_z^2 = \underline{a} \text{Var}\{\underline{x}\} \underline{a}^\top = \underline{a} \Sigma_{xx} \underline{a}^\top.$$

We will discuss the propagation of variances in linear models of many variables more generally in section 14.7.

varianssien
kasautuminen

¹⁴Sometimes we see in older texts $\text{Var}\{\underline{x}\} = m_x^2$, $\text{Var}\{\underline{y}\} = m_y^2$, $\text{Cov}\{\underline{x}, \underline{y}\} = m_{xy}$, as back then, typewriting Greek letters was difficult.

¹⁵... because there may be a more complicated statistical dependence that does not





2.5 Geodetic observables



2.5.1 Angles, directions

kolmiomittaus-
verkko

Traditionally, the large national or continental geodetic networks have always been *triangulation networks*. *Directions* are measured between the network stations, and the measurements in the network are adjusted: from the measured values, small contradictions caused by the uncertainty of measurement are computationally removed, leading to the solution of the triangulation. Back in time, distance measurement was difficult over larger distances, as the only working technique was measurement using mechanical means, such as measuring tapes and rods.

astemittaus

perusviiva

In triangulation, the procedure is that in a network, all feasible directions are measured, plus *one single distance*. From these measured values, all the other distances and point locations are computed. This is how Snellius carried out his famous grade measurement between Bergen op Zoom and Alkmaar, cities 130 km apart, by measuring directly only the baseline 87 *roeden* long, built by him in a meadow! The Lapland grade measurement was similar: the baseline was built during winter 1736 on the ice of the Torne river, and all other measurements were directions between the points of the triangulation network.

kolmiomittaus

17

mittapöytä

18

The basic idea of triangulation can be explained using the *plane table*.¹⁷ At point *A* on the terrain, a plane table is erected, a sheet of transparent paper placed on it, and all directions to the terrain points *B*, *C*, *D* and *E*¹⁸ are drawn. Move to point *B*, and draw a similar set of directions to targets *A*, *C*, *D* and *E*. In the office, the sheets of paper are placed on top of each other. The result is a miniature image of the true geometry of the terrain — at a scale of $m = ab : AB$. So, if the distance *AB* in the terrain is 3 km, and on the superposed papers $ab = 30$ cm, then the scale of the “map” thus obtained is 1 : 10 000.

Nowadays, directions are measured using a *theodolite*, a precise instrument for measuring horizontal and vertical angles, and creating the map takes place computationally.

show as correlation.

¹⁶We often use the transpose notation to make it easier to write a column vector in running text.

¹⁷German *Messtisch*, French *planchette*, Dutch *meettafel*, Swedish *lantmätartavla*, Danish, Norwegian *målebord*.

¹⁸The drawing is made easier by using an *alidade*, German *kippregel*.



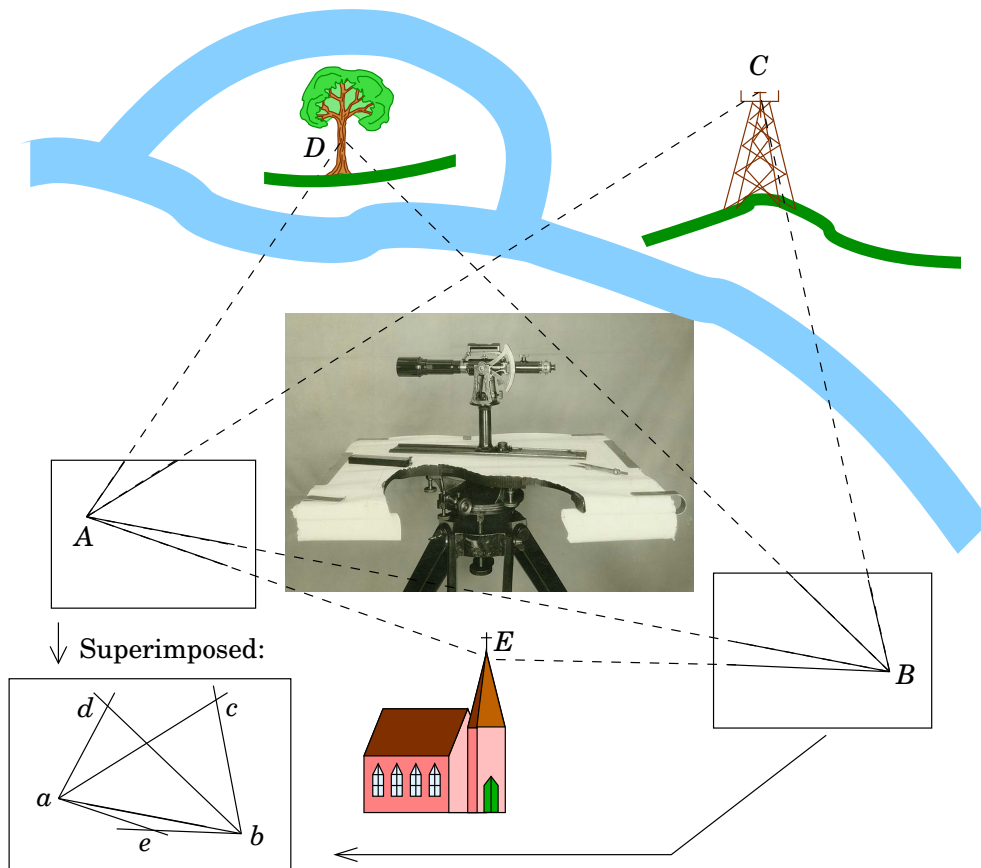


FIGURE 2.9. Triangulation by means of a plane table and alidade. Centre image [Wikimedia Commons, Plane table with alidade.](#)

In photogrammetry there is a three-dimensional analogue to the plane table method, *stereo model restitution*, which however nowadays is also invariably realised fully digitally. Figure 2.10 shows how from two aerial images a stereo model is formed using a viewing device. Similarly, in a stereo restitution instrument, a model is formed, inside which a *floating mark* may be moved around three-dimensionally. The co-ordinates of the mark are continuously output to a computer hooked up to the instrument, and a map can be drafted immediately.

avaruus-
mittamerkki

Aerotriangulation is also based on forming stereo models.

ilmakolmiointi

A fairly modern geodetic technique based on the idea of the plane table but in three dimensions, is Yrjö Väisälä's¹⁹ *stellar triangulation*. This technique, in which the stellar background is used for measuring directions, is discussed in section 17.3.

19

tähti-
kolmiomittaus

¹⁹Yrjö Väisälä (1891–1971), “the Wizard of Tuorla”, was a Finnish astronomer, physicist, geodesist, metrologist, builder of telescopes, finder of comets and asteroids, recreational sailor and Esperanto practitioner.



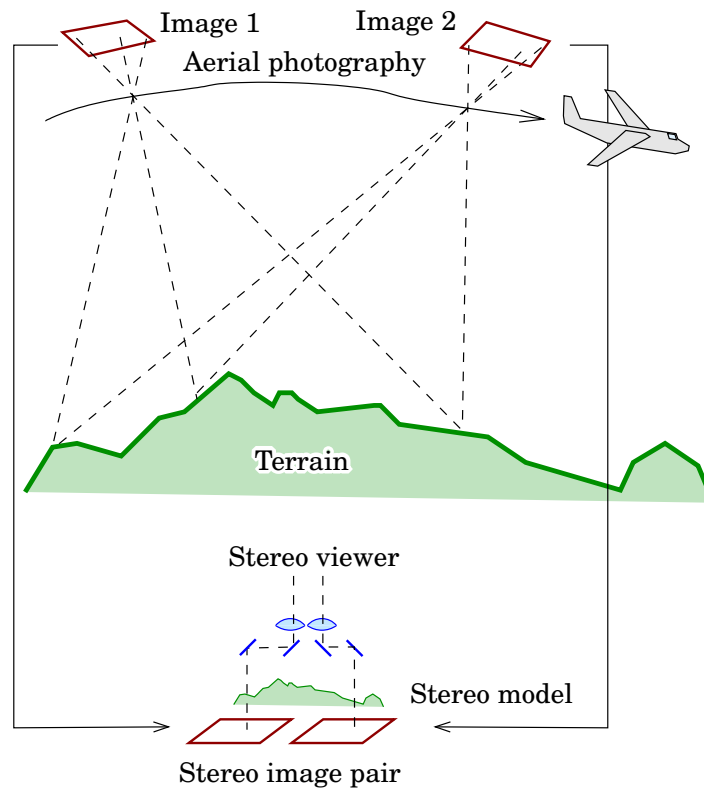


FIGURE 2.10. Forming a stereo model in photogrammetry.



2.5.2 Distances, distance differences

In the past, distances could only be measured mechanically, using measuring wires, tapes, or rods. The precision achieved in the measurements was often impressive thanks to careful procedures.

In local surveying, steel measuring tapes continue to be used, as they are inexpensive and easy to use and carry along. One should only remember to clean and grease them after use. The lengths are 20–60 m. The precision measuring wires of old were made of invar, a steel alloy with a very low thermal expansion coefficient.

Today, distance measurements are done electronically or electro-optically. Only a straight line of sight between points is needed. The devices may use microwaves (Tellurometer) or more often visible light (Kern Mekometer), laser light (Geodimeter),²⁰ or, in the case of most modern range-finders (distance measurement devices) and total stations, infrared light produced by a light-emitting diode (LED). The effect of atmospheric delays on signal propagation must always be carefully taken

²⁰

etäisyysmittari
takymetri
hohtodiodi

²⁰Geodimeter™ is a Swedish invention and the name is a trademark of Trimble AB, Danderyd, Sweden.



into account in processing the observations.

Terrestrial geodetic measurement instruments are able to measure both horizontal and vertical angles as well as slant ranges. They are called *electronic tacheometers*,²¹ and they thus combine the properties of a theodolite and a range-finder into one integrated, fully electronic, highly automated instrument. 21

Satellite geodesy also uses electronic distance measurement. The GNSS, Global Navigation Satellite Systems — like the Global Positioning System GPS — in broad use today are based on distance measurements made by microwaves — more precisely, distance *difference* measurements.²² 22 Satellite lasers on the other hand (like the instrument at the Metsähovi research station) measure the travelling time from the observation station to a satellite reflecting the light, and back.

The advantage of using electronics is, that the incredibly high precision of frequency measurement is thus harnessed as distance measurement precision. When the precision of frequency measurement can well be $1 : 10^{12}$, it is understandable that with GPS, we measure intercontinental distances with a relative precision of even $1 : 10^9$. Satellite techniques are even more precise than terrestrial ones, because most of the signal propagation takes place outside the denser parts of the atmosphere.

2.5.3 Potential differences, levelling

Measurement of geopotential differences is traditionally done by *levelling*: one measures the height difference between two points in metres, and converts this difference into a potential difference using local gravity g . If the height difference is ΔH and the difference in geopotential is ΔC , we use the equation²³ 23

$$\Delta C = g \Delta H.$$

The measurements are repeated along a line, and the differences ΔC_i obtained are summed up. From the lines, a *levelling network* is built, in which the potential differences calculated around closed loops must sum to zero. The network is adjusted using this condition.

²¹Tacheometer, Greek *rapid measurer*. Called more often *total stations*.

²²*Hyperbolic positioning systems*, like Decca, Transit/Doppler, and also GPS and the other satellite positioning systems, are based on distance difference measurement.

²³This is an example of “work is force times path”. Work (per unit of mass) is ΔC , path is ΔH , and force (per unit of mass: according to Newton’s law $F = ma$, *acceleration*) is g .

In subsection 4.3.1 a few more exotic height determination alternatives to traditional levelling are presented, many of which are based on the direct physical comparison of geopotential values, for example using a fluid surface.



2.6 About co-ordinates

In geodesy we use *co-ordinates* in order to describe the figure and size of the Earth and to determine the locations of points on or near the Earth's surface.

The co-ordinate reference systems used in geodesy are generally *three-dimensional*, as the Earth is a three-dimensional object in a three-dimensional space. As an example we may mention latitude, longitude, and height (φ, λ, h) , which let us present the location of a point in an intuitive way.

“Two-dimensional” co-ordinate reference systems are really *map projection co-ordinates*, derived quantities not of direct interest to geodesy. They belong more to the field of cartography, although they are used very broadly in applied surveying. For example, on older Finnish topographic maps one encounters **KKJ** co-ordinates, which are map-plane co-ordinates, the kind of co-ordinates (x, y) used on a map sheet, directly measurable by ruler.

In addition to spatial co-ordinates, *time*, for describing processes of change, and in physical geodesy, as a kind of co-ordinate, *geopotential*, the potential of the gravity field of the Earth, are also used. See subsection 4.1.1.



2.7 Why we use plane co-ordinates

Because the Earth is a three-dimensional object, geodesy is a three-dimensional science. The Earth and points associated with her are located in three-dimensional space, and the task of geodesy is to determine and present these locations using three-dimensional co-ordinates, such as (X, Y, Z) . Modern measurement systems, like the global positioning system **GPS**, are able to directly measure three-dimensional co-ordinates.

Geodesy, however, is also an applied branch of science and engineering serving people. Humankind lives, forced by gravity, in a quasi-two-dimensional “subspace”, in which there is freedom of movement almost



only in the horizontal direction, along the surface of the Earth. Furthermore, the important medium of communication called *paper* is unconditionally two-dimensional, and maps are commonly drawn on paper!

For this reason, *plane co-ordinates* are very generally used in geodesy and surveying: rectangular, two-dimensional co-ordinates in the horizontal plane.

There exist a multitude of practical plane co-ordinate reference systems suitable for surveying work. The main differences between them are:

- The location of the *origin* and the *orientation* of the axes. The origin, the starting point where $x = y = 0$, must be known. Generally the axes are x to the north and y to the east, but not always, and not necessarily accurately.
- The *technology of determination*, in other words, geocentricity: modern plane co-ordinates are obtained from geocentric, three-dimensional co-ordinates produced using [GNSS](#).

Geodetic plane co-ordinates are, in fact, *map projection co-ordinates*: they are calculated, using map projection formulas, from originally three-dimensional co-ordinates through geodetic latitude and longitude (φ, λ) .

In a very small area, like a building site, no proper map projection is needed. In those kinds of areas, plane co-ordinates may be understood as a special, rectangular case of *topocentric* (observation-site centric) *co-ordinates*.



2.8 Co-ordinates of location in three dimensions

Three-dimensional co-ordinates can be of different types, for example *rectangular* or *geodetic*, also called geographical. The most common co-ordinate reference system used is the *three-dimensional rectangular* or *Cartesian* (X, Y, Z) system. Often it is also geocentric.

Geocentricity means literally, that the origin coincides (to a certain accuracy) with the centre of mass of the Earth. In addition, the Z axis of a geocentric system points along the Earth's axis of rotation, to the celestial North Pole.

In a geocentric co-ordinate reference system, the direction of the X axis is in principle arbitrary and thus conventionally agreed. For a geocentric system on the Earth's surface, the *meridian which contains the direction of the plumb line at the Royal Observatory Greenwich* serves as an international standard²⁴ for fixing the direction of the X axis.

luotiviiva
24





FIGURE 2.11. The Greenwich meridian for tourists. [Wikimedia Commons, The prime meridian \(Greenwich\)](#).

The X axis lies both in the plane of the Greenwich meridian and in the equatorial plane, and is thus perpendicular to the Z axis. The Y axis in turn is perpendicular to both the Z and X axes, so that all three axes are mutually perpendicular. See figure 2.12.

Three-dimensional rectangular co-ordinates X , Y and Z may be generally applicable, but they are not particularly intuitive. For example, the co-ordinates of the GPS antenna at the Metsähovi research station are, expressed in the Finnish EUREF-FIN reference frame (see subsection 3.2.3):

$$X = 2892571.1204\text{m}, \quad Y = 1311843.2621\text{m}, \quad Z = 5512633.9521\text{m}.$$

The numbers are interesting looking, but do not give a very enlightening or easy to grasp answer to the question “where is Metsähovi?” . . .

As a first step to more practical co-ordinates, we construct *geodetic co-ordinates*. First we construct mathematically a reference ellipsoid,

²⁴The treaty of Washington DC of 1884 made the Greenwich meridian the world’s zero or reference meridian. At the same time, a “world time” or universal time was agreed: Greenwich Mean Time, GMT. The basic idea was that civil times of countries would differ from GMT by a whole number of hours, for Finland +2h in winter (EET) and +3h in summer (EEST). Without this time-zone system, international traffic (by sea, air, or telephone) would be cumbersome.



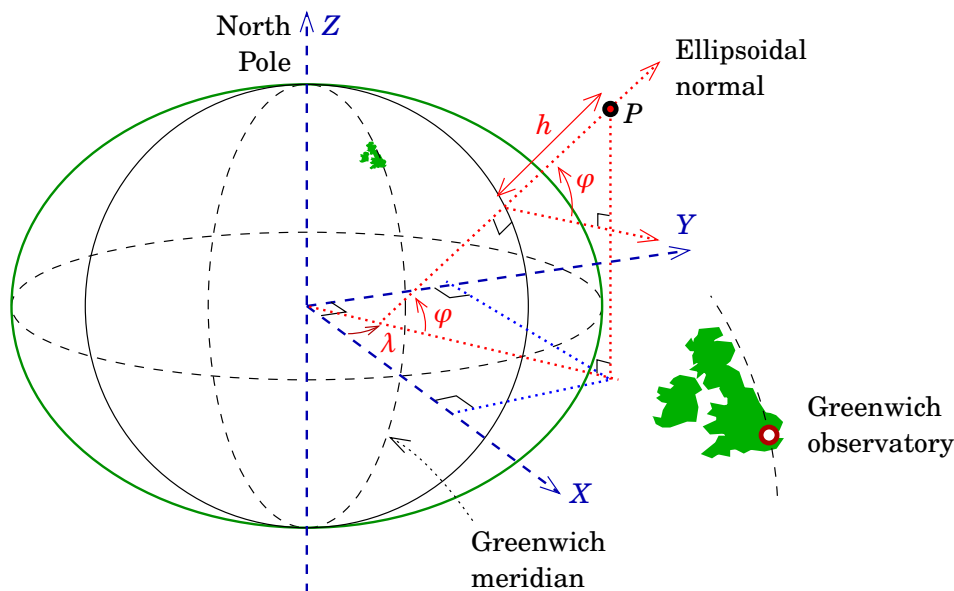


FIGURE 2.12. Rectangular and geodetic co-ordinates.

a suitably flattened ellipsoid of revolution the measures of which are reasonably close to those of the real Earth.²⁵

A point is projected along the *ellipsoidal normal* onto the surface; the distance of projection, h , is the *ellipsoidal height*, the *direction angles* of the projection line or *ellipsoidal normal*²⁶ are φ , the *geodetic latitude* reckoned from the equator, and λ , the *geodetic longitude*, reckoned from the Greenwich meridian. The triplet (φ, λ, h) — or often only the pair (φ, λ) — is referred to as *geodetic co-ordinates*. The geodetic co-ordinates of, for example, Metsähovi are

$$\varphi = 60^\circ 13' 2''.89046, \quad \lambda = 24^\circ 23' 43''.13336, \quad h = 94.568 \text{ m}, \quad (2.5)$$

... and this tells people already a whole lot more about where this point is located!

There is a simple mathematical relationship between rectangular and geodetic co-ordinates: each can be *converted* into the other without losing precision. They are equivalent²⁷ presentations of the location of a point:

²⁵For example, the GRS80 reference ellipsoid: equatorial radius 6378137.0m, polar radius 6356752.3141m (some 21 km shorter), and flattening 1 : 298.257222101.

²⁶The ellipsoidal normal generally does *not* go through the centre of the ellipsoid! See figure 2.12.

²⁷So, if (X, Y, Z) is given, then (φ, λ, h) can be calculated, and the reverse is also true.





FIGURE 2.13. Geodetic co-ordinates.

$$(X, Y, Z) \begin{array}{c} \xleftarrow{\text{equation}} \\ \xrightarrow{\text{equation}^{-1}} \end{array} (\varphi, \lambda, h)$$

... and both are geocentric and three-dimensional. The differences between them are:

- Rectangular co-ordinates are more convenient to handle in numerical work.
- Geodetic co-ordinates are more intuitive, closer to the human understanding.

In addition to geocentric co-ordinates, *topocentric* co-ordinates often are encountered; co-ordinates reckoned relative to some local origin. Topocentric co-ordinates are naturally produced by measurements in which the location of the measuring instrument becomes the origin of the co-ordinate system: one measures a direction — *azimuth* or horizontal direction angle, and *elevation* angle above the horizon — and a *distance* from the instrument's location to the points to be measured.

korkeuskulma



Self-test questions

1. What is a measurement unit, a measured quantity, and its dimension? Describe and explain the differences.
2. “Three score years and ten”. How much is this in gigaseconds?



3. Scour the Internet for examples of confusion around physical units causing damage, loss or accidents.
4. The mean temperature of the Earth for 2016 was an estimated 14.8°C . How much is this in kelvin?
5. Explain random errors, gross errors, and systematic errors.
6. Consider a pack of 52 playing cards, with the numbers 2–10, a jack counted as 11, a queen as 12, a king as 13, and an ace as 1. What is the expectancy for a card drawn at random from the pack?
7. Look up the equation for the standard normal distribution's *probability density* $p(x)$. What is the function value $p(x)$ for $x = \mu \pm \sigma$? And for $x = \mu$? And what is the ratio $p(\mu \pm \sigma)/p(\mu)$?
The *standard normal distribution* is the normal distribution for the parameter values $\mu = 0$ and $\sigma = 1$.
8. Would you expect that the incidence of bush fires in Australia and the thickness of snow cover in Lapland would be correlated? Anticorrelated? Uncorrelated? Why?
9. Why do we use also two-dimensional (plane) co-ordinates in geodesy?
10. How is geocentricity defined?
11. What is the advantage of using geodetic co-ordinates (φ, λ, h) over rectangular co-ordinates (X, Y, Z) ?



Exercise 2–1: Co-ordinates and a street address

A riddle for seafarers:

1. To what street address do the geodetic co-ordinates in the accompanying picture, figure 2.13, refer?
2. What is the general nature of the object at the address? Give a picture!
3. Are the co-ordinates geocentric? On which reference ellipsoid?
4. Find the “Easter egg” in the [index](#) of this text.





Map projections, datums and transformations

3

Le 24 février dernier, la Chambre des députés adoptait un projet de loi, dû à l'initiative parlementaire et ayant pour objet de fixer à nouveau l'heure légale de notre pays. Il était ainsi formulé : *l'heure légale, en France et en Algérie, est l'heure, temps moyen, de Paris, retardée de 9 minutes 21 secondes.*

How France legislated in 1898 civil time in France and Algeria to be Greenwich Mean Time. [Dastre \(1898\)](#)



3.1 Map projections

Although geodetic co-ordinates are more user friendly than rectangular ones, even better co-ordinates, closer to the human user, can be produced by using the method of *map projection*.

The practice of depicting the Earth's surface on a two-dimensional plane or map is an old one. *Cartography* is the science that has grown around this art. If the area to be depicted is small, the method of depiction is straightforward and error-free: local horizontal rectangular landscape co-ordinates (x, y) can be depicted, through a *scale*, onto the paper map plane. We speak of a *plan*.¹ In most countries, including Finland, the x axis points north and the y axis east.

For a larger area, we use a *map projection*, a mathematical method to map the location, or latitude and longitude (φ, λ) , of a point to the map plane. Thus we can draw a graphical representation of the Earth's surface onto paper, a *map*. Many digital applications, like [CAD](#) software,

¹Swedish *plankarta* (archaic; nowadays the term refers to zoning maps), Dutch *plattegrond*.

that do not actually even require the use of a paper map, are nevertheless based on the intuitive use of the map plane.

In the simplest approach we use the co-ordinate pair (φ, λ) directly as map co-ordinates x and y : $x = S\varphi$, $y = S\lambda$, S being the nominal scale. This is a pathetic solution, because

- The co-ordinates φ and λ are in degrees, angular units, whereas the map co-ordinates have to be in metric units.
- One degree of longitude λ expressed in kilometres diminishes toward the poles. At the latitude of Helsinki, a degree of longitude is only 55 km, when at the equator it is 111 km. A slightly better approach is to use the co-ordinate pair $(\varphi, \lambda \cos \varphi)$.

karttaprojektio-
oppi

Better solutions are offered by *map projection science*. Thus we may map the parameter pair (φ, λ) on the surface of the reference ellipsoid onto the pair (x, y) in the map plane in a sensible way. Unfortunately a method that depicts everything exactly correctly does not exist. *Something is always distorted* ([Wikipedia, Theorema Egregium](#)).

In map projections we always make approximations. We always lose something. There are no projections that do not distort anything or have the same scale throughout the map plane. A projection is chosen *according to the purpose of use*, so that something that is important to the user is preserved: the shapes of objects, their surface areas, some distances, the compass direction.

kulmatarkka

- If angles and ratios of distances are preserved, we speak of a *conformal* or angle-preserving projection. In this case both the linear scale — distances — and the scales of surface areas are distorted, except at some special points of the map.

pinta-alararkka

- If surface area is preserved, we speak of an *equivalent* or equal-area projection. Here, angles and shapes are distorted, again with the exception of special points or lines.

etäisyystarkka

- If distances are preserved, we speak of an *equidistant* projection. A projection can be equidistant only along certain lines, never everywhere.
- In navigation it is important that compass directions are preserved. In that case, *loxodromes* or rhumb lines ([Wikipedia, Rhumb line](#)), lines of constant compass direction, are straight. The normally oriented *Mercator* projection has this property.

isoympyrä

- If great circles are depicted as straight lines, we have the *gnomonic*



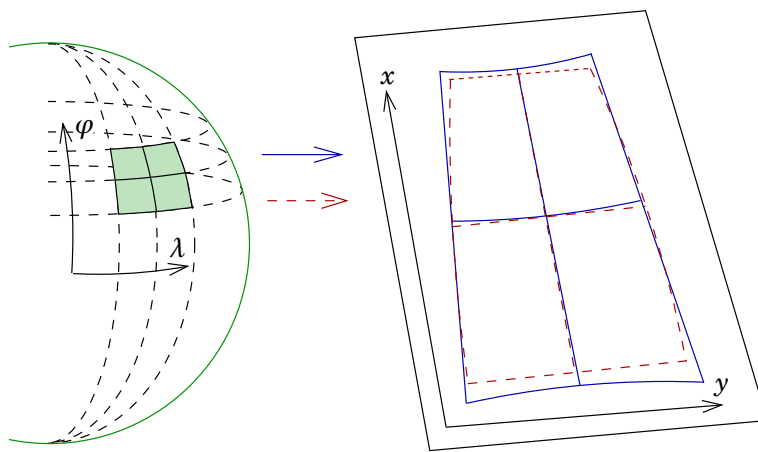


FIGURE 3.1. Depicting the curved surface of the Earth to the map plane using different projections. Something will always be distorted!

projection. It is used in aviation and in connection with meteor observations — the track of an aircraft on the Earth's surface or the track of a meteor on the sky is mapped to a straight line, which may be drawn with a ruler.

It is important to be aware that *there is no "correct" projection!* The choice is dictated by the purpose of use and the distortions considered acceptable. In a way they are all "wrong". On the other hand, they are also all useful.²

Philosophically one may be of the view, that map co-ordinates (x, y) are not actually geodetic quantities. Genuinely geodetic quantities are always three-dimensional. Map projection co-ordinates are *derived quantities*.

A high-level introduction into the mathematics of map projections is offered by [Grafarend et al. \(2014\)](#).

3.2 The various co-ordinate solutions used in Finland

In Finland, like everywhere on Earth, several different co-ordinate reference frames³ are in use for historical reasons. This complicates the use of geographic information and co-ordinate materials. It is a precondition for the use of co-ordinate data, that we know in which co-ordinate reference frame the material is. If necessary, one *transforms* from one frame to

²"Essentially, all models are wrong, but some are useful" – George E. P. Box (1919–2013), British statistician.

³Here, we do not yet make any clear distinction between co-ordinate reference systems and reference frames. Find more on this in section 3.8.



FIGURE 3.2. The systematic shift between the road network and the aerial photograph base could have something to do with the use of different co-ordinate reference frames. Google Earth™. [Google terms of use](#). Map data © 2009 Google. Image © 2010 DigitalGlobe, © 2010 Tele Atlas, © 2010 Europa Technologies.



another before use.

The most fundamental distinction between different co-ordinate reference frames is, whether we have a “traditional” one, created before the satellite era using traditional geodetic measurement methods, or a “modern”, geocentric, one created using satellite positioning technology. In the following, we explain more about co-ordinate reference frame alternatives and the transformations between them.

paikkatieto

However, even if it is clear in principle what co-ordinate frame geospatial data is in, small differences in how the data was processed and connected to known locations on the ground may lead to visible discrepancies in the resulting co-ordinates. See figure 3.2 for an example.



3.2.1 National Map Grid Co-ordinate System (KKJ)

kolmiomittaus

In Finland, the old, in its time official, and now largely obsolete, co-ordinate reference frame is **KKJ**, the *National Map Grid Co-ordinate System* (Parm, 1988). The system was created based on the results of the national triangulation effort. It is also a good example of the use of a *map projection* for mapping a national territory onto a two-dimensional



map plane. More about this map projection in subsection 3.3.1.

3.2.2 Helsinki system (VVJ)

The *Helsinki system*, or **VVJ**, was the predecessor of **KKJ**. The national triangulation was a huge project (Puupponen and Järvinen, 2008), lasting over half a century. The users of accurate co-ordinates, however, could not wait. Thus, **VVJ** was created “on the fly” in support of practical mapping work. The co-ordinates were calculated stage after stage, as the Finnish triangulation progressed from Southern Finland towards the north. Not until the 1970s, when the national triangulation was almost completed and its complete adjustment carried out, was the **KKJ** system formed as a nationally unified system. tasoitus

The co-ordinate differences between the Helsinki system and **KKJ** are up to a few metres (Häkli et al., 2009, page 18).

Ancient as the Helsinki system may be, it was still in use in many municipalities in the 2010s: for example Tuusula did not move away from it — directly to **EUREF-FIN** — until 2013...

3.2.3 The EUREF-FIN reference frame

With the spread of satellite positioning, as early as in the 1990s it was possible to determine three-dimensional co-ordinates for everywhere on Earth, in a system that was *geocentric* on the several-centimetre level: the origin of the three-dimensional co-ordinates (X, Y, Z) coincides to that level of precision with the centre of mass of the Earth. In addition, the Z axis of a geocentric co-ordinate system points in the direction of the Earth’s axis of rotation, whereas the X and Y axis directions lie within the equatorial plane.

Geocentric systems are *global*. In Europe, the International Association of Geodesy **IAG**⁴ created the geocentric reference system **ETRS89**, the European Terrestrial Reference System 1989. The system has been *realised* by measurements many times: the first Europe-wide realisation, or *co-ordinate reference frame*, was **EUREF89**, which was based on a space geodetic measurement campaign (Overgaauw et al., 1994). Later, Finland created its own national realisation under the name **EUREF-FIN**,⁵

⁴This was done by its European reference-frame subcommission **EUREF**. The name **EUREF** was proposed by the Danish geodesist Knud Poder. Knud Poder (1925–2019) was an eminent Danish land surveyor and mathematical geodesist and a pioneer in computational geodesy.

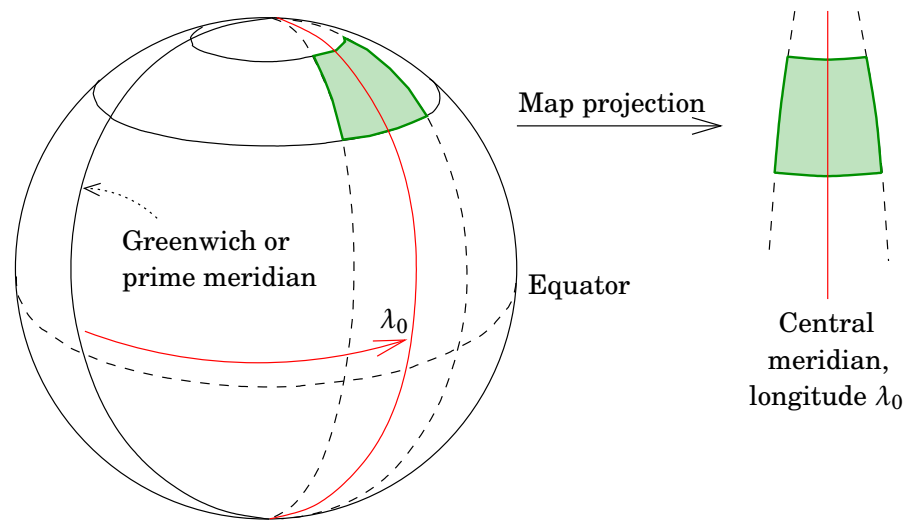


FIGURE 3.3. Mapping the curved Earth's surface as a narrow zone onto a plane. This is the principle of both the Gauss–Krüger and UTM projections. Both projections are used in Finland. The distortions remain acceptably small only in a relatively narrow area.

which has in recent decades been taken into use by the national mapping agencies and most other actors.

In subsection 3.3.3 we explain more about the map projections chosen for use with the EUREF-FIN reference frame.



3.3 Map projections used in Finland

In Finland we encounter currently two different plane or projected coordinate reference frames: the old **KKJ**, and the new system based on **EUREF-FIN**. In the old **KKJ**, the Gauss–Krüger projection was used, in the new system is used, in addition to Gauss–Krüger, also the **UTM** (Universal Transverse Mercator) map projection. See figure 3.3.



3.3.1 The map projection system of the **KKJ**

projektiokaista

In order to map the Earth's surface to the map plane without large distortions, Finland was divided into six **KKJ** projection zones, every one of which has its own co-ordinate frame in the projection plane. The zones

⁵The often-used name **WGS84** refers to the system — with actually half a dozen realisations — maintained by the US defence authorities. The system's latest realisations are consistent with **ITRF** frames on the centimetre level. Consistency with the various **ETRS89** realisations is poorer than that, although better than a metre. See [Malys et al. \(2016\)](#). Often the name **WGS84** is used, erroneously, as shorthand for all of these.

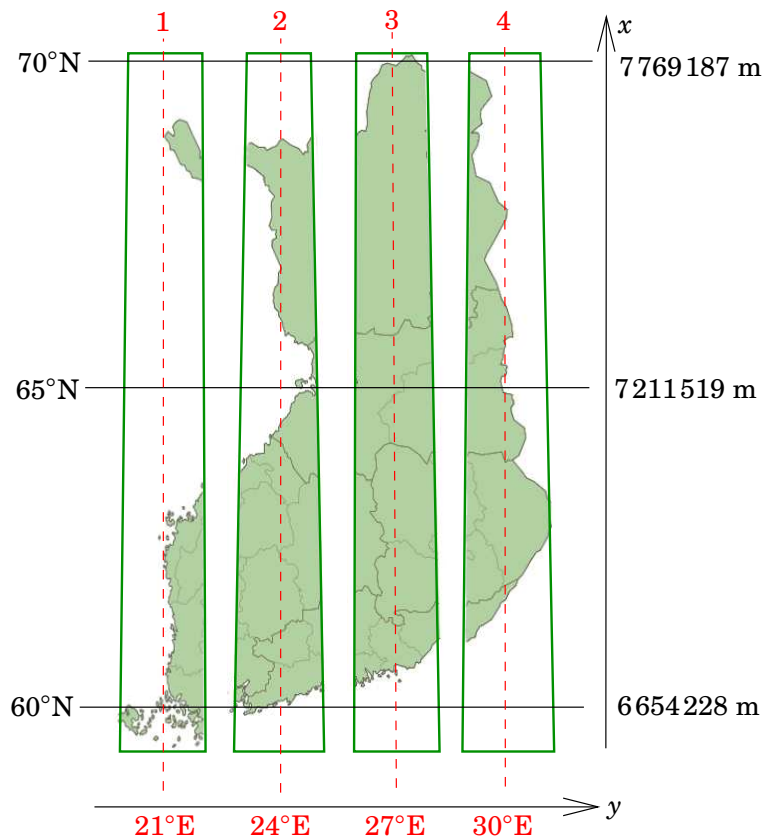


FIGURE 3.4. The zone division of the Finnish **KKJ** system's Gauss–Krüger projection. The zones 0 (central meridian 18°) and 5 (33°) have been left out.



were numbered from zero to five: the central meridians of the zones are at longitudes 18°, 21°, 24°, 27°, 30° and 33° east.

The co-ordinates within a projection zone are x (northing) and y (easting). The projection used is *conformal* and goes by the name of Gauss–Krüger, one type of transversal Mercator projection. The reference ellipsoid used in the projection calculations is the Hayford or International Ellipsoid of 1924.

In the several zones, the same co-ordinates appear multiple times. Therefore, to obtain unique values, the zone number is prepended to the y co-ordinate as its first decimal (except when it was 0). See figures 3.4, 3.5. This multi-zone system is referred to as the *basic co-ordinate system* of the **KKJ**.

The x co-ordinate origin is on the equator, and the values grow along the central meridian. Because of this, the x co-ordinates in Finland are in the range 6 600 000–7 800 000 m.

poikittainen

perus-
koordinaatisto



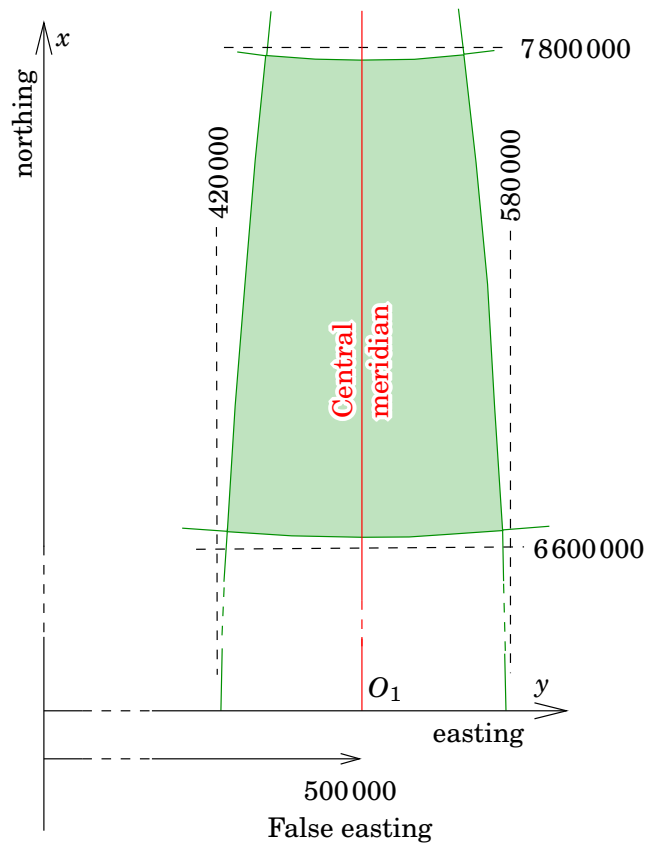


FIGURE 3.5. The geometry of one zone of the Finnish **KKJ** system (stretched in the east-west direction).

The y co-ordinate describes *distance from the central meridian*. In order to avoid negative y co-ordinates, 500 km is added to them, so that a point on the central meridian has a y co-ordinate of 500 000 m (*false easting*). The y co-ordinates could theoretically be in the interval 0–1 000 000 m; in practice, however, the range of values is 420 000–580 000 m, due to the narrowness of the zones at Finnish latitudes.

For example, according to figure 3.5, the bottom-left corner co-ordinates of the area of the first zone would be

$$x = 6\,600\,000.000\text{ m}, \quad y = 1\,420\,000.000\text{ m}.$$

Similarly the bottom-left corner of the third zone:

$$x = 6\,600\,000.000\text{ m}, \quad y = 3\,420\,000.000\text{ m}.$$

On the central meridian of the projection, the distortion is zero: the nominal scale of the map applies exactly.

In the case where **KKJ** co-ordinates were used within a small area, the leftmost digits were often left off, as they are all the same. Thus



one obtained *truncated co-ordinates*, which may still be encountered in municipal calculation documents. katkaistut koordinaatit

The map projection system employed by the Helsinki system (VVJ) was similar to that of KKJ, except that the notation was $y = 21^{\circ}420000$, so the y co-ordinate was prefixed with the longitude of the central meridian itself.

3.3.2 The KKJ's Uniform Co-ordinate System

In addition to KKJ, for small-scale maps⁶ only, the *Uniform Co-ordinate System* (Finnish acronym: YKJ) of KKJ was taken into use, which mapped the whole territory of Finland using one and the same Gauss–Krüger projection with a central meridian of 27° . As there is only one zone, nothing is prepended to the y co-ordinate. 6
yhtenäis-
koordinaatisto

3.3.3 The map projections of EUREF-FIN

In connection with the new reference frame we use EUREF-FIN map projection co-ordinates (JUHTA, 2016b), which thus are *plane co-ordinates* (x, y). The new map projections, which use the geocentric GRS80 reference ellipsoid, are, depending on the map scale, either the familiar Gauss–Krüger, or alternatively UTM, Universal Transverse Mercator, a projection type that is also conformal and is in broad international use.

- For *small-scale* maps which depict all of or a large part of Finland, one chooses the new ETRS-TM35FIN⁷ projection co-ordinate frame, which is based on the three-dimensional EUREF-FIN co-ordinate frame and the UTM projection for central meridian 27° E. This replaces the old KKJ Uniform Co-ordinate System. The projection also forms the basis for the whole Finnish map-sheet division. 7
- ETRS-TM35FIN is also used for topographic maps.
- For *large-scale* maps, intended for local use, one uses the Gauss–Krüger projection like before, but with a zone width of 1° . These

⁶A small-scale map is a map the *scale number* M of which is *large*, if the scale is $1 : M$. A scale of $1 : 1000000$ is small: even large objects look small on the map, but the area mapped is large. A scale of $1 : 2000$ is large: even small details are well discernible, but the area covered by the map is small.

⁷The name is a tongue and memory breaker. Alternative names circulating in the user community are “Finnish UTM” and “Finnish Map Grid”.

kaavoitus

projections are named **ETRS-GK n** , in which n is the degree number of the central meridian, for example 23°E, in which case the name of the projection is **ETRS-GK23**. For practical reasons, however, this rule is bent; for example, in order to get a whole municipality mapped onto a single zone. Due to the choice of projection type, these maps are useful for applications where no larger distortions of the scale can be accepted, like zoning and infrastructure construction, where map co-ordinates are moved straight to construction projects as terrain co-ordinates.

The **UTM** system is also much used internationally. **UTM** differs from Gauss–Krüger in two ways:

- The scale on the central meridian is 0.9996 rather than 1.0. This means that on the central meridian the map depicts details 400 ppm (*parts per million*) smaller than they ought to be based on the map's nominal scale. That amounts to 40 cm for every kilometre.
- The width of the projection zones is 6°, not 3° or 1°. This means that the scale distortion, which on the central meridian amounts to –400 ppm, turns positive going to the edges of the zone, to about +1000 ppm, calculated on the equator. At Finnish latitudes the distortion remains quite a bit smaller than that.

These two choices aim to keep the scale distortion over the whole area of the zone within certain bounds, in spite of the zone's large width.

Both Gauss–Krüger and **UTM** are *conformal* projections: angles and length ratios are preserved locally, so circles map to circles and squares to squares — but only if they are small.⁸



3.3.4 *The triangulated affine transformation of the Finnish National Land Survey*

The Finnish National Land Survey provides on its web-site a service that allows one to transform point co-ordinate sets between the older **KKJ** (actually, the Uniform Co-ordinate System **YKJ**) and the new **ETRS-TM35FIN**. The service is based on a division of the Finnish territory into triangles using a Delaunay⁹ triangulation, and an affine or bilinear

⁸If the square is too large, the nonlinearity of the map projection will cause it to be mapped to a curved quadrangle. A large circle maps to a circle only for the stereographic projection, otherwise it too will be deformed.



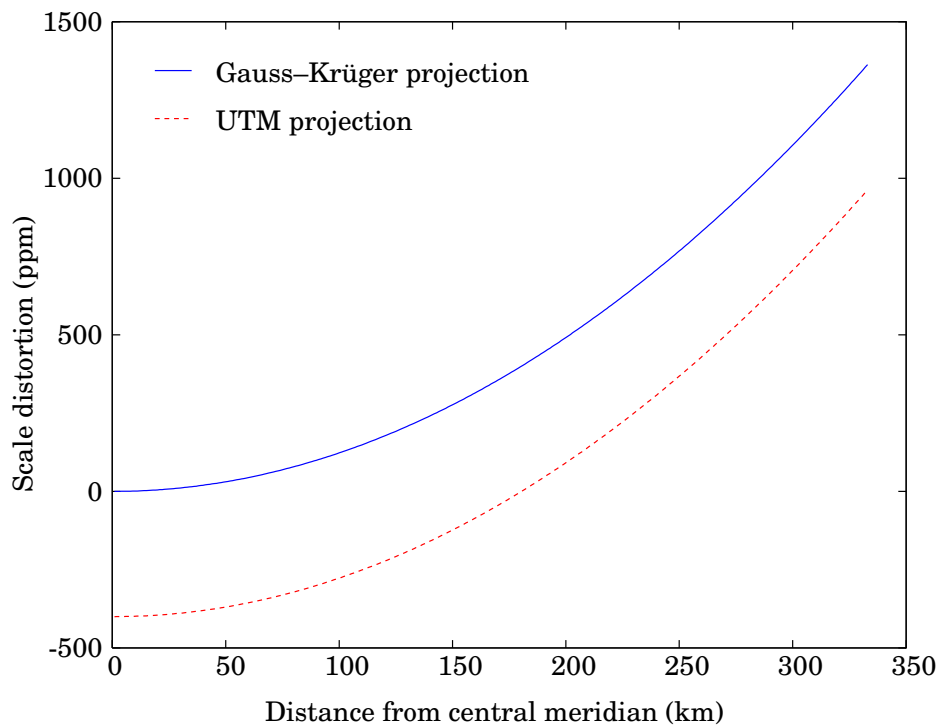


FIGURE 3.6. Scale distortion of Gauss–Krüger and UTM projections.

transformation within every triangle. The transformation is linear within every triangle and continuous across triangle boundaries, JUHTA (2016b). See figure 3.7.

The method is simple: for the triangle $\triangle ABC$ we define first for the vertices A , B , and C the co-ordinate differences Δx and Δy between the co-ordinate frames:

$$\left. \begin{aligned} \Delta x_i &\stackrel{\text{def}}{=} x_{i,\text{YKJ}} - x_{i,\text{ETRS-TM35FIN}} \\ \Delta y_i &\stackrel{\text{def}}{=} y_{i,\text{YKJ}} - y_{i,\text{ETRS-TM35FIN}} \end{aligned} \right\} i = A, B, C.$$

After this, we define *barycentric co-ordinates* for an arbitrary point P , co-ordinates (x, y) :

$$p^A \stackrel{\text{def}}{=} \frac{\omega^A}{\omega}, \quad p^B \stackrel{\text{def}}{=} \frac{\omega^B}{\omega}, \quad p^C \stackrel{\text{def}}{=} \frac{\omega^C}{\omega},$$

with (the vertices A , B , and C of the triangle are assumed to be in

⁹Boris Nikolayevich Delaunay (1890–1980) was a Russian mathematician and mountaineer. His grandson Vadim Nikolayevich Delaunay (1947–1983) was a poet who participated in the 1968 Red Square demonstration against the occupation of Czechoslovakia. Apparently not related to the French celestial mechanist Charles-Eugène Delaunay (1816–1872).

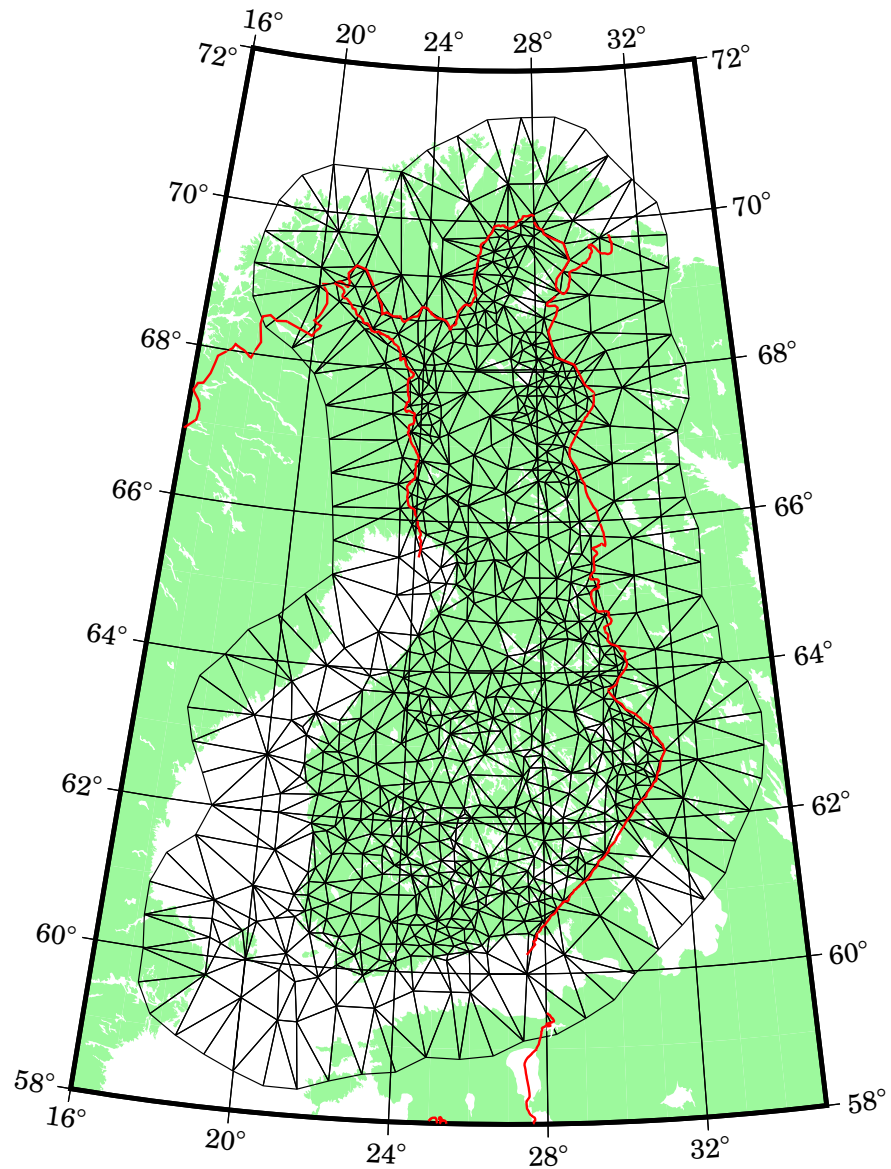


FIGURE 3.7. The triangulated affine transformation of the Finnish National Land Survey. Triangle network used.

clockwise order):

$$\omega^A \stackrel{\text{def}}{=} \frac{1}{2} \det \begin{bmatrix} 1 & 1 & 1 \\ x_B & x_C & x \\ y_B & y_C & y \end{bmatrix}, \quad \omega^B \stackrel{\text{def}}{=} \frac{1}{2} \det \begin{bmatrix} 1 & 1 & 1 \\ x_C & x_A & x \\ y_C & y_A & y \end{bmatrix},$$

$$\omega^C \stackrel{\text{def}}{=} \frac{1}{2} \det \begin{bmatrix} 1 & 1 & 1 \\ x_A & x_B & x \\ y_A & y_B & y \end{bmatrix}, \quad \omega \stackrel{\text{def}}{=} \frac{1}{2} \det \begin{bmatrix} 1 & 1 & 1 \\ x_A & x_B & x_C \\ y_A & y_B & y_C \end{bmatrix},$$

all determinants. Note that ω^A , ω^B , and ω^C , like p^A , p^B , and p^C , are *linear functions* of the co-ordinates (x, y) of the arbitrary point P . If we



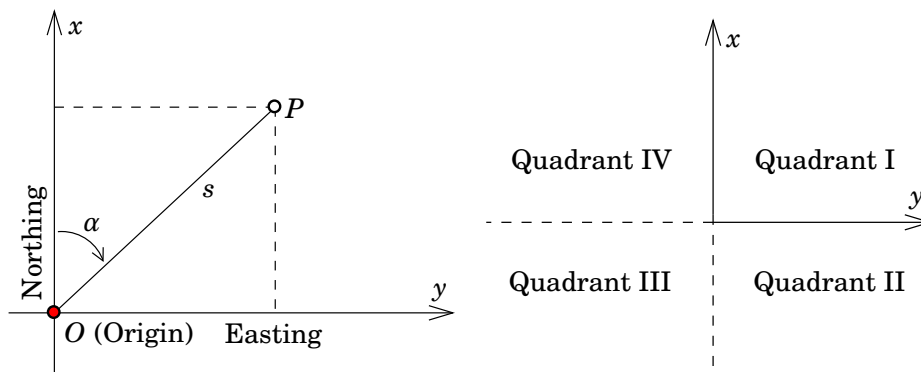


FIGURE 3.8. Geodetic plane co-ordinates and the quadrants of the plane I–IV.

write, for point P , the co-ordinate corrections as

$$\begin{aligned} \Delta x_P &\stackrel{\text{def}}{=} p^A \Delta x_A + p^B \Delta x_B + p^C \Delta x_C, \\ \Delta y_P &\stackrel{\text{def}}{=} p^A \Delta y_A + p^B \Delta y_B + p^C \Delta y_C, \end{aligned} \tag{3.1}$$

we are carrying out a bilinear interpolation between the points A , B , and C .

It is easy to show that

- Equations 3.1 reproduce the corrections for the vertices: $\Delta x_P \rightarrow \Delta x_A$ if $P \rightarrow A$, etc.
- The corrections are *continuous across triangle edges*: for example, on edge AB , a straight line, the functions $(\Delta x_P, \Delta y_P)$ are linear along the edge, and reproduce $(\Delta x_A, \Delta y_A), (\Delta x_B, \Delta y_B)$ at both ends. They must be the same for both triangles sharing the edge AB .
- As a result of this, the interpolation is continuous for the whole triangulation cover.

kolmiointi

3.4 More about plane co-ordinates

The plane co-ordinates used in geodesy differ a little from the familiar mathematical (x, y) system. See figure 3.8. While in mathematics, the x axis points to the right and the y axis to the upper edge of the paper, in geodesy it is the habit that the x axis points north (“northing”) and the y axis east (“easting”).

In addition to rectangular co-ordinates (x, y) , polar co-ordinates (α, s) are used. Sometimes the symbols used are (A, s) . In geodesy, the *azimuth* or horizontal direction angle α (or A) turns clockwise from the north,¹⁰

¹⁰In astronomy, sometimes, especially in older texts, the azimuth turns from the south



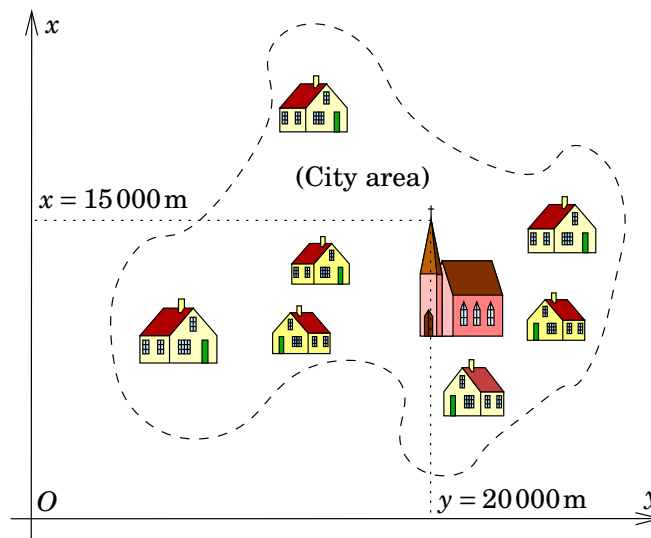


FIGURE 3.9. A local co-ordinate frame.

that is through the east, unlike in mathematics. s is the distance from the co-ordinate origin O .

The following trigonometric relations exist between rectangular and polar co-ordinates:

$$\begin{aligned} y &= s \sin \alpha &\implies \sin \alpha &= \frac{y}{s}, \\ x &= s \cos \alpha &\implies \cos \alpha &= \frac{x}{s}, \\ \tan \alpha &= \frac{\sin \alpha}{\cos \alpha} = \frac{y}{x} &\implies \alpha &= \arctan \frac{y}{x} + k \cdot 200^{\text{g}}, \quad x \neq 0. \end{aligned}$$

In this equation, the small integer k is chosen such, that the result α lies in the *correct quadrant*: $\arctan(y/x)$ is always in the interval $(-\frac{\pi}{2}, +\frac{\pi}{2})$, that is in the quadrants I or IV. This is easiest to verify by making a sketch.

The Pythagoras theorem yields the distance s :

$$s = \sqrt{x^2 + y^2}.$$



3.4.1 Local co-ordinates

Local, often old, co-ordinate frames were long used in many Finnish municipalities, and one comes across them in older documents.

to the west, that is also clockwise. In geodesy, too, practices vary: one must always check.

The word “azimuth” originates from the Arabic *as-sumût*, “directions”.



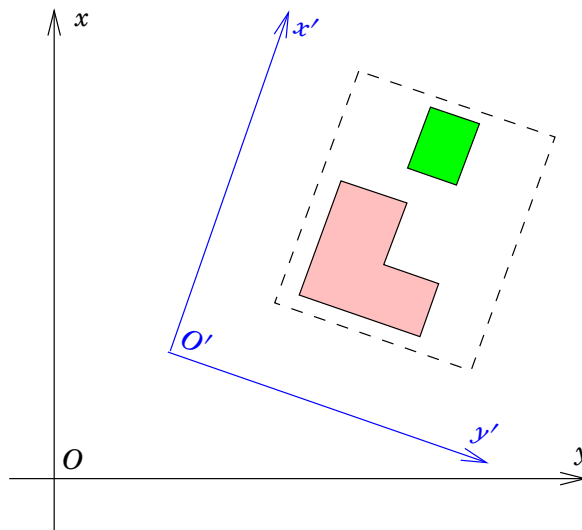


FIGURE 3.10. Temporary co-ordinates.

The origin is generally located so, that everywhere in the municipality there are only positive x and y values. Often the origin is just a computational point without realisation on the terrain, for example “truncated” [KKJ](#) co-ordinates.

The connection to the national frame may be that one also knows the co-ordinates of a church spire or some other landmark in the [KKJ](#) system. Then, one can transform local and national co-ordinates into each other by adding constant shifts, *translations*, to both co-ordinates x and y .

siirto

In more precise work, one point is not enough: a sufficient number of common points is needed, and their co-ordinates must be known in both the local and the national system.

3.4.2 Temporary co-ordinates

Sometimes it makes sense to use in measurements a temporary co-ordinate frame that deviates from the general system. Even the directions of the axes may deviate from the customary north and east directions.

A temporary or project-specific co-ordinate frame is only used during measurement or, for example, during a construction project. Later on, the co-ordinates may be transformed to a more permanent, local or national, correctly oriented system.

The origin and axes orientation can be chosen in accordance with the measurement at hand, for example along the walls of a building.

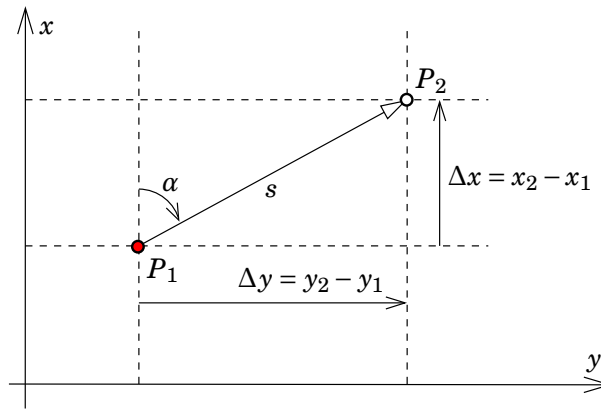


FIGURE 3.11. The forward geodetic problem in the plane.



3.5 The geodetic forward and inverse problems

geodeettinen
päätehtävä

The “*direct or forward geodetic problem*” means the determination of the co-ordinates of an unknown point, when the co-ordinates of a starting point and both the azimuth (horizontal direction) and the distance from the starting point are given.

In the general case, on an arbitrary curved surface, the forward geodetic problem has no easy solution. On a sphere, however, a closed — though not particularly simple — solution already exists. On the surface of a reference ellipsoid the solution is obtained numerically. An on-line service is offered by, among others, the US National Geodetic Survey on their web-site [NGS, Computation utilities](#).

In a plane co-ordinate system, two-dimensionally, the forward geodetic problem is simpler, as we shall see next.



3.5.1 The forward geodetic problem in the plane

Let there be *in the plane* two points P_1 and P_2 , figure 3.11. The plane co-ordinates (x_1, y_1) of starting point P_1 , as well as the azimuth α and length s of vector $\overrightarrow{P_1P_2}$, are given. The co-ordinates (x_2, y_2) of the unknown point P_2 are to be calculated.

The solution is obtained as follows:

$$\begin{aligned}\sin \alpha &= \frac{\Delta y}{s} \implies \Delta y = s \sin \alpha, \\ \cos \alpha &= \frac{\Delta x}{s} \implies \Delta x = s \cos \alpha,\end{aligned}$$

and

$$x_2 = x_1 + \Delta x = x_1 + s \cos \alpha,$$

$$y_2 = y_1 + \Delta y = y_1 + s \sin \alpha.$$



Example Given point A with co-ordinates

$$x_A = 6\,800\,000\text{ m}, \quad y_A = 400\,000\text{ m}.$$

If the distance from point B is $s = 2828.427\text{ m}$, and the azimuth (direction angle) $\alpha = 50^\text{g}$, solve the forward geodetic problem for the points A and B .

Solution

$$\Delta x = s \cos \alpha = 2828.427\text{ m} \cdot \cos(50^\text{g}) = 2000\text{ m},$$

$$\Delta y = s \sin \alpha = 2828.427\text{ m} \cdot \sin(50^\text{g}) = 2000\text{ m},$$

$$x_B = x_A + \Delta x = 6\,800\,000\text{ m} + 2000\text{ m} = 6\,802\,000\text{ m},$$

$$y_B = y_A + \Delta y = 400\,000\text{ m} + 2000\text{ m} = 402\,000\text{ m}.$$



3.5.2 The inverse geodetic problem in the plane

The *inverse geodetic problem* means the determination of the azimuth (horizontal direction) and distance between two given points.

geodeettinen
käänteistehtävä

Let there be given again two points in the plane, P_1 and P_2 (figure 3.11). Let their rectangular co-ordinates be (x_1, y_1) and (x_2, y_2) . To be calculated are α and s .

Solution

$$s = \sqrt{\Delta x^2 + \Delta y^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

$$\tan \alpha = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}. \quad (3.2)$$

We rather do not use $\sin \alpha = \Delta y/s$ or $\cos \alpha = \Delta x/s$ for determining α : the sine formula becomes imprecise when $\alpha \approx 100^\text{g} + k \cdot 200^\text{g}$, and the cosine formula when $\alpha \approx 0^\text{g} + k \cdot 200^\text{g}$, $k \in \mathbb{Z}$.

In those points, the functions $\sin \alpha$ and $\cos \alpha$ are *stationary*: a large change in α causes only a small change in the function value $\sin \alpha = \Delta y/s$, and therefore, from the given values Δy and s , precise as they may be, one can calculate α only imprecisely.

It is better to calculate α from equation 3.2:

$$\alpha = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) + k \cdot 200^\text{g}, \quad k = \begin{cases} 0 & \text{if } (x_2 - x_1) > 0, \\ 1 & \text{if } (x_2 - x_1) < 0, \end{cases}$$

$$\alpha = \begin{cases} 100^\text{g} & \text{if } x_1 = x_2 \text{ and } y_2 > y_1, \\ 300^\text{g} & \text{if } x_1 = x_2 \text{ and } y_2 < y_1. \end{cases}$$



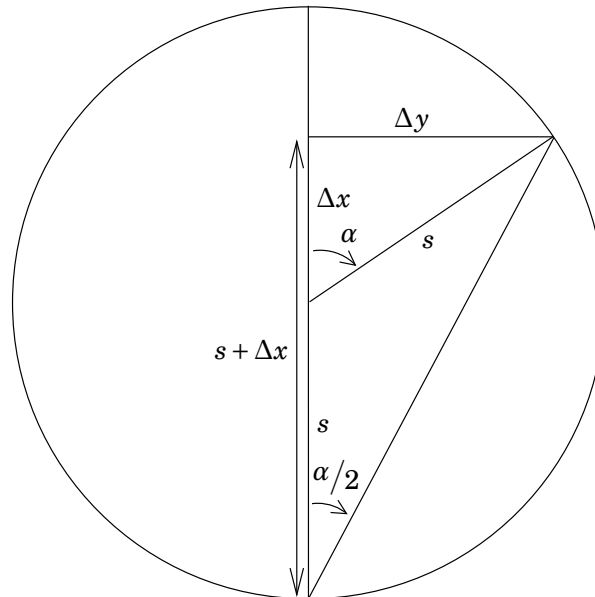


FIGURE 3.12. The half-angle formula for the arc tangent.

It must be remembered here that the values of the arctan function are, according to the definition, always in the interval $(-\frac{\pi}{2}, +\frac{\pi}{2})$. The correct α value may however well be outside this interval, like when $(x_2 - x_1) < 0$. This is why there is the conditional term $k \cdot 200^g$.

Many programming languages offer the function $\text{atan2}(x, y)$ with two arguments, which also finds the correct quadrant automatically.

- ¹¹ A more elegant solution is to use the *half-angle formula*:¹¹

$$\alpha = 2 \cdot \left(\frac{\alpha}{2} \right) = 2 \arctan \frac{\Delta y}{\Delta x + s} = 2 \arctan \frac{\Delta y}{\Delta x + \sqrt{\Delta x^2 + \Delta y^2}}.$$

See figure 3.12.

Example Given point A: $x_A = 6\,800\,000\text{ m}$, $y_A = 400\,000\text{ m}$, and point C, co-ordinates $x_C = 6\,793\,000\text{ m}$, $y_C = 407\,000\text{ m}$, solve the inverse geodetic problem for points A and C.

Solution

$$\Delta x = x_C - x_A = 6\,793\,000\text{ m} - 6\,800\,000\text{ m} = -7000\text{ m},$$

$$\Delta y = y_C - y_A = 407\,000\text{ m} - 400\,000\text{ m} = +7000\text{ m}.$$

¹¹This formula too breaks down in the edge case $\Delta y = 0$ and $\Delta x < 0$, because then $\alpha = 2 \arctan 0/0$ is undefined, when the correct solution is $\alpha = 200^g$.



1. The traditional method:

$$\begin{aligned}\alpha_{AC} &= \arctan \frac{\Delta y}{\Delta x} + k \cdot 200^{\text{g}} = \\ &= \arctan(-1) + k \cdot 200^{\text{g}} = -50^{\text{g}} + k \cdot 200^{\text{g}}.\end{aligned}$$

The correct solution is apparently

$$\alpha_{AC} = -50^{\text{g}} + 200^{\text{g}} = 150^{\text{g}}.$$

2. The half-angle formula:

$$\begin{aligned}\alpha_{AC} &= 2 \arctan \frac{\Delta y}{\Delta x + \sqrt{\Delta x^2 + \Delta y^2}} = \\ &= 2 \arctan \frac{7000}{-7000 + 7000\sqrt{2}} = \\ &= 2 \arctan \left(\frac{1}{\sqrt{2} - 1} \right) = 2 \cdot 75^{\text{g}} = 150^{\text{g}}.\end{aligned}$$

And

$$s_{AC} = \sqrt{\Delta x^2 + \Delta y^2} = 7000 \text{ m} \cdot \sqrt{2} = 9899.495 \text{ m}.$$



3.6 The similarity co-ordinate transformation

The similarity¹² or Helmert transformation is a transformation between two rectangular co-ordinate frames, usually in the plane, or two-dimensionally. It is encountered very often in practical measurement and computation tasks, when bodies of co-ordinate material in two or more co-ordinate frames need to be combined for joint use. In the general case we find a sufficient number of *common points* — often benchmarks — from the area, the co-ordinates of which are known in both systems, and carry out an *adjustment*. The special case in which we have only two common points, which is minimally sufficient for determining the transformation, is simpler.

yhden-
muotoisuus-
muunnos
12

kiintopiste

We know the co-ordinates of two benchmarks *A* and *B* in *both* coordinate systems:

$$(x_A, y_A), (x_B, y_B), \quad (u_A, v_A), (u_B, v_B).$$

¹²In Swedish: *likformighetstransformation*, in German: *Ähnlichkeitstransformation*, in French: *similitude*.





FIGURE 3.13. Friedrich Robert Helmert (1841–1917) was a great German geodesist and a developer of adjustment calculus and its theory. [Humboldt University Berlin \(2017\)](#).

In addition, we are given a set of points with co-ordinates only in the (u, v) system:

$$(u_1, v_1), (u_2, v_2), \dots, (u_i, v_i), \dots, (u_n, v_n).$$

The problem now is to compute a similarity transformation for this point field:

$$(u_i, v_i) \mapsto (x_i, y_i), \quad i = 1, \dots, n.$$

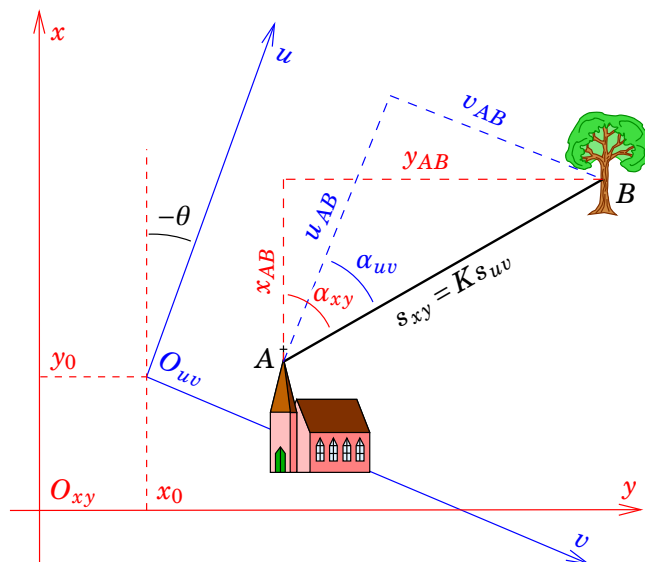


FIGURE 3.14. A similarity or Helmert co-ordinate transformation in the plane.



See figure 3.14. The transformation is carried out in the following steps:

1. A *shift*, or translation, of the origin $O_{uv} \mapsto O_{xy}$, translation parameters (x_0, y_0) . The components of the origin translation vector \mathbf{t} are $(-x_0, -y_0)$ in the (x, y) frame.
2. A *rotation* of the whole (u, v) co-ordinate axes frame by an angle θ kierto. The rotation angle of the axes is positive clockwise, but must be *subtracted* from the directions between points when going from (u, v) to (x, y) .
(In the figure, θ is negative.)
3. A *scale transformation* of the (u, v) co-ordinates to the scale of the (x, y) co-ordinates, by multiplying with the scale factor or *scale ratio* K .

The Helmert transformation in the plane is also called a *four-parameter transformation*. The parameters are x_0 , y_0 , θ , and K .

The general form of the Helmert transformation is (figure 3.14):

$$\begin{aligned} x &= x_0 + K \cos \theta \cdot u + K \sin \theta \cdot v, \\ y &= y_0 - K \sin \theta \cdot u + K \cos \theta \cdot v, \end{aligned} \quad (3.3)$$

in matrix form — see appendix A:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + K \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \quad (3.4)$$



3.7 Determining the transformation parameters

Determining the transformation parameters unambiguously requires at least *four* “observations” — for example, the total of four co-ordinates of two points (x_A, y_A) , (x_B, y_B) . Then we obtain four equations:

$$\begin{aligned} x_A &= x_0 + K \cos \theta \cdot u_A + K \sin \theta \cdot v_A, \\ y_A &= y_0 - K \sin \theta \cdot u_A + K \cos \theta \cdot v_A, \\ x_B &= x_0 + K \cos \theta \cdot u_B + K \sin \theta \cdot v_B, \\ y_B &= y_0 - K \sin \theta \cdot u_B + K \cos \theta \cdot v_B, \end{aligned}$$

again in matrix form:

$$\begin{bmatrix} x_A \\ y_A \\ x_B \\ y_B \end{bmatrix} = \begin{bmatrix} 1 & 0 & v_A & u_A \\ 0 & 1 & -u_A & v_A \\ 1 & 0 & v_B & u_B \\ 0 & 1 & -u_B & v_B \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ K \sin \theta \\ K \cos \theta \end{bmatrix}.$$



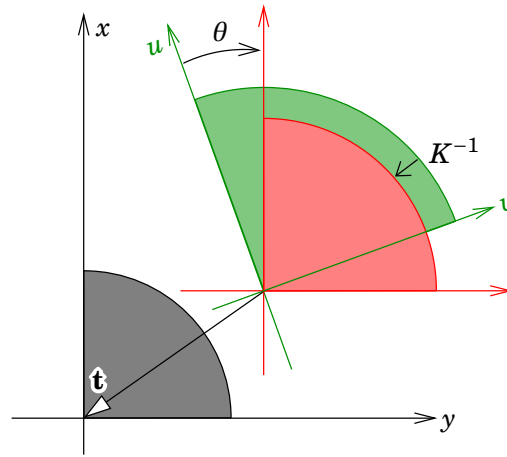


FIGURE 3.15. The stages of the Helmert transformation in the plane in terms of changes of the co-ordinate axes: axis rotation angle θ , scaling of axis units K^{-1} , translation vector of the origin \mathbf{t} . Note that the changes to the origin location, axes directions and axis unit lengths are precisely the opposite of what happens to point co-ordinates and numerical direction and distance values between points.



If the old co-ordinates (u_A, v_A) , (u_B, v_B) of the same points A and B are also known, one may from these equations solve four transformation parameters (x_0, y_0) and $(K \sin \theta, K \cos \theta) \mapsto (K, \theta)$ *uniquely*.

Difference transformation Subtraction yields

$$\begin{aligned} \begin{bmatrix} x_{AB} \\ y_{AB} \end{bmatrix} &= \begin{bmatrix} x_B - x_A \\ y_B - y_A \end{bmatrix} = \\ &= \begin{bmatrix} v_B - v_A & u_B - u_A \\ -(u_B - u_A) & v_B - v_A \end{bmatrix} \begin{bmatrix} K \sin \theta \\ K \cos \theta \end{bmatrix} = \\ &= \begin{bmatrix} v_{AB} & u_{AB} \\ -u_{AB} & v_{AB} \end{bmatrix} \begin{bmatrix} K \sin \theta \\ K \cos \theta \end{bmatrix}, \end{aligned}$$

with the logical definitions

$$\begin{aligned} x_{AB} &\stackrel{\text{def}}{=} x_B - x_A, & y_{AB} &\stackrel{\text{def}}{=} y_B - y_A, \\ u_{AB} &\stackrel{\text{def}}{=} u_B - u_A, & v_{AB} &\stackrel{\text{def}}{=} v_B - v_A. \end{aligned}$$

Arrange the terms cleverly anew:

$$\begin{bmatrix} x_{AB} \\ y_{AB} \end{bmatrix} = \begin{bmatrix} v_{AB} & u_{AB} \\ -u_{AB} & v_{AB} \end{bmatrix} \begin{bmatrix} K \sin \theta \\ K \cos \theta \end{bmatrix} =$$



$$= K \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} u_{AB} \\ v_{AB} \end{bmatrix}. \quad (3.5)$$

This equation, the *difference transformation*, applies for arbitrary point pairs.

The scale ratio or scale factor is obtained with Pythagoras' theorem:

$$K = \frac{s_{xy}}{s_{uv}} = \frac{\sqrt{x_{AB}^2 + y_{AB}^2}}{\sqrt{u_{AB}^2 + v_{AB}^2}}.$$

The rotation angle is¹³ (positive clockwise, i.e., turning from the x axis to the y axis): kiertokulma
13

$$\theta = \alpha_{uv} - \alpha_{xy} = \arctan \frac{v_{AB}}{u_{AB}} - \arctan \frac{y_{AB}}{x_{AB}}. \quad (3.6)$$

The translation vector is computed starting from the Helmert difference transformation, equation 3.5. Let us look at the point pair A and O_{uv} , the origin of the old (u, v) co-ordinate system. The points have co-ordinates (x_A, y_A) and (x_0, y_0) in the new system, and co-ordinates (u_A, v_A) and $(u_0, v_0) = (0, 0)$ in the original system. Then siirtovektori

$$u_{AO} = u_A - u_0 = u_A, \quad v_{AO} = v_A - v_0 = v_A,$$

and also

$$x_{AO} = x_A - x_0, \quad y_{AO} = y_A - y_0.$$

The difference transformation now gives

$$\begin{bmatrix} x_A - x_0 \\ y_A - y_0 \end{bmatrix} = K \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} u_A \\ v_A \end{bmatrix},$$

so the translation vector is

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_A \\ y_A \end{bmatrix} - K \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} u_A \\ v_A \end{bmatrix},$$

just the translation parameters we are after.

For an arbitrary point (x, y) the Helmert transformation equations are now, after all transformation parameters have been solved for:

$$\begin{aligned} x &= x_0 + K \cos\theta \cdot u + K \sin\theta \cdot v, \\ y &= y_0 - K \sin\theta \cdot u + K \cos\theta \cdot v, \end{aligned} \quad (3.3)$$

the same equations 3.3 already given above.

¹³Forget for a moment the quadrant problem. In principle, a term $k \cdot 200^\circ$ should be added, with k a small integer.



Symbolic matrix form The Helmert transformation in matrix form is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + K \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \quad (3.4)$$

This can be written compactly:

$$\bar{\mathbf{x}} = \bar{\mathbf{x}}_0 + KR\bar{\mathbf{u}}, \quad (3.7)$$

in which the definitions of the vectors (column matrices) and matrix are

$$\bar{\mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} x \\ y \end{bmatrix}, \quad \bar{\mathbf{x}}_0 \stackrel{\text{def}}{=} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \quad \bar{\mathbf{u}} \stackrel{\text{def}}{=} \begin{bmatrix} u \\ v \end{bmatrix}, \quad R \stackrel{\text{def}}{=} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}.$$

Often one writes $K = 1 + m$, in which m is the *scale distortion*. Usually this number is small and is expressed in the unit ppm (parts per million).

The transformation equations 3.4 and 3.7 are called the similarity or *Helmert transformation* in the plane.

Example

1. Given the co-ordinates of points A and B in the (u, v) co-ordinate system:

$$\begin{aligned} u_A &= 0 \text{ m}, & v_A &= 0 \text{ m}, \\ u_B &= 1500 \text{ m}, & v_B &= 1500 \text{ m}, \end{aligned}$$

and in the (x, y) co-ordinate system:

$$\begin{aligned} x_A &= 2000 \text{ m}, & y_A &= 3000 \text{ m}, \\ x_B &= 3500.150 \text{ m}, & y_B &= 4500.150 \text{ m}. \end{aligned}$$

Assuming that the transformation between the (u, v) and (x, y) systems is a Helmert transformation:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + K \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}, \quad (3.4)$$

calculate its parameters K , θ , x_0 , and y_0 .

2. Given the co-ordinates of point C in the (u, v) system:

$$u_C = 1000 \text{ m}, \quad v_C = 2000 \text{ m}.$$

Calculate x_C and y_C .



Solution

1. We see immediately that

$$u_{AB} = 1500 \text{ m}, \quad x_{AB} = 1500.150 \text{ m},$$

$$v_{AB} = 1500 \text{ m}, \quad y_{AB} = 1500.150 \text{ m}.$$

From this we infer visually, with the help of the *difference transformation* 3.5, that $K = 1.0001$ and $\theta = 0$.

After this, for point A:

$$x_A = x_0 + 1.0001 \cdot u_A \implies x_0 = x_A - 1.0001 \cdot u_A = 2000 \text{ m},$$

$$y_A = y_0 + 1.0001 \cdot v_A \implies y_0 = y_A - 1.0001 \cdot v_A = 3000 \text{ m}.$$

2. Calculate

$$x_C = x_0 + 1.0001 \cdot u_C = 2000 \text{ m} + 1000.1 \text{ m} = 3000.1 \text{ m},$$

$$y_C = y_0 + 1.0001 \cdot v_C = 3000 \text{ m} + 2000.2 \text{ m} = 5000.2 \text{ m}.$$

**3.8 Datums and datum transformations**

Geodetic co-ordinates are not just mathematical quantities. Measurements are carried out between points on the terrain, and their co-ordinates are computed, based on given starting or *datum points*. The choice of starting points is always to some extent arbitrary. Every choice made creates what geodesists call a *geodetic datum*. In other words, when geodetic measurements are made on a part of the Earth's surface using a certain set of measurement points, and co-ordinates are conventionally assigned to starting points chosen from this set, we get, in real life, a solution that represents only a *realisation* of a certain reference system.¹⁴

14

A co-ordinate reference frame or *datum* is generally established locally. When it meets another, similarly established (but based on different starting points) frame, the co-ordinates of the same points are generally not the same. For example, in the places where the Finnish and Swedish

¹⁴In English, we call the formal definition a *co-ordinate reference system*, whereas we call the realisation on the terrain a *co-ordinate reference frame*. For example, **ETRS** = *European Terrestrial Reference System* and **ETRF** = *European Terrestrial Reference Frame*. In Finnish, the corresponding terms are gaining traction, too: *vertausjärjestelmä* against its realisation or *vertauskehys*.



precise levelling networks meet at the border in the Torne river valley, we get two different height values for the same point, *which are both correct*.

tasoverkko In the case of horizontal networks we also speak of datums, *horizontal datums*: where networks meet at borders, the horizontal co-ordinates (φ, λ) of the same point are generally not precisely the same in both datums. The differences are, for classical triangulation networks, of the order of seconds of arc.

For the purpose of transforming the co-ordinates of points in one datum into co-ordinates of another datum, the literature offers *datum transformation* equations.



3.8.1 Example: height network

Let us look at the example of height measurement, *levelling*. The Finnish official height system until September 25, 2007, **N60**, was based on the height of a certain point. This starting or datum point is a granite pillar located in the garden of Helsinki astronomical observatory, and a certain polished surface on that pillar. Here a role is played by historical accident, by the fact of Helsinki being the Finnish capital. The height value of the datum point was chosen such, that the heights of points were rather precisely reckoned from mean sea level in Helsinki harbour at the start of 1960. Scientifically the choice of Helsinki was arbitrary.

pääkiintopiste The new Finnish height system, **N2000**, uses as its starting or datum point the fundamental benchmark PP2000 at the Metsähovi research station — 40 km to the west of Helsinki — the height value of which was chosen so that the heights are relative to the Amsterdam mean sea level **NAP**, one of the oldest in the world. The choice of Amsterdam is also the product of history, not science.

The national precise levelling has brought official heights to everywhere in the country. It is clear that the precision of a calculated point height in this system will depend on the point's distance from Helsinki. A height in Kevo, Northern Lapland, will be clearly more poorly known than a height in Jyväskylä. And the height of Turku is somewhat imprecise, because the levelling from Helsinki to Turku was not absolutely precise. On the other hand, the heights measured for points in the Helsinki area are very precise, because the *datum point*, be it Observatory Hill or Metsähovi, is so nearby.

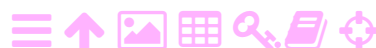




FIGURE 3.16. Fundamental benchmark PP2000 of the N2000 height datum at the Metsähovi research station. Established by the Finnish Geodetic Institute, today the Finnish Geospatial Research Institute (FGI) of the National Land Survey of Finland.



Imagine for a moment, that not Helsinki but Turku were the capital of Finland,¹⁵ and that a benchmark in the wall of Turku Cathedral were chosen as the datum point for the Finnish height system. In that case, all heights of points close to Turku would be very precise, but the points in the Helsinki area would be similarly imprecise as points in the Turku area are in the present system, as the levelling between Turku and Helsinki is somewhat imprecise.

The precision picture depends on the viewpoint, on the chosen *datum*.

Figure 3.17 shows a levelling network of four points. The height differences AB , BC , CD , and DA are given — measured. Furthermore, the heights above mean sea level of coastal points A and B , measured by a mareograph or tide gauge are also given.

1. First we adjust the levelling loop, table 3.1a.

¹⁵Not hard to imagine, as this is how it was 1809–1812, and informally before that during the Swedish imperial era.



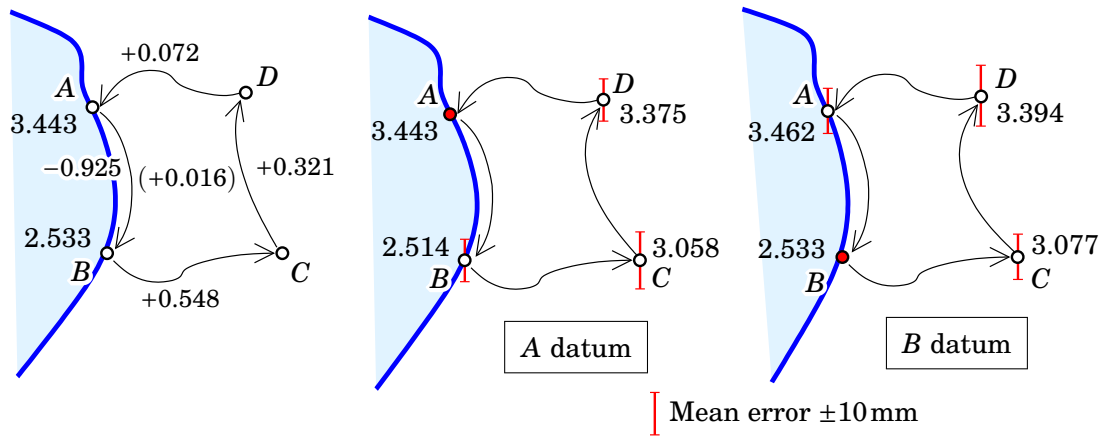


FIGURE 3.17. Alternative vertical datums A and B.

- 16
2. Use point A as *datum point* and calculate the heights of the points,¹⁶ table 3.1b.
 3. Do the same calculation, but now using point B as the datum point, table 3.1c.

It is seen that in the latter case all calculated heights are greater by 0.019m. The height *differences* are of course the same. The difference



TABLE 3.1. Alternative vertical datums A and B.

(a)
Adjusting the levelling loop

Interval	Observed	Correction	Adjusted
AB	-0.925	-0.004	-0.929
BC	+0.548	-0.004	+0.544
CD	+0.321	-0.004	+0.317
DA	+0.072	-0.004	+0.068
Closing error	+0.016	↑	0.000

(b) Point heights, datum A			(c) Point heights, datum B		
Point	Height	Mean error	Point	Height	Mean error
A	3.443	±0.000	A	3.462	±0.010
B	2.514	±0.010	B	2.533	±0.000
C	3.058	±0.014	C	3.077	±0.010
D	3.375	±0.010	D	3.394	±0.014

¹⁶The mean errors in the table are made up though realistic-looking.



of 0.019m is precisely the “difference of height differences” of points A and B between the two methods: (1) levelling plus adjustment, and (2) mareographs. The difference stems from the uncertainty, or “measurement error”, of the levelling and of the mareograph measurements, and from the circumstance that the true mean sea surface is not a level or equipotential surface.

The *datum difference* between datum A and datum B is 0.019m.

The *datum transformation* is

$$H_i^{(B)} = H_i^{(A)} + 0.019\text{m},$$

more generally

$$H_i^{(B)} = H_i^{(A)} + \left(H_A^{(B)} - H_A^{(A)} \right).$$

Every adjustment produces precision estimates of the computed values, or *mean errors*. If we assume that, as a given value, the height value of the datum point is *errorless* (mean error zero), the mean errors of the point heights will grow moving away from the datum point. Table 3.1 contains (invented) mean errors behaving in just this way. They are also drawn in figure 3.17 as error bars. We see that the precision behaviour of the network depends on the datum-point choice.



3.8.2 Horizontal co-ordinate datums

As already said, horizontal networks can also have different datums, *horizontal datums*. Figure 3.18 depicts one network both in datum AB and in datum PQ .

- The AB datum is created by taking already known approximate co-ordinate values for points A and B as the *formal truth*, and adjusting the whole network without changing these. In this way one calculates the co-ordinates of the other points, including P and Q , in the same AB datum.
- The PQ datum is again created in the same way, but by fixing the co-ordinates of points P and Q to prior given approximate values, and adjusting the network while keeping these fixed. Thus one also solves the locations of the other points in the network, but now in the PQ datum.

The figure shows that the point co-ordinates are different in the AB and PQ datums. The *shape* of the whole network is nevertheless the same, independently of whether datum AB or PQ is chosen for calculating



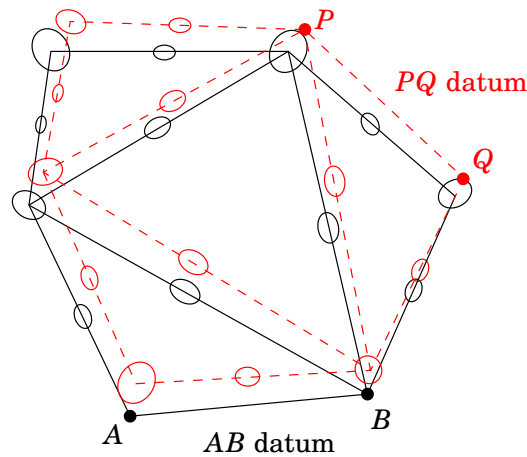


FIGURE 3.18. Two different datums of a horizontal network, the AB and PQ horizontal datums with their starting or datum points. Point error ellipses and inter-point (relative) error ellipses have been drawn. Note how the error vanishes in the datum points, the co-ordinates of which are part of the datum definition and thus assigned values.



the solution. In this case, the transformation between the datums is a similarity or *Helmert* transformation.

verkkotasoitus The prior known co-ordinates for the points A , B , P , and Q originate generally from previous network adjustments, astronomical position determinations, or are read from a map: they are *approximate values* of the co-ordinates. Determining a datum is thus the same as choosing the datum points: points the approximate co-ordinates of which are taken as the *formal truth* in the network adjustment.

It would be a coincidence if the computation of the network in the AB datum produced the same co-ordinates as computation in the PQ datum. The differences between the co-ordinates computed in different ways are comparable in magnitude to the *goodness of approximation* of the approximate co-ordinates used. The differences are often so small, that the transformation parameters are close to zero or unity. In the Helmert transformation 3.3:

$$x = x_0 + K \cos \theta \cdot u + K \sin \theta \cdot v,$$

$$y = y_0 - K \sin \theta \cdot u + K \cos \theta \cdot v,$$

the rotation angle θ is so small, that $\sin \theta \approx \theta$ and $\cos \theta \approx 1$. Writing $K = 1 + m$, with m the scale distortion, m is also a small number, and we



obtain

$$\begin{aligned}x &= x_0 + (1 + m)u + (1 + m)\theta v \approx x_0 + u + mu + \theta v, \\y &= y_0 - (1 + m)\theta u + (1 + m)v \approx y_0 - \theta u + v + mv,\end{aligned}$$

as a matrix equation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} m & \theta \\ -\theta & m \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix},$$

an elegant equation in which the second and third terms on the right-hand side are *small*, because they contain only the small transformation parameters x_0 , y_0 , m , and θ . Thus, the co-ordinate differences $x - u$ and $y - v$ are also small, as we observed already above.



3.9 Map projections and height systems in a three-dimensional world

Although the real Earth and her gravity field are three-dimensional phenomena, which we can describe and handle correctly only in three dimensions, nevertheless, means of description are very widely used that are based on “two-plus-one-dimensional” thinking. For this are used map projections and height systems, which together describe the world by means of 2D + 1D co-ordinates (x, y, H) .

Although we have here three co-ordinates, one cannot speak of genuinely three-dimensional co-ordinates because, on the one hand, (x, y) and on the other, H , are not comparable.

Among ordinary people — and even among land surveyors — there is a conceptual model of a “shoebox world”: rectangular, the sides oriented in the northern and eastern directions, and the height co-ordinate being simply the distance from the bottom of the shoebox, “sea level”.

It would be easy to be judgemental about this way of thinking. Remember, however, that *in a small area, the shoebox model is* — may well be — *an acceptable approximation*. For example, inside cities, plan maps and rectangular co-ordinates are used without bad repercussions. The question of the acceptability of this approximation demands careful analysis.

If the (x, y, H) representation is acceptable, then it is simpler, as a representation of location and height, than the Earth’s true geometry, true locations with their rectangular geocentric co-ordinates and true



heights with their geopotential numbers. However, in the general case it is actually a more complicated way of representing location and height: the complexity of map projections and the gravity field pervades all co-ordinates thus defined. Errors caused by misconceptions are easy to make.

Therefore in precise scientific work in geodesy, *geocentric, three-dimensional co-ordinates* and *geopotential numbers* should always be used. “Plane co-ordinates” — more precisely, map-projection co-ordinates — and metric heights should always be seen as *derived quantities*, on the basis of which no precise computations should be attempted.

See the following diagram, in which the signs “ \longleftrightarrow ” designate operations used:

$$(X, Y, Z) \xleftrightarrow{\text{reference ellipsoid}} (\varphi, \lambda, h) \begin{cases} (\varphi, \lambda) \xleftrightarrow{\text{map projection}} (x, y) \\ h \xleftrightarrow{\text{geoid model}} H \end{cases} \quad (3.8)$$

Reference ellipsoid A co-ordinate conversion between rectangular and geodetic co-ordinates is a mathematical, exact operation. The choice of reference ellipsoid is, however, arbitrary. Today, [GRS80](#) is the standard. In Finland, however, historically the Hayford or International Ellipsoid of 1924 has also been used.

Map projection A mathematical, exact operation. Many alternatives are on offer.

Geoid model The height type may be orthometric, normal (or variants of those two) or dynamic. A *geoid model* (or similar) is always needed, see section [4.1](#).

In diagram [3.8](#) on the left are the more abstract quantities, whereas on the right are the more concrete quantities, closer to daily life.



3.10 The time co-ordinate

In geodesy we use, in addition to co-ordinates of place, other co-ordinates as well. The first of these is *time*. Time describes the changes happening in the Earth, the research into which belongs to the field of *geodynamics*, on which more in section [18.1](#). With the measurement precision of modern geodesy, the Earth *lives* and *changes* continuously:



- Due to the solid-Earth tide, the ground below our feet moves periodically, twice a day, up and down — even in Finland a couple of decimetres. We do not notice this of course, because a more stable reference point is lacking: everything around us moves up and down with us.
- Because of plate tectonics, all continental plates move evenly. The velocity of motion is of the order of a few centimetres per year, and can be precisely monitored with GNSS technology.
- The rotation of the Earth is irregular. With space geodetic observation techniques it is possible to follow the variations in the direction of the Earth's rotation axis, both relative to the solid Earth — polar motion — and relative to the celestial sphere — precession and nutation — and variations in rotation rate — LoD, length of day. These phenomena together are called “Earth orientation parameters” (EOP). vuorokauden pituus
- In Fennoscandia, Canada and elsewhere, the Earth's surface is rising slowly after the last ice age, *glacial isostatic adjustment* (GIA).
- There are other, more local motions too, some of them caused by human activity.



Self-test questions

1. How are map projections classified based on what they distort and what they preserve?
2. What is the main distinction between “traditional” and “modern” co-ordinate reference frames?
3. What are the differences between the old KKJ datum and the new EUREF-FIN datum for the territory of Finland?
4. Name and describe geodetic plane co-ordinates and geodetic polar co-ordinates. How do you convert between them?
5. Describe the geodetic forward problem in the plane, and its solution.
6. Describe the geodetic inverse problem in the plane, and its solution.
7. Describe the Helmert transformation in the plane. How many free parameters does the transformation contain? Describe them.
8. What is a datum, and how is one established?



9. Describe briefly the various phenomena that geodynamics studies.



Exercise 3–1: Distances

We find the co-ordinates of two points in Finland: the GNSS fundamental station Metsähovi (METS), and the GNSS station at the Sodankylä Geophysical Observatory (SODA). The EUREF data centre (ROB, EUREF Permanent GNSS Network) offers the following approximate ITRF co-ordinates:

Station		X (m)	Y (m)	Z (m)
METS	1	2892571.00	1311843.28	5512634.01
SODA	2	2200147.00	1091638.20	5866870.60

From this, we

1. compute the latitude and longitude φ and λ , again using the NLS online service [Finnish National Land Survey, Paikkatietoikkuna](#). The software asks for 3D Cartesian (rectangular) co-ordinates in the ETRS89 system, which is not quite the same as approximate ITRF; we just pretend it is.
2. From these, we compute the distance, on the surface of the GRS80 reference ellipsoid, between the points, solving the *inverse geodetic problem*. Use the NGS web-site [NGS, Computation utilities](#) to obtain the distance s_1 .
3. Instead of a reference ellipsoid, we may use a spherical approximation, with the mean Earth radius being $R = 6371.008$ km. The angular distance between METS and SODA can be computed from their latitude and longitude, using the equation¹⁷ (cosine rule on the sphere)

17

$$\cos \psi = \sin \varphi_1 \sin \varphi_2 + \cos \varphi_1 \cos \varphi_2 \cos(\lambda_2 - \lambda_1).$$

And then, the metric distance along the surface of the sphere is

$$s_2 = \psi R.$$

jänne-etäisyys

4. Alternatively, the *chord distance* in spherical approximation:

$$s_3 = 2R \sin\left(\frac{1}{2}\psi\right).$$



5. We return to the [NLS](#) web-site ([Finnish National Land Survey, Paikkatietoikkuna](#)) and convert the co-ordinates (φ, λ) to [ETRS-TM35FIN](#), a two-dimensional *map projection system*.
6. Now, using these two-dimensional map projection co-ordinates (x, y) for METS and SODA, we compute the distance between the two stations using the Pythagoras theorem in two dimensions:

$$s_4 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

7. Now, as the icing on the cake, compute the *chord* (three-dimensional) *distance*, also using Pythagoras, from the original rectangular co-ordinates:

$$s_5 = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2}.$$

All these distances are different. Some of the differences are small, some substantial. Complete the exercise by explaining where all these differences come from, and *draw a figure* with all distances marked.

Do all calculations in *millimetres* rounding accuracy.

¹⁷Smart readers will note that this equation produces, for short distances ψ , a cosine close to unity, leading to a loss of accurate digits when ψ is recovered by the arccos function.

It is possible to convert the equation to a “half-angle version” in the following way: substitute

$$\begin{aligned}\cos \psi &= 1 - 2 \sin^2 \frac{\psi}{2}, \\ \cos(\lambda_2 - \lambda_1) &= 1 - 2 \sin^2 \frac{\lambda_2 - \lambda_1}{2},\end{aligned}$$

yielding

$$\begin{aligned}1 - 2 \sin^2 \frac{\psi}{2} &= \sin \varphi_1 \sin \varphi_2 + \cos \varphi_1 \cos \varphi_2 \left(1 - 2 \sin^2 \frac{\lambda_2 - \lambda_1}{2} \right) = \\ &= (\sin \varphi_1 \sin \varphi_2 + \cos \varphi_1 \cos \varphi_2) - 2 \cos \varphi_1 \cos \varphi_2 \sin^2 \frac{\lambda_2 - \lambda_1}{2} = \\ &= \cos(\varphi_2 - \varphi_1) - 2 \cos \varphi_1 \cos \varphi_2 \sin^2 \frac{\lambda_2 - \lambda_1}{2} = \\ &= 1 - 2 \sin^2 \frac{\varphi_2 - \varphi_1}{2} - 2 \cos \varphi_1 \cos \varphi_2 \sin^2 \frac{\lambda_2 - \lambda_1}{2},\end{aligned}$$

from which

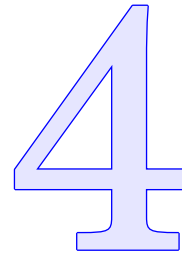
$$\sin^2 \frac{\psi}{2} = \sin^2 \frac{\varphi_2 - \varphi_1}{2} + \cos \varphi_1 \cos \varphi_2 \sin^2 \frac{\lambda_2 - \lambda_1}{2}, \quad (3.9)$$

the half-angle version of the spherical cosine rule which is well behaved for points that are close together (compared to the size of the Earth).





Height measurement and the levelling instrument



Lattamiehentie, 01260 Vantaa
Lattamiehentie, 80100 Joensuu
Vaakitsijantie, 90650 Oulu

Lattamiehentie, “Rodman Road”, *Vaakitsijantie*, “Leveller Road”, street addresses



4.1 Height, geopotential, and the geoid

Heights express the locations of points in the vertical direction, the direction of the local *gravity vector*, the vertical or plumb line.

luotiviiva

Intuitively this is based on the naïve “shoebox model” of the Earth’s figure, where height is the third co-ordinate, the straight, metric distance from the bottom of the shoebox, sea level.

The shoebox model is also called the “flat Earth approximation”: somewhere below the land surface there is a reference surface, assumed to be a plane, that coincides with mean sea level. Height is the distance in metres from this level surface.

In reality, the Earth is not flat and the reference surface is curved, even undulating. The reference surface is called the *geoid*. It is a *level* or *equipotential surface* of the Earth’s gravity field, a surface on which every point has the same value for the *geopotential*, the potential of the Earth’s gravity field.

The direction of gravity, the *plumb line*, is everywhere *perpendicular* onto this surface. The distance of a point from this surface, measured along the plumb line, is called its *orthometric height*. Thus, orthometric height has a simple geometric interpretation, and it is of course a *metric* quantity.



4.1.1 *The geopotential*

Physical geodesy is the branch of geodesy concerned with the gravity field and gravity potential of the Earth. The geopotential W could be taken as the fifth “co-ordinate” after the three co-ordinates of place X , Y , Z , and time. The geopotential describes the *energy level* of points in relation to sea level. This corresponds to the popular conception of “height”. We have a habit, mostly appropriate in daily life, of expressing height as a metric quantity. . . but what really interests us is the potential *energy* that comes with height. The effect of gravity on our daily activities is so strong, that “height determination”, the study, determination, and presentation of the geopotential, forms a large part of practical geodesy and surveying.

Water “Height”, or potential, represents *energy*. The energy may be recovered or stored — hydro-power. The energy may also be destructive — floods — for which one has to be prepared. Sewers have to have a sufficient slope so they will work.

Air The levels of equal pressure in the air follow fairly precisely the levels of the geopotential. The phenomenon is exploited in barometric height determination. An aircraft also measures its “height” using an air-pressure sensor.

Traffic Gravity affects the planning of traffic routes. Slopes may not be too steep, and limited variations in potential, or energy level, along the route are desirable. In the case of waterways, this happens automatically in a natural way.

The geopotential is closely related to gravity. Surfaces having the same value for the geopotential, *equipotential surfaces*, are what are ordinarily called “horizontal surfaces”. A freely flowing fluid — sea water, lake water, air — will settle along an equipotential surface. In the sea, the hydrostatic pressure is a constant along the equipotential surfaces of the gravity field, just like, in the atmosphere, barometric pressure — at least approximately: disturbances are due to salinity and temperature variations in water and temperature variations in air, and to the currents these cause.

suolaisuus



4.1.2 *Metric heights*

Heights are “humanised” in a similar manner to what map projections do for location data. One invents a way in which height “above sea



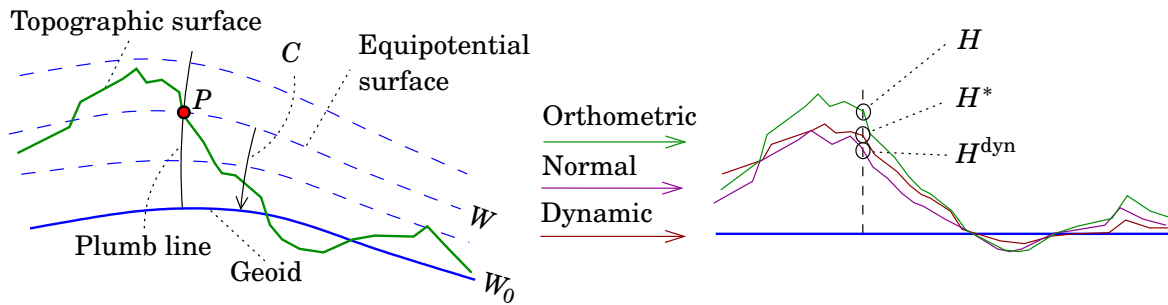


FIGURE 4.1. Different height types map geopotential numbers C in different ways to metric heights (differences exaggerated).

level” can be expressed as a *metric* quantity, as a height H from some *reference surface*, usually mean sea level. The first step is always to compute the difference between the geopotential values of the point under consideration and of mean sea level, the *geopotential number* C of the point. This number is positive from mean sea level upwards.

Unfortunately, just like with map projections, there is no solution that is in all respects satisfactory. Something is always distorted.

In the same way as with map projections, there are also different *height types*, like

- orthometric height H
- normal height H^*
- dynamic height H^{dyn} .

They all have their good and bad properties.

Earlier, in the section on geometric co-ordinates (section 2.8), we became acquainted with co-ordinates bound to the reference ellipsoid, of which one was the *height from the reference ellipsoid*, h . This co-ordinate describes the location of a point in a vertical direction and is in a way also the height of the point. It is nevertheless reckoned from the *reference ellipsoid*, a surface that is not physically useable as a reference surface in daily life. Moreover, it does not describe the energy level with respect to sea level, like H , H^* , and H^{dyn} (and of course geopotential numbers C) do.

If all this appears at this stage difficult and theoretical, it may be worthwhile to come back to this section later in the course and read it again.



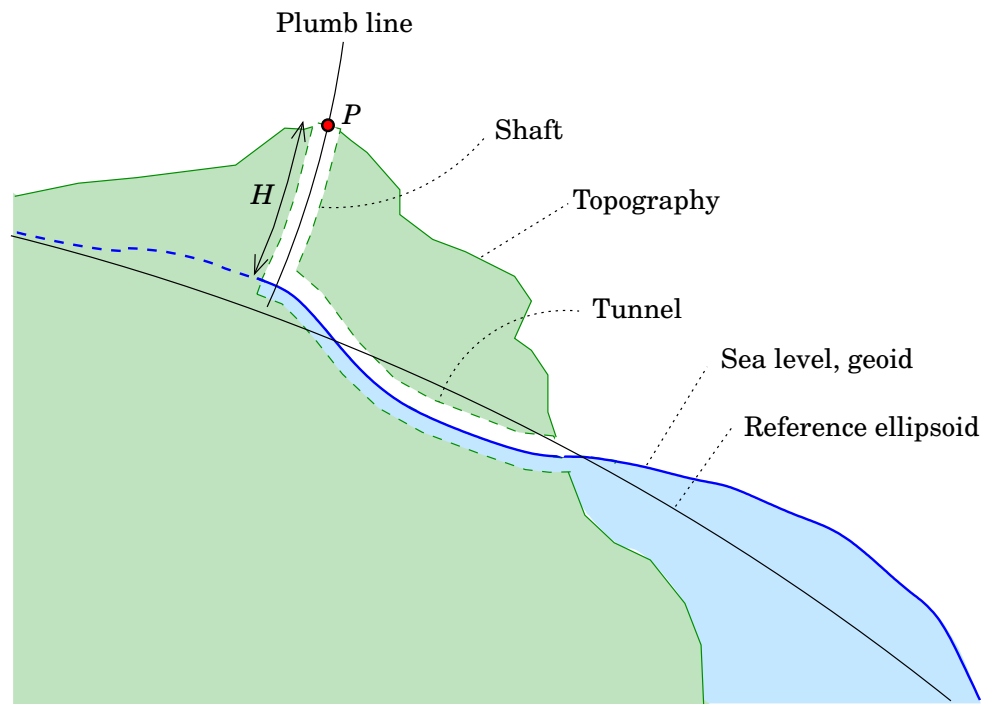


FIGURE 4.2. *Orthometric heights* are metric distances from the *geoid*, the water surface that would form if sea water could freely move under the topography in an imaginary tunnel network. In that case, orthometric heights H could be directly measured along the *plumb line* through a shaft like the one depicted.



4.2 Orthometric height

Orthometric (Greek “correctly measured”) heights H correspond the most precisely to our concept of “height above sea level”. They are in principle just metric heights above the *geoid*. The *geoid* is that equipotential surface of the gravity field which on average is on the same level as mean sea level. In other words, mean sea level continued under the land masses.

If we could excavate a network of tunnels under the continents (figure 4.2) on the level of the sea surface, the water would spread throughout the network in such a way, that its surface would be a physical realisation of the *geoid*. A point’s orthometric height would be its *distance from this fluid surface*. This direct physical interpretation is the reason why many geophysicists, and many countries — among them Finland until 2007 — have chosen to use an orthometric height system.

Building a tunnel network like the one depicted is of course not practical. Inland, the *geoid* is realised *computationally*, by carrying out the



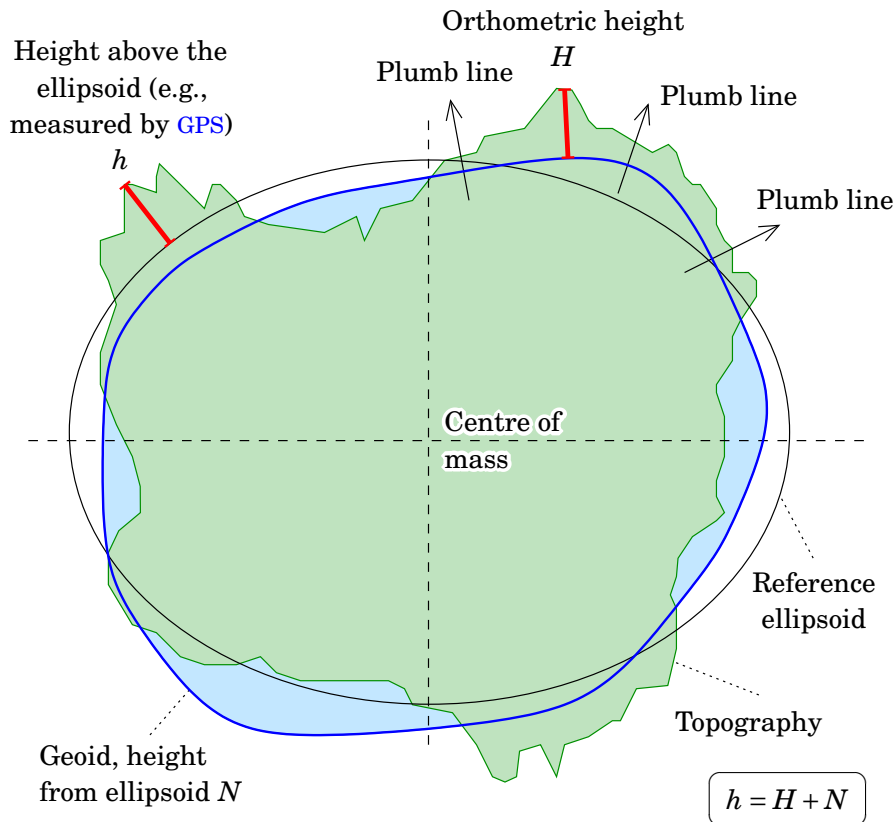


FIGURE 4.3. Important reference surfaces and height concepts.

calculation of a height or *levelling network*, starting from a chosen coastal point or set of coastal points. Thus, we disseminate orthometric heights throughout the country, to everywhere that the levelling network extends to.

In Finland up until 2007, the official height system or *vertical datum* was **N60**, the zero point of which was Helsinki's mean sea level at the start of 1960. **N60** heights are to a good approximation orthometric. In 2007, the **N2000** height system was taken into use, with the zero level as the mean sea level according to the official Amsterdam or **NAP** (*Normaal Amsterdamse Peil*, Amsterdam Ordnance Datum) datum. **N2000** heights are *normal heights*, the definition of which differs a little from that of orthometric heights. For Finland, the difference is of little practical significance.

4.3 Height determination and levelling

4.3.1 Exotic methods for height determination

The most direct way of measuring height differences is to realise an



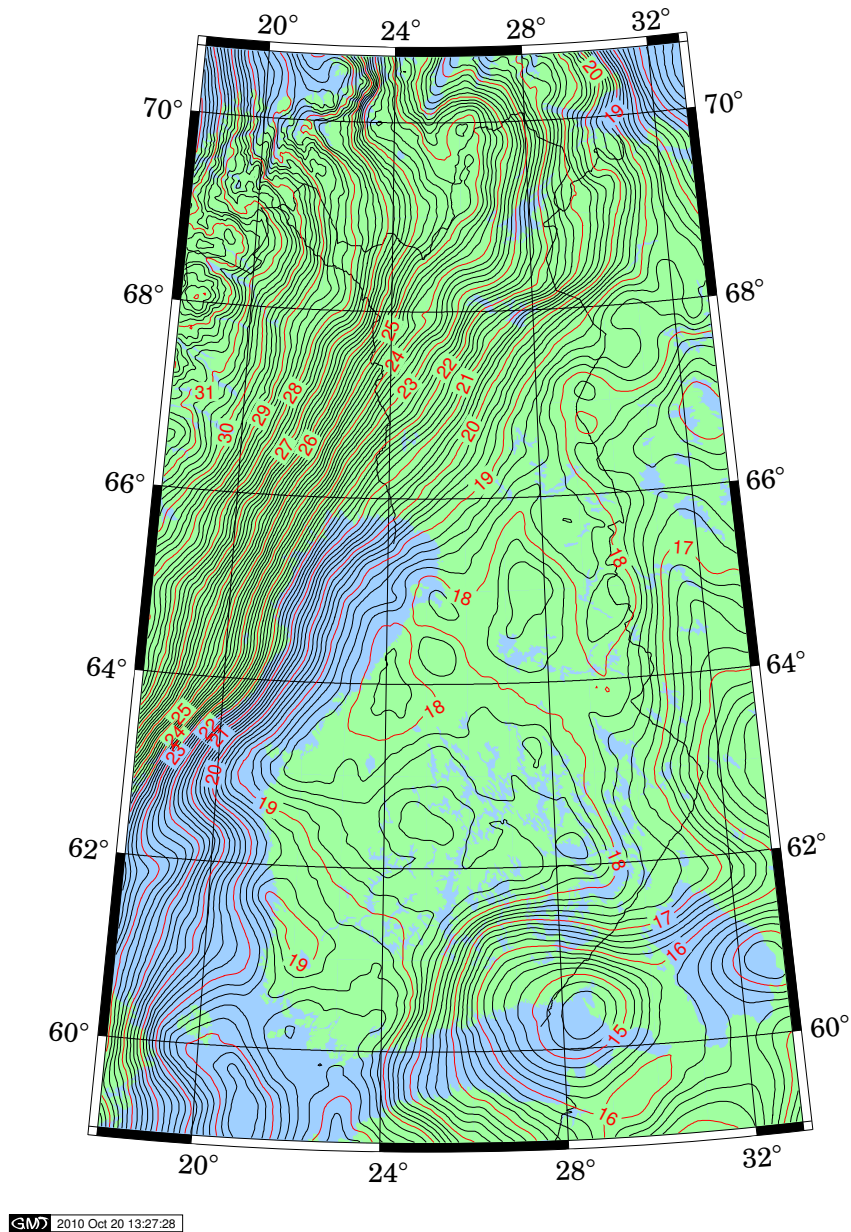


FIGURE 4.4. The Finnish geoid model **FIN2000** (data © National Land Survey of Finland). This map shows the heights of the geoid above the geocentric **GRS80** reference ellipsoid. Unit m.

equipotential surface of the gravity field by means of a *fluid surface*. This is how one can transfer geopotential values from one place to another.

- In Denmark and the Netherlands *hydrostatic levelling* has been employed, in which a long hose or tube filled with distilled water is used to transfer the level of the geopotential (the “height”) from one island to another or between island and mainland. The dis-



tances over which measurements have been done have been tens of kilometres.

- By using *water gauges* in interior waters, heights may also be transferred hydrostatically. As in the water-hose technique, here the air pressure difference between the terminals must also be taken into account, as well as the effects of wind and currents. The method, which works best under an ice cover, has been tested, for example, in the Netherlands (IJsselmeer, Rijkswaterstaat 1996–97, [Reijnoudt, 1996](#)) and Finland. vesiasteikko
- *Barometers* have also been traditionally used for measuring height differences. A careful procedure taking into account natural air-pressure variations due to weather will yield a best-case accuracy of about one metre. See [Heiskanen and Härmälä \(1963\)](#) pages 84–87. ilmapuntari

The geophysical modelling of sea currents has been attempted, for example, in the Åland Sea.

Trigonometric traverse levelling must also be mentioned here, see [figure 6.8 \(Takalo, 1995\)](#). linjavaaitus

A certain *hi-tech* method for measuring potential differences uses precise atomic clocks and the slowing of clocks predicted by general relativity theory. At the time of writing there exist *optical lattice clocks*, atomic clocks operating at optical frequencies, which should have the relative precision of $1 : 10^{18}$ required for one-centimetre precision.

Another *hi-tech* method which is already in use ([Gruber et al., 2014](#)) is the construction of precise, high-resolution geopotential models, which may be used to calculate a point's precise geopotential immediately when satellite positioning has determined the point's precise geocentric location. The already completed satellite gravity mission [GOCE](#) (2009–2013) is key to this.

4.3.2 Levelling

The already mentioned *levelling* technique is the standard method for determining heights referred to mean sea level. Levelling measures the height difference between two points, [figure 4.5](#). The distance between points measured in one measurement set-up is short. By chaining point intervals into a traverse levelling, height differences between points at great distances from each other may be determined, and thus heights mapped over wide areas.



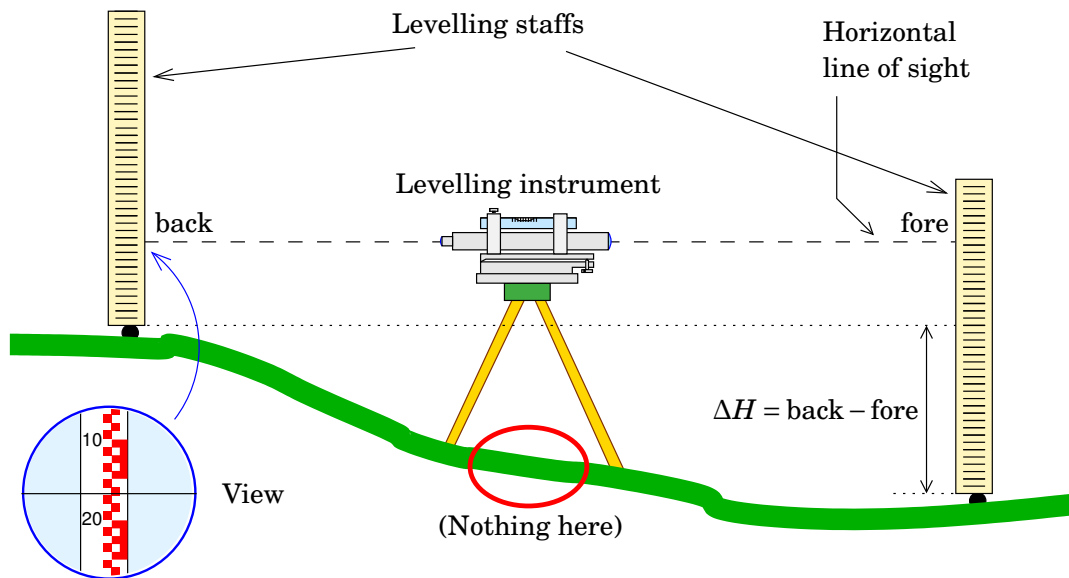


FIGURE 4.5. The geometry of levelling.

In Finland, like elsewhere, the levelling network covers the whole country and provides the opportunity to determine the heights of points in the *height system* of the network. There is a *hierarchy* for levelling networks: the precise-levelling network, which was measured and maintained in Finland by the Finnish Geodetic Institute, covers the whole country but is sparse, the loops of the network being hundreds of kilometres long. The lower-order levellings by the National Land Survey densify this network, and local actors — municipalities, builders — connect their own levelling networks to this system. This brings official heights within the reach of all users.

korkeuskäyrä The height contours appearing on topographic maps are also in the
maastokartta official system, earlier **N60**, nowadays **N2000**.



4.3.3 Creating a height system

The height differences ΔH provided by geometric traverse levelling may be added together only *within a small area*, where local gravity is constant. In larger areas, the height differences ΔH must first be converted to geopotential differences ΔC :

$$\Delta C = g \cdot \Delta H,$$

in which g is local gravity. After that, it holds for geopotential differences



that, around a closed loop,

$$\sum_{\text{closed loop}} \Delta C = 0,$$

although for raw height differences

$$\sum_{\text{closed loop}} \Delta H \neq 0!$$

In other words, while the sum of height differences $\sum_A^B \Delta H$ depends on the path chosen from A to B — and is thus not unambiguous — the sum of potential differences $\sum_A^B \Delta C$ is independent of the choice of path. Being unambiguous, the geopotential is better suited as the basis for the height *system* of an area.

4.4 The levelling instrument (“level”)

4.4.1 Construction and function

Levelling (*geometric levelling*, figure 4.5) depends on a horizontal line of sight: the optical axis of the levelling instrument’s measuring telescope, or *sight axis*, is horizontal. To achieve this, the instrument has a *spirit level*. Both the spirit level and the telescope are connected to the body of the instrument.

A traditional levelling instrument (figure 4.6) includes among other things a measuring telescope, a small circular or *bull’s-eye level* for approximate levelling, and a precise tubular level. The instrument also has a tripod and footscrews. The optical axis of a well-adjusted levelling instrument, the *sight axis* — the line defined by the crosshairs in the eyepiece — is parallel to the horizontal as defined by the level, the *horizon*.

At every instrument station, the levelling instrument must be levelled anew. Many levelling instruments have a separate *lifting screw* for precisely levelling the instrument. This is done before every forwards and every backwards measurement.

The accurate parallelity of the sight axis and the tubular level’s horizon is achieved using an *adjustment screw* when checking up the instrument.

4.4.2 Classification

Levelling instruments are classified according to their *accuracy*, *purpose of use*, and *construction*, in order of increasing accuracy, table 4.1.



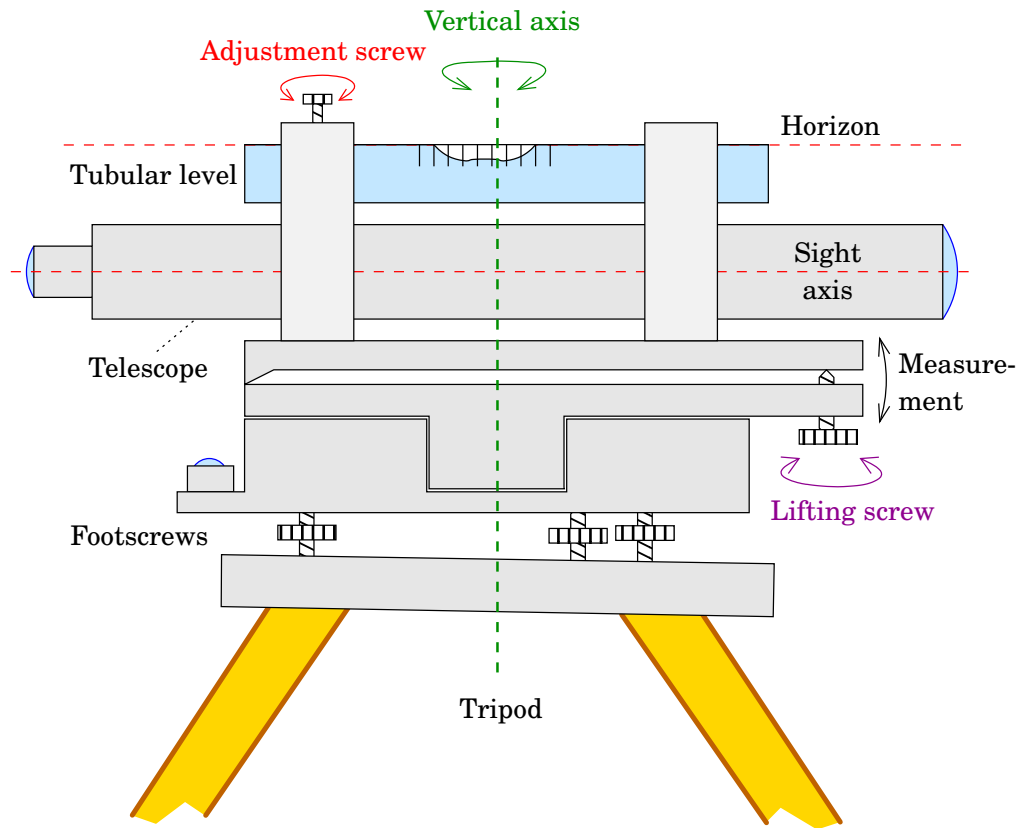


FIGURE 4.6. Levelling instrument.

We speak of *precise levelling*, if the precision on a one kilometre double run is below 0.5 mm. Instruments of this accuracy class have an optical levelling micrometer, see figure 4.17. The size and magnification of the measuring telescope grows with accuracy class, from 20× to 40× magnification. The levelling staff or rod is also chosen to correspond to the accuracy class, see table 4.2.

4.5 The measuring telescope

A *measuring telescope* is found in many optical geodetic measuring in-

TABLE 4.1. Classification of levelling instruments.

Instrument type	Levelling type
Builder's level	Construction, earthwork levelling
Engineer's level	Engineering, construction levelling
Precise level	Base network levelling
High-precision level	Precise levelling



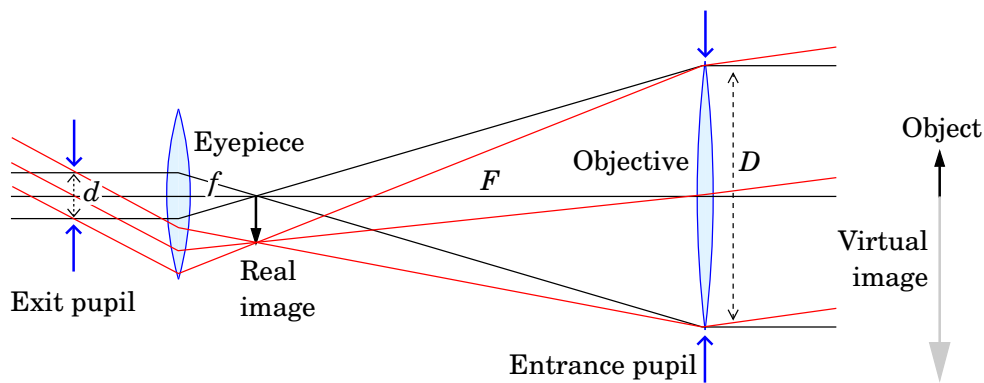


FIGURE 4.7. Telescope.

struments, like the theodolite, chapter 5, and the levelling instrument.

A *telescope* is an optical device presenting an enlarged, and thus more detailed, image of a remote object. Telescopes for use at night-time also serve to collect as much light as possible from dim objects, in order to present a bright image.

The telescope in figure 4.7 has an objective of diameter D , also called the *entrance pupil* or aperture, and a focal length F . An eyepiece is used with a focal length f . If the object is at infinity, the objective first forms a real image of it in the focal plane, which the eyepiece projects as an enlarged virtual image to infinity. The *magnification* equals the ratio $\mu = F/f$.

All the light leaving the telescope passes through the *exit pupil*, the diameter of which is $d = D/\mu$. The exit pupil is the image of the entrance pupil formed by the eyepiece. For example, for binoculars specified as 7×50 , we have $D = 50$ mm, $\mu = 7\times$, and $d \approx 7$ mm. Young people with a pupil diameter at night of 8 mm can thus place their eyes at the exit pupils and capture all incoming light.

The tasks of the measuring telescope are to

- give a sharp image of the aiming target, the levelling staff
- form the sight axis by placing the crosshairs¹ which are in the eyepiece, onto the image of the far away levelling staff.

Both tasks demand precise *focusing*, the movement of lenses along the

¹The crosshairs are nowadays usually engraved onto a glass plate, a *reticule*. Still in the 1930s, threads from a spider's nest web were used! See Bedini (2005a,b).

Crosshairs were apparently invented by the astrometrist William Gascoigne (1612–1644), who accidentally noticed the suitability of spider silk for this purpose. He was killed in combat in the English Civil War.

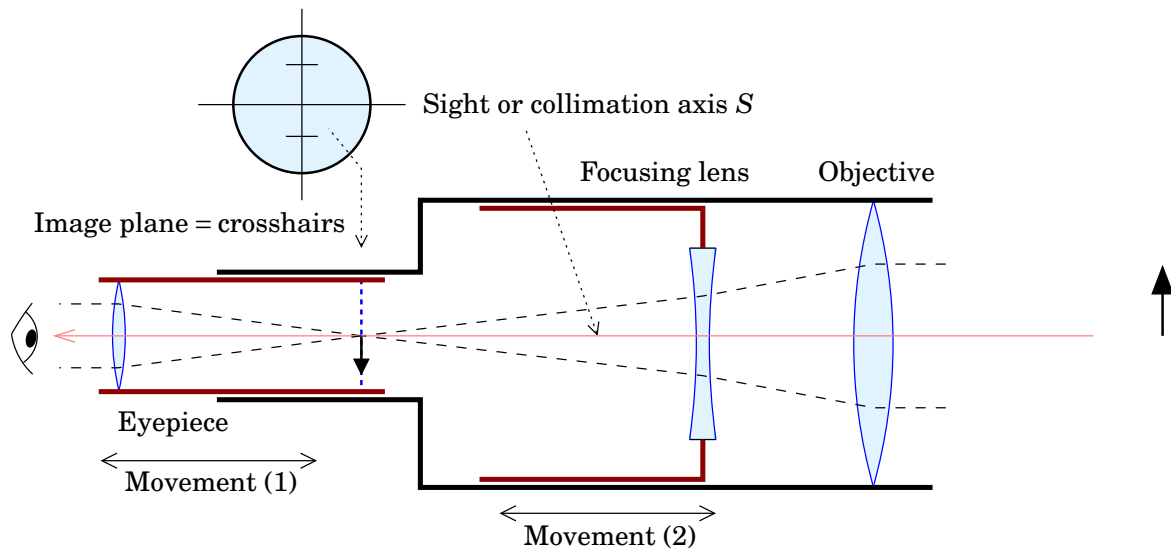


FIGURE 4.8. Measuring telescope and focusing. The object of study is to the right (at a great distance!), the observer's eye to the left. A measuring telescope is used in many optical geodetic measuring instruments, such as the theodolite, chapter 5, and the levelling instrument.

telescope axis, in order to achieve a sharp image of both the object and the crosshairs. See figure 4.8.

Focusing is done as follows, usually by turning actuator rings or screws on the telescope:

1. The eyepiece is turned so that the image of the crosshairs becomes sharp.
2. The focusing element of the instrument is turned so that the image of the target also appears sharp.

In this case, the image of the target formed by the objective, the focal plane of the eyepiece,² and the plane of the crosshairs all coincide.

In levelling, commonly equal distances to the front and back staffs are chosen. If this is not possible because of the terrain, one should *focus carefully* at every instrument station before every reading. If not, one may get *parallax*: the apparent direction of the telescope's optical axis will depend on the position of the observer's eye in relation to the eyepiece. Observing through a poorly focused telescope also causes eye fatigue.

²This only applies if the target is at infinity and the eye of the observer is error-free. More precisely, the eyepiece + the possible eyeglasses of the observer + the living optics in their own eye project a sharp image of crosshairs and target image onto the retina.

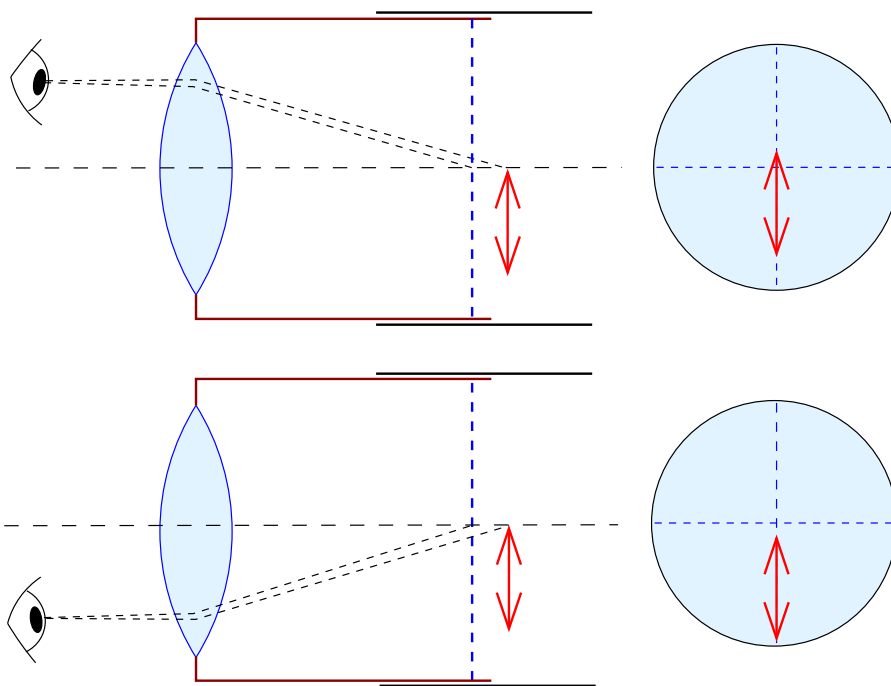


FIGURE 4.9. Parallax of a measuring telescope. If the image and crosshairs are not in the same plane, moving the eye with respect to the eyepiece will cause them to move with respect to each other.



One must always focus carefully!

Eyeglasses can be taken off if they are ordinary glasses and not cylindrical (astigmatism) or prismatic³ (heterophoria), because near- or farsightedness can be corrected by focusing the eyepiece.



4.6 The tubular level

The construction of a tubular spirit level is explained in figure 4.10. The adjustment screw seen in the picture is only used for adjusting the instrument, rarely in the field. Its purpose is to get the level axis and the sight axis of the telescope (figure 4.8) precisely parallel. The level axis is horizontal when the bubble is in the middle, a precondition for measurement.

The task of the tubular level is to help the observer bring the sight axis of the levelling instrument into the horizontal plane, perpendicular to local gravity. With it, the instrument is levelled in the measuring direction in connection with every reading.

³Prismatic glasses are relevant only when using binoculars.

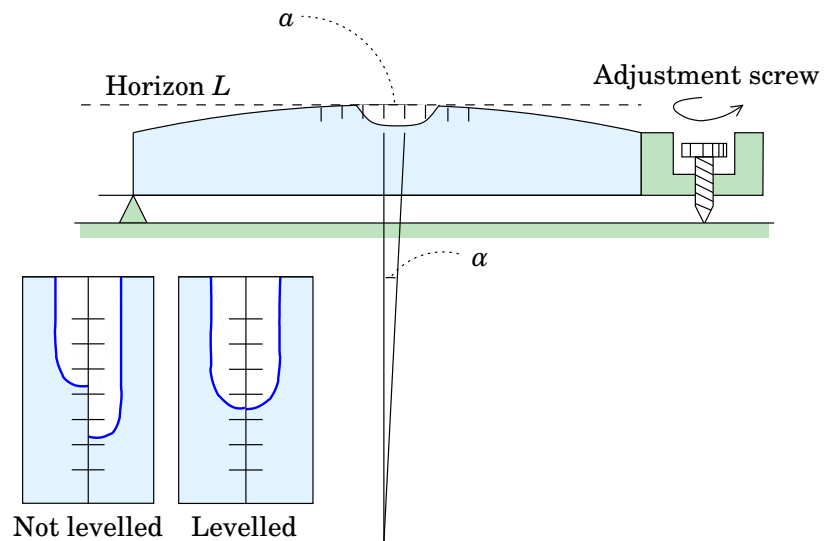


FIGURE 4.10. Tubular spirit level. Above, the construction, left below, the bubble seen through the prism system in a *coincidence level*. The angle α is the sensitivity.



The distance a is the interval between the level's graduations. Generally $a \sim 2\text{ mm}$. The *sensitivity* of the level is expressed by the angle α .

In a *coincidence level*, a system of reflective prisms is used to show the opposite heads of the bubble side by side, improving the precision of the levelling achieved.



4.7 Checking and adjusting a levelling instrument



4.7.1 Field check

At certain time intervals one must verify that the sight axis S of the levelling instrument is *parallel* with the axis, or horizon, L of the level. Due to the influence of the environment, every instrument “lives” and changes, for example, with variations in temperature and air pressure, and due to handling and wear.

A check is carried out using levelling observations (field check): the measuring distance ℓ is chosen as 25–50 m, depending on weather conditions: the measurements are best done during overcast weather.

The checking method is based on the circumstance that, when measured from A , the measurement result, that is the height difference, is correct, whereas when measured from B , the measurement result contains an error $2v$, in which v is the error caused by the difference in



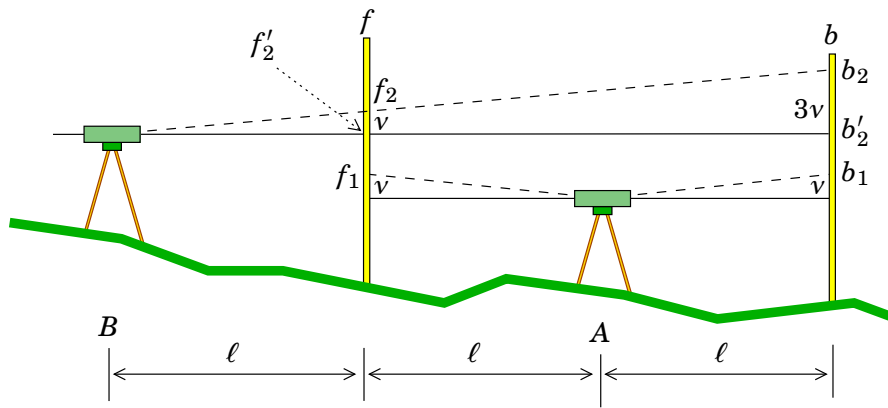


FIGURE 4.11. The geometry of a field check (Kukkamäki method).

direction between the sight axis and level horizon, at staff distance ℓ . We readily obtain

$$\begin{aligned} \Delta h_A &= (b_1 - v) - (f_1 - v) = b_1 - f_1, \\ \Delta h_B &= (b_2 - 3v) - (f_2 - v) = b_2 - f_2 - 2v. \end{aligned}$$

These height differences are *identical*. From this condition we solve for v :

$$b_1 - f_1 = b_2 - f_2 - 2v \implies v = \frac{1}{2} ((b_2 - f_2) - (b_1 - f_1)).$$

The figure shows that

$$f'_2 = f_2 - v, \quad b'_2 = b_2 - 3v.$$

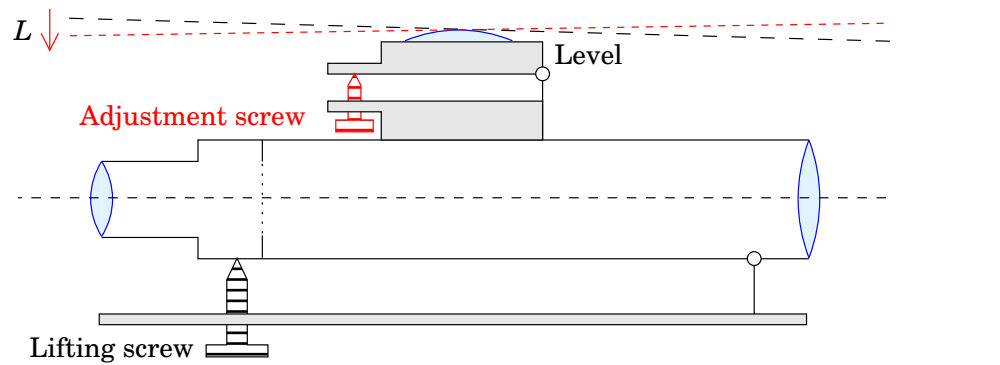
These are readings that can now be calculated. This enables the correction of the difference in direction between the sight axis and level horizon in the field, using the *adjustment screw* meant for just that.

4.7.2 Adjusting a levelling instrument

When, as a result of the field check, it is known that the staff reading is off by an amount v , one proceeds as follows to adjust the instrument:

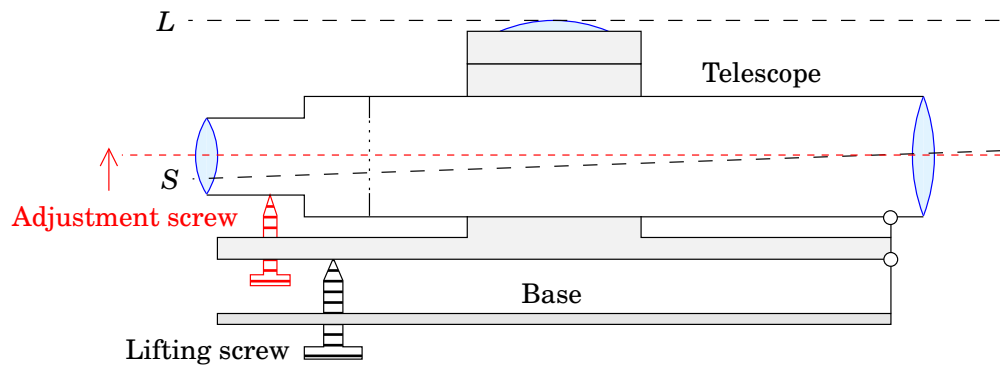
- The instrument has an adjustment screw for the *tubular level*, which tilts the level with respect to the telescope (figure 4.12a). The assembly of telescope and tubular-level is attached to the base by a *lifting screw*.
 1. Level the instrument first approximately using the footscrews, then precisely using the lifting screw, all the time keeping an eye on the tubular level.
 2. Take the staff reading f .





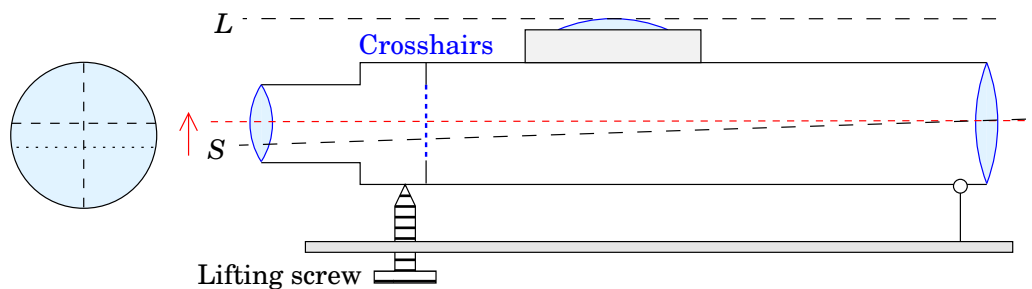
(a)

The tubular level has an adjustment screw



(b)

The telescope has an adjustment screw



(c)

The crosshairs plate can be adjusted



FIGURE 4.12. Adjusting the horizon of a levelling instrument.

3. Move, using the *lifting screw*, to the staff reading $f' = f - v$, where v comes from the field check.

Note! Do *not* use the footscrews for this, because then the instrument may also tilt in the transversal direction — the footscrews are not in line with the sight axis of the measuring telescope. Their only role is the approximate levelling of the instrument using the bull's-eye level.



4. Now the bubble in the tubular level is no longer in the middle. Use the *adjustment screw* for the level to get the bubble in the middle again. After that, $L \parallel S$.
- The instrument has an adjustment screw for the *measuring telescope*: the telescope tilts with respect to the level (figure 4.12b). The assembly of telescope and tubular level is again attached to the base with a *lifting screw*.

An equivalent, popular technical solution is an adjustment screw that shifts the crosshairs plate in the vertical direction within the image plane (figure 4.12c).

1. Level the instrument.
2. Take a staff reading f .
3. Move, using the telescope's (or crosshairs glass plate's) *adjustment screw*, to the new staff reading $f' = f - v$.
4. The tubular-level bubble is still in the middle!



4.8 Self-levelling instrument

Self-levelling or automatic levelling instruments use *gravity* to obtain a horizontal sight direction.

Early models used gravity to level the whole measuring telescope according to the pendulum principle (figure 4.13). It is clear that such an instrument is difficult to use under field conditions due to disturbances caused by wind and observer proximity.

In today's instruments, only a prism or mirror guiding the light is used as a pendulum. It is suspended inside the telescope: a *pendulum compensator* is built into the telescope. In order to function, the instrument must already be approximately levelled, for example with the aid of a bull's-eye level.

The principle of the self-levelling levelling instrument is shown in figure 4.14.

The principle of operation of the pendulum compensator is explained conceptually in figure 4.15, in which the path of the light beam has been folded open. The figure shows the situation in a co-ordinate frame attached to the measuring telescope.

A small tilt of the telescope away from the horizontal causes a tilt of the incoming light beam of α . In order for the image to remain in the same



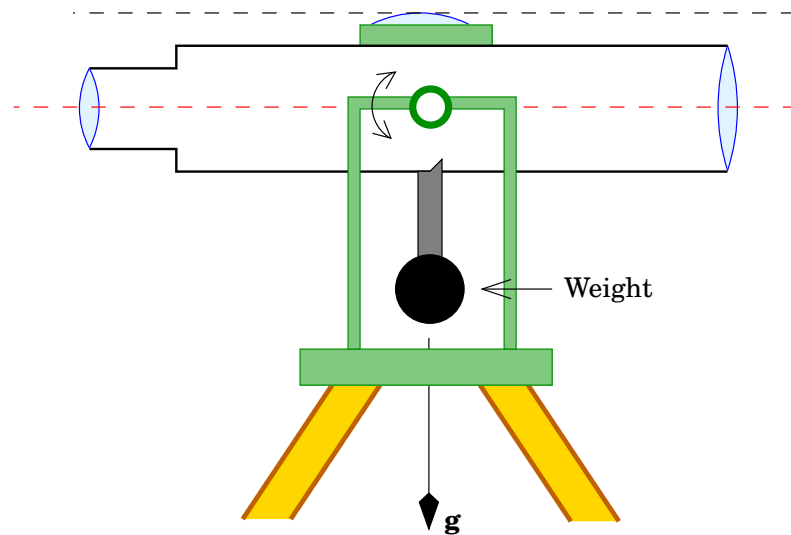


FIGURE 4.13. Principle of operation of an early self-levelling levelling instrument. An instrument built in this way is not very practical.

place in the image plane of the telescope, the compensator bends the light beam by an amount 2α , assuming that the distance between the objective and compensator s is equal to that between the compensator and image plane, in other words, the compensator is precisely in the middle between them. A freely suspended mirror turns, relative to the telescope, by an amount α , and the direction of the reflected beam changes by an amount 2α , just as intended. See [Kahmen and Faig \(1988\)](#) pages 334–336.

The strength of compensator instruments is their ease of use. However, in the early days there were technical issues, such as magnetism of the mirror suspension ([Kukkamäki and Lehmuskoski, 1984](#)). These problems

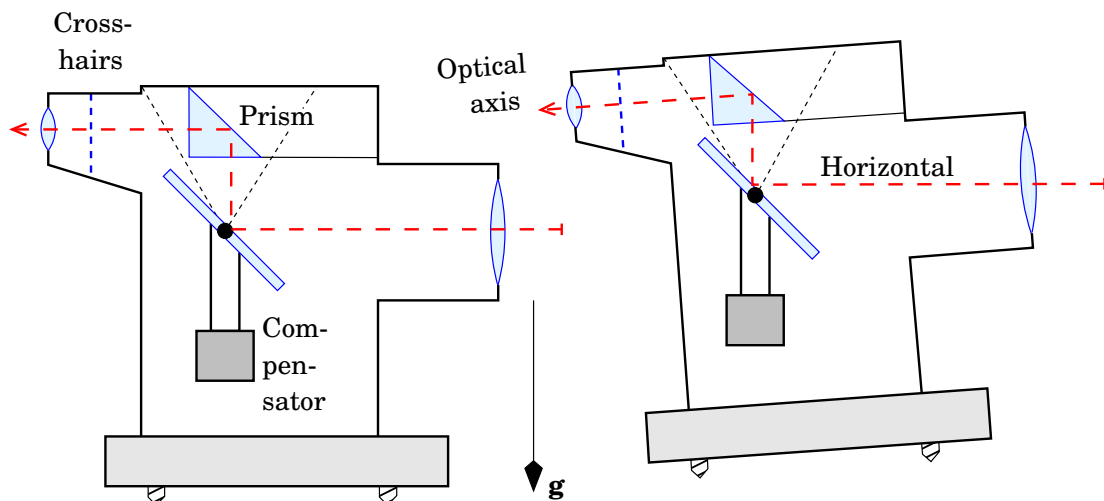


FIGURE 4.14. A modern self-levelling levelling instrument.



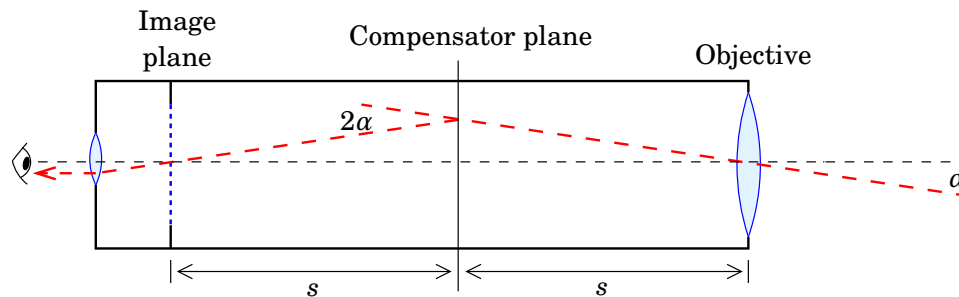


FIGURE 4.15. Principle of operation of the pendulum compensator.

appear to have been solved.

4.9 Digital levelling instrument

Nowadays digital levelling instruments are generally used, as the automation of measurement that they bring with them saves costs. The measurements are stored directly into the instrument's memory, and the necessary checks are done immediately.

The staff used together with a digital levelling instrument bears a *bar-code* that, at least in principle, is no different than that found on merchandise. Thanks to it, height values can be machine read using the instrument's **CCD** sensor and processor system. As a side result, a crude staff distance is also obtained, as well as a warning signal if the fore and back distances are too different.

viivakoodi

Unlike a traditional levelling staff, on which the measurement always uses the edges of no more than two graduation lines, with a digital or bar-code staff always a *whole area* is used, of size 30cm in the case of the Zeiss (Trimble) DiNi12 instrument. This has both advantages and disadvantages.

Advantage In the measurements, some sort of average over the edges of many staff graduations is used. Therefore, the accuracy of manufacture of the graduations and the accuracy of calibration of the staff are less critical. The staffs last longer in a useable state.

Disadvantages

- The whole interval on the staff that is being used should be visible. In forested areas, this may cause problems.
- The calibration must always be done as a *system calibration*: the instrument and staff are calibrated as a “black box”, together. On the other hand, by combining the calibration



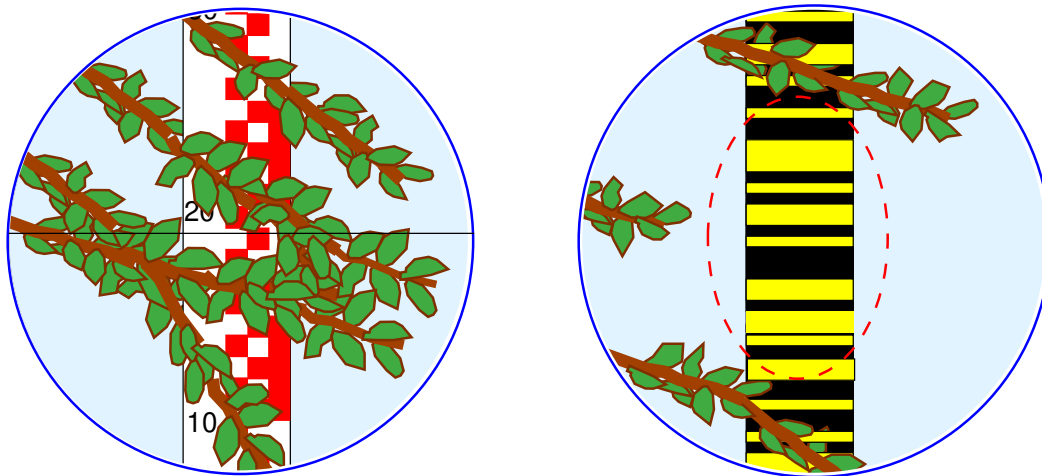


FIGURE 4.16. With a traditional levelling staff, one may also read through foliage. A bar-code staff causes confusion more easily.



of staff *graduations* and system calibration, one may reconstruct how the instrument weights the graduations it uses, and thus get the black box ajar.

The digital method of levelling is in widespread use, even in precise levelling, and has been the subject of active research (Takalo et al., 2001; Takalo and Rouhiainen, 2004).



4.10 The levelling staff

The *levelling staff* or *levelling rod* is a scale, in Finland a metric one, with which the height differences between two points are measured using a levelling instrument. There are many alternatives for the graduation of the scale, figure 4.17:

jaotus

“E” graduation Simplest of all. Its weakness is that a bright white small square appears a little larger than a dark red small square — the Helmholtz⁴ brightness illusion. When one interpolates the millimetres visually, small systematic effects easily occur.

4

Chessboard graduation Here, the above mentioned weakness has been corrected.

Line graduation Used in precise levelling. The line interval is 10 or 5 mm.

⁴Hermann Ludwig Ferdinand von Helmholtz (1821–1894) was a German physician and physicist and a student of vision.

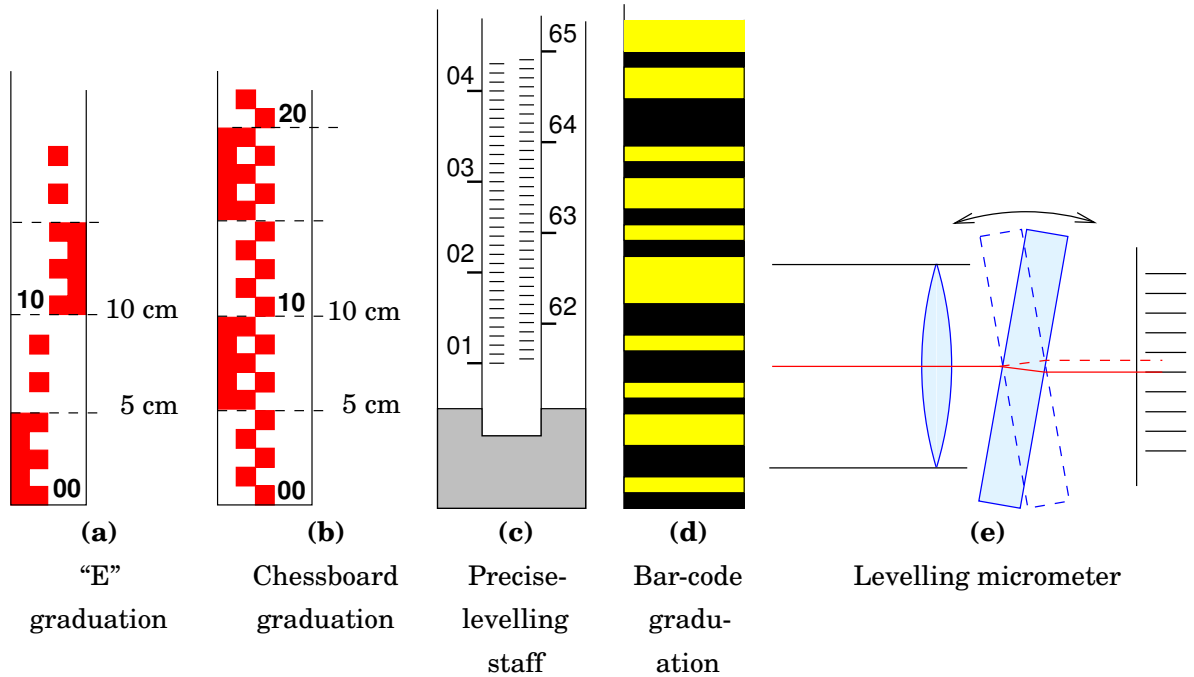


FIGURE 4.17. Graduation alternatives for the staff scale: “E” graduation, chessboard graduation, precise-levelling staff, bar-code graduation. On the right, an optical levelling micrometer to be placed in front of the objective.

Instruments for precise levelling have an *optical levelling micrometer*, with which a better accuracy is obtained than by visually interpolating the graduations. The micrometer in front of the telescope objective contains a turnable glass plate, with which one places the horizontal crosshair on one of the graduation lines. The reading from the line gives the crude value, the turning scale of the micrometer plate completes it to a precise reading.

Bar-code graduation To be used with digital levelling instruments.

The staffs are often manufactured from wood, better ones from aluminium. In precise levelling, an “invar staff” is used, on which the graduations are painted on an invar tape, which is mounted on a wooden or aluminium-alloy frame, and kept under tension by a spring.⁵ For precise-levelling staffs, a graduation of half a centimetre is used — the

⁵The force of the spring is known — 200 N — and its effect on the length of the invar tape is computable. The coefficient of thermal expansion of invar is close to zero, and metal, unlike wood, is insensitive to moisture.



TABLE 4.2. Classification of levelling staffs.

Type of staff	Length (m)	Other
Simple	3–5	Foldable or telescoping, wood or aluminium, two to four parts, “E” graduation
Base-network levelling staff. In German <i>Zweiskalenlatte</i>	3	Stiff, wooden, chessboard or line graduation, double scale or scale on both sides (reversal staff), bull’s-eye level
Precise-levelling staff	3	Wooden or aluminium frame, invar tape on which double scale
Bar-code staff	3	Aluminium, invar tape. Used with digital levelling instrument
Industrial staff		Double-scale staff of extreme precision
Self-calculating staff		See subsection 4.11.2

“staff unit” is 5 mm. There are two graduation scales, slightly shifted with respect to each other, as a double-check on reading mistakes and to randomise reading errors. Sometimes the scales are on different sides of the staff: a “reversion staff”.

kääntölatta

Upmarket levelling staffs always sport a built-in bull’s-eye level. The staff must be precisely vertical at the moment of reading!

Staffs are classified according to their purpose of use, see table 4.2, which describes Central European practice.

A *self-calculating* staff or area-levelling staff has the following properties, see figure 4.21:

- The graduation increases from above going downwards.
- At the lower end is an adjustable foot, which may be pulled out on a known point, so that the correct fraction of a metre becomes visible. After that, one goes into the terrain to map any number of point heights.

In high-precision work, the staffs need to be regularly calibrated, at least before and after the field season. Digital levelling instruments and the bar-code staffs they use should be calibrated *as a system*.

kiintopiste

Suitable staff supports may be benchmarks, temporary wooden spikes hammered into the ground, and the like. Standard staff supports are

kiila

depicted in figure 4.18. A levelling spike is used when the soil is soft,



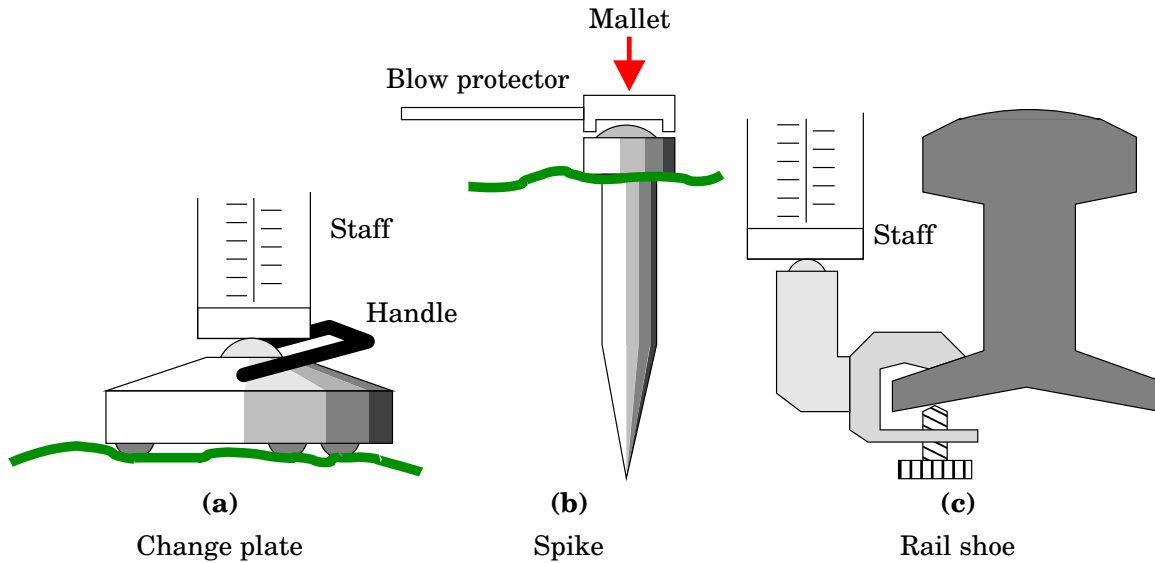


FIGURE 4.18. Various temporary levelling-staff supports: change plate (German: *Frosch* — “frog” —), levelling spike, rail shoe.

whereas a change plate is used on a hard substrate. A rail shoe is used when levelling along railways. It may be left in place when a train speeds past. During the first and second precise levellings of Finland, many levelling lines ran along railways, but since, more and more, highways have been used.

kilpikonna

4.11 Levelling methods

4.11.1 Traverse levelling

Line or traverse levelling is used in base-network measurements. Its objective is to bring the official height system close to all users in the nation, to be used for example as the reference level for detail surveys.

runkomittaus

The *precise-levelling network* is the national height base network. It is realised hierarchically by successive network densifications, section 8.3.

kartoitus-
mittaus
korkeusrunko

The heights are computed for the user community in the national height system. The heights for the third and last Finnish precise levelling were still computed in the N60 system; today the national height system is N2000. All levelling network computations are however first carried out with *geopotential numbers*, from which then either orthometric (N60) or normal (N2000) heights are calculated.

Traverse levelling is done by summing the measurements between



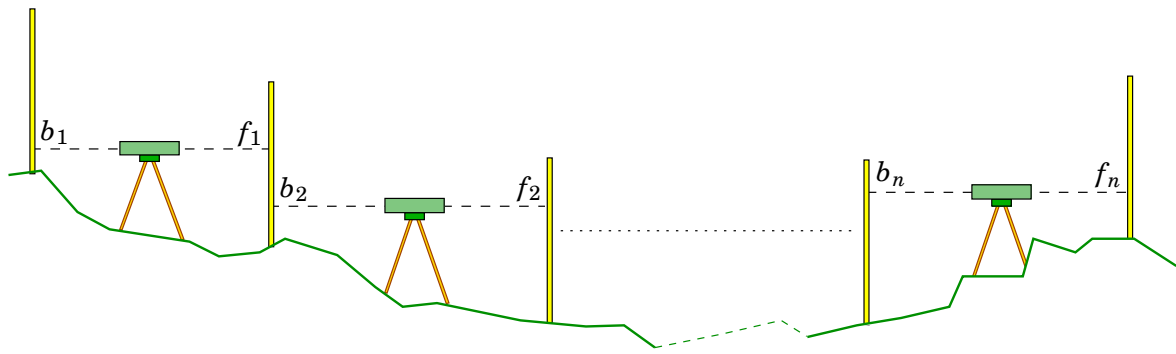


FIGURE 4.19. Traverse levelling.

⁶ successive instrument stations (figure 4.19) into a height difference:⁶

$$\Delta H = (b_1 - f_1) + (b_2 - f_2) + \dots + (b_n - f_n) = \sum_{i=1}^n (b_i - f_i).$$

piikki

- The line runs from known point to known point. Sometimes this is impossible, and we speak of a “spike”. In this case we measure carefully in both directions: *check*.
- A well-planned levelling network incorporates all measurements and points into closed loops: *check*.
- In order to minimise the impact of weather- and instrument-related errors, one makes the fore and back staff distances as equal as possible: $\ell_{\text{back}} \approx \ell_{\text{fore}}$. The staff distances also may not be too long, for example in precise levelling, 50m, however depending on weather conditions. In overcast weather one can use longer staff distances; in sunny weather with strong shimmer, staff distances have to be shortened (Kääriäinen, 1966).
- If a levelling line runs along a railway or highway, *safety arrangements* must be in order.



4.11.2 Area levelling

pintavaaitus

Area levelling is explained in figure 4.20. With this method, the height situation of a whole area is mapped using one known starting point. Among the points to be measured, there must be at least one other known point: *check*.

⁶The formula is exact only within a small area where gravity is constant. See section 4.1.

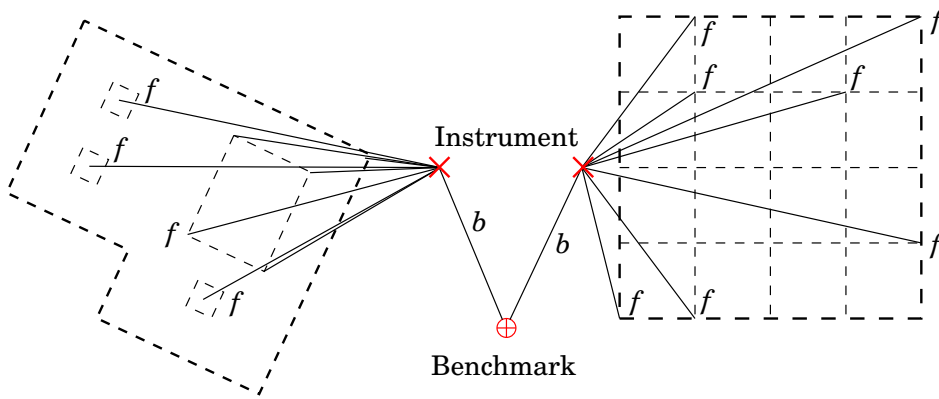


FIGURE 4.20. Area levelling. One backwards (b) observation, many forwards (f) observations.

Area levelling is an ideal opportunity to use a *self-calculating levelling staff*, see figure 4.21. At the starting point, the leg of the staff is pulled out until in the levelling instrument one sees the correct decimal fraction of a metre: if the height of the point is known to be 12.75 m, the leg is pulled out until the observer sees in the telescope the number $\langle n \rangle.750$, with $\langle n \rangle$ the number of whole metres. The leg is screwed tight, and the staff is transferred to the first observation point. The observer must keep track of the whole metres; the fractions show directly in the telescope.

The results of area levelling are needed and used for

- mapping the height situation on a building site before starting to construct the foundation
- creating digital terrain models locally and at high resolution
- calculating earthwork volumes to be moved.

maamassa

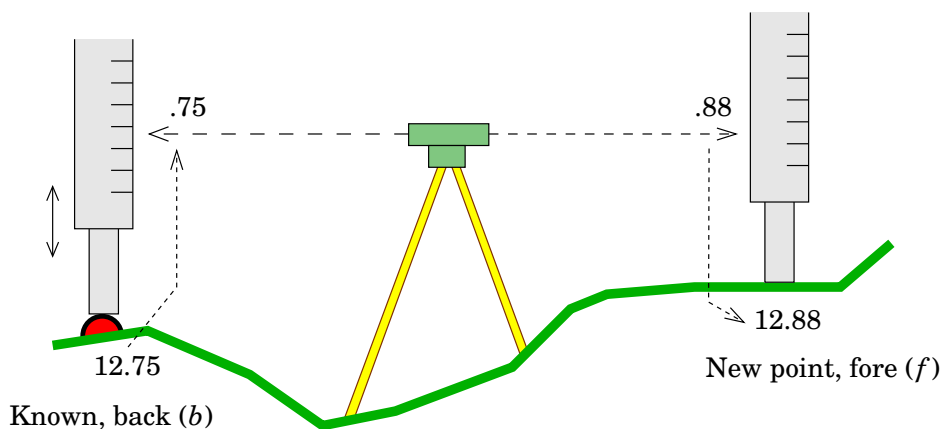


FIGURE 4.21. Self-calculating levelling staff. How the leg of the staff is set to the right length.



For the area levelling to be useful, the approximate horizontal locations of the measured points must have been established by some other means.

Proper *checks* are important: in addition to the starting point, other points with known heights should be included in the measurement. An erroneous starting height would propagate in full to the whole area, which would be, in sewer construction for example, a fatal and expensive mistake.



4.11.3 Technical levelling

Installation measurement in industry and construction belongs to the field of *engineering geodesy*.

- Extreme case: the [CERN Large Hadron Collider \(LHC\)](#) in Geneva, circumference 27km, precision on the level of millimetres ([Schrock, 2014](#)).
- Paper machines, shipyards.
- Road construction, bridges, tunnels, railways.

Deformation measurement and monitoring For this purpose, levelling is only one of many methods: deformations are usually three-dimensional. See section [18.1](#).

- Extreme case: post-glacial land uplift.
- Deformations caused by the pumping of natural gas, petroleum, or drinking or irrigation water, anthropogenic land subsidence. Venice.
- Deformations of dams, reservoirs, bridges, high steel.
- Old buildings and archeological artefacts, the tower of Pisa and the good ship *Vasa* ([Vasamuseet, Deformation Monitoring](#)).

maan
vajoaminen
vesiallas



4.11.4 Levelling of profiles and cross-sections

Profiles and cross-sections are measured in connection with construction projects, especially of roads, railways and canals.

A profile is a longitudinal section of the Earth's surface, usually along a planned trajectory of a road, railway or waterway.⁷

The work starts with setting out the trajectory into the terrain.

7
maastoon
merkintä

⁷In this case, a parallel line is used.



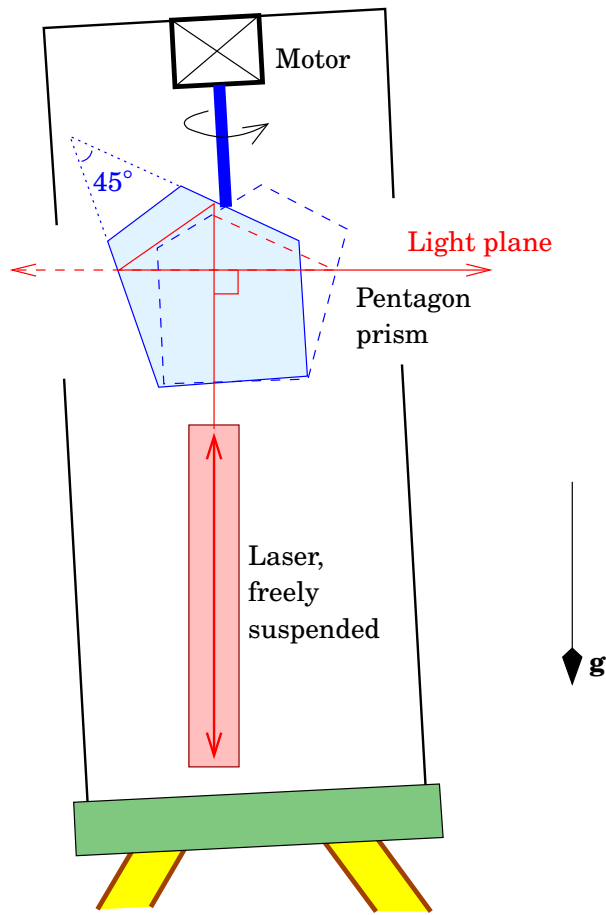


FIGURE 4.22. Principle of operation of a laser level.

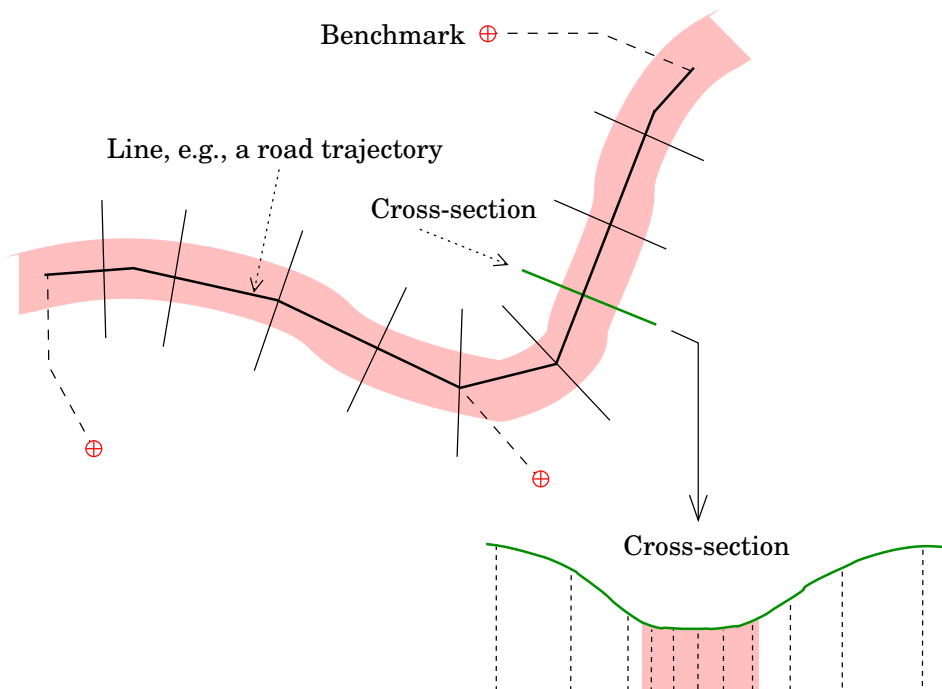


FIGURE 4.23. A profile and cross-sections.



Markers are placed at 25, 50 or 100 metre intervals, as well as at sharp terrain features. The markers are numbered: the number signs are placed outside the work area. The heights are measured with a builder's level.

At least two tie measurements to benchmarks of the height base network in the area are made, at the start and end points of the profile. If the use of two benchmarks is not possible, the loop must be closed as a double-check by measuring back and forth. Closing errors are corrected in proportion to the levelled distances.

poikkileikkaus **A cross-section** is a transversal section of the Earth's surface, perpendicular to the trajectory of the profile. At corners, the angle is split equally. Cross-sections are typically 20–50m long. The purpose of cross-sections is to support the planning work and enable the calculation of earthwork masses to be moved. The height measurement of the cross-sections is carried out in the same manner as area levelling. The point density is chosen in accordance with the _____ terrain and the intended use.



4.11.5 The laser level

lasertaso Often used in area levelling are *laser levels*, figure 4.22. Laser levels are instruments stabilised by a pendulum compensator, which projects laser light through a rotating *pentagon prism*⁸ into the environment, in order to form a horizontal plane.

These instruments are handy on building sites, where they realise a horizontal plane, which the user can make visible with a stick. Spreading sand, laying floors, or bricklaying a wall along a straight line become easier. A staff equipped with a suitable sliding receiving device will directly give the height of the point on which it stands.

A laser level can also usually be quickly reconfigured to produce a vertical plane of laser light. This is also useful on building sites.



Self-test questions

1. What is the reference surface for orthometric heights?

⁸Inside the prism are two reflective surfaces with an angle between them of 45°. The angle between the incoming and outgoing rays is thus always 90° or 100°.





FIGURE 4.24. Metsähovi research station from the air. Google Maps™. **Google terms of use.** Images © 2019 Google, © 2019 Maxar Technologies. Map data © 2019 Google.



2. What is the relationship between the geoid surface and deviations of the plumb line?
3. How is the *sight axis* of a levelling instrument (or a theodolite) defined?
4. How does one focus a measuring telescope?
5. What is parallax, and how does one avoid it?
6. What are the graduation alternatives for levelling staffs?
7. Name three standard levelling-staff supports.
8. In levelling, what is the role of the rodman, the person holding the staff?
9. Describe the Kukkamäki method of field testing a levelling instrument. What is the purpose of the test?
10. Describe how an automatic or self-levelling levelling instrument works. Make a drawing!
11. Describe how a laser level works.





Exercise 4–1: Heights

Earlier we gave the geographical or geodetic co-ordinates (on the [GRS80](#) reference ellipsoid) of the [GPS](#) antenna at Metsähovi research station, equation 2.5:

$$\varphi = 60^\circ 13' 2''.89046, \quad \lambda = 24^\circ 23' 43''.13336, \quad h = 94.568 \text{ m.} \quad (2.5)$$

1. Use the geoid model given on the [UNAVCO](#) web page ([UNAVCO, EGM96 geoid calculator](#)), which is good to ± 25 cm on the Finnish territory, to determine the geoid height N at Metsähovi. Alternatively use [Karney](#), which also gives the newer [EGM2008](#) model.
2. Compute the height H above “sea level”, as represented by the [EGM96](#) model. The relationship you need is given in figure 4.3.
3. From the air, Metsähovi research station looks like figure 4.24. The [METS GNSS](#) antenna is mounted on the tall steel grid mast to the right.

pääkiintopiste

The height, in the N2000 system, that is, above sea level, of the fundamental benchmark PP2000, figure 3.16, is 54.4233 m. Assume that this is also the approximate height of the terrain at the foot of the steel grid GNSS mast, ± 1 m.

Question: how tall is this steel grid mast?

4. Go to Google Maps and measure the latitude and longitude of the foot of the mast. Compare with the above values for latitude and longitude. How large are the differences? Discuss.





The theodolite



Our money's all spent, to the deuce it went!
The landlord, he looks glum,
On the tap-room wall, in a very bad scrawl,
He has chalked to us a sum.
But a glass we'll take, ere the grey dawn break,
And then saddle up and away —
Theodolite-tum, theodolite-ti, theodolite-too-ral-ay.

From the folk song “*The Old Survey*”, around 1905 ([The Institute of Australian Culture, 2012](#))

THE THEODOLITE WAS probably invented by Leonard Digges. The invention was published by his son Thomas in his land-surveying textbook *Pantometria* in 1571. The origin of the name is unclear. The first theodolites did not yet have a telescope, which was possibly not invented — or at least attempted to be patented — until 1608 by Hans Lippershey in the Netherlands.

The theodolite is an angle measurement instrument which measures horizontal and vertical angles relative to the local horizon (horizontal plane) and plumb line (direction of gravity). Modern instruments called electronic tacheometers or total stations also measure distances. Here, we shall call all of these “theodolites”.

luotiviiva



5.1 Horizontal angles and zenith angles

Theodolite measurements are always made inside the Earth's gravity field. The vertical axis of the theodolite is levelled to be in the direction of the local plumb line. Directions and direction differences may be expressed, in these natural instrument co-ordinates, as horizontal and

koje-
koordinaatit

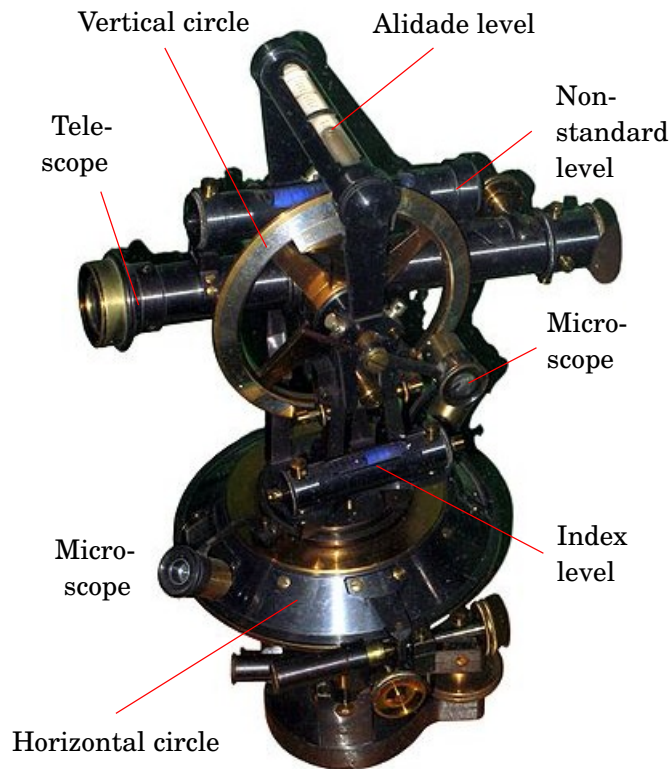


FIGURE 5.1. An old-fashioned theodolite. Note the external horizontal and vertical circles and reading microscopes. Base image [Wikimedia Commons, Exploration theodolite](#).

vertical angles.

Let there be (figure 5.2) a difference in direction, as seen from the instrument, between points A and B . The points are projected along the plumb line onto the local horizontal plane of the instrument, yielding points A' and B' . The difference in direction between points A' and B' , the angle α , is the *horizontal angle* between points A and B . Both the horizontal directions and horizontal angles are counted positive in the clockwise direction.

The angles ζ_A and ζ_B are the *vertical* or *zenith angles* of points A and B .

Horizontal angle The angle α formed by the projections OA' and OB' of the rays OA and OB in the horizontal plane.

Zenith angle The angle formed by the plumb line and the ray OA (ζ_A) or OB (ζ_B). The zenith angle is always positive.

The *plumb line* or vertical, the direction of a freely suspended plumb wire,



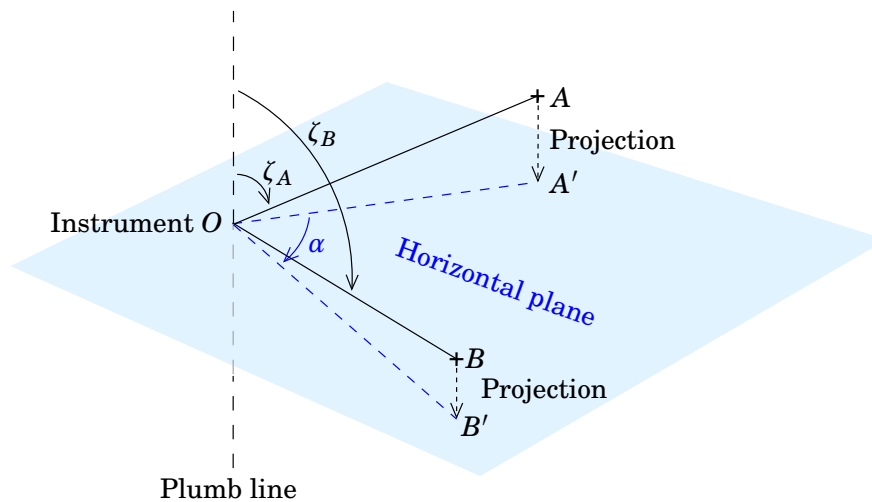


FIGURE 5.2. Horizontal angle and zenith angle.

is the local direction of the Earth's gravity vector. The plumb line points at the centre of mass of the Earth, but *only approximately*.¹

1



5.2 The axes of a theodolite

A theodolite has *three axes*, figure 5.3:

- the vertical axis or *standing axis*, which must be directed along the local plumb line or gravity vector
- the horizontal axis or *trunnion axis*, around which the telescope turns
- the sight or collimation axis, the axis of the telescope tube passing through the crosshairs of the eyepiece.

tähtäysakseli
hiusviiva-
ristikko
okulaari
vaaka- ja
pystykehä

A theodolite has two *circles*, the horizontal and vertical circle, figure 5.3.

Theoretical objective for a well-adjusted theodolite:

- H_z , V and S intersect at the same point.
- $V \perp H_z$. If not, the error is called *trunnion-axis tilt*.
- $S \perp H_z$ (but not $S \perp V$! Why not?). The error is called *collimation*

tappikaltevuus

¹Approximately, mainly because the Earth is an ellipsoid of revolution and not a sphere. The deviation of the plumb line from the direction to the Earth's centre of mass due to the flattening is at its largest as much as $11'$ at latitudes $\pm 45^\circ$.

The plumb line is also ever so slightly curved. In addition, there are local, varying deviations of the plumb line from the surface normal to the ellipsoid of revolution, of an order of magnitude ranging from a few seconds of arc in even terrain to over a minute of arc in the mountains.



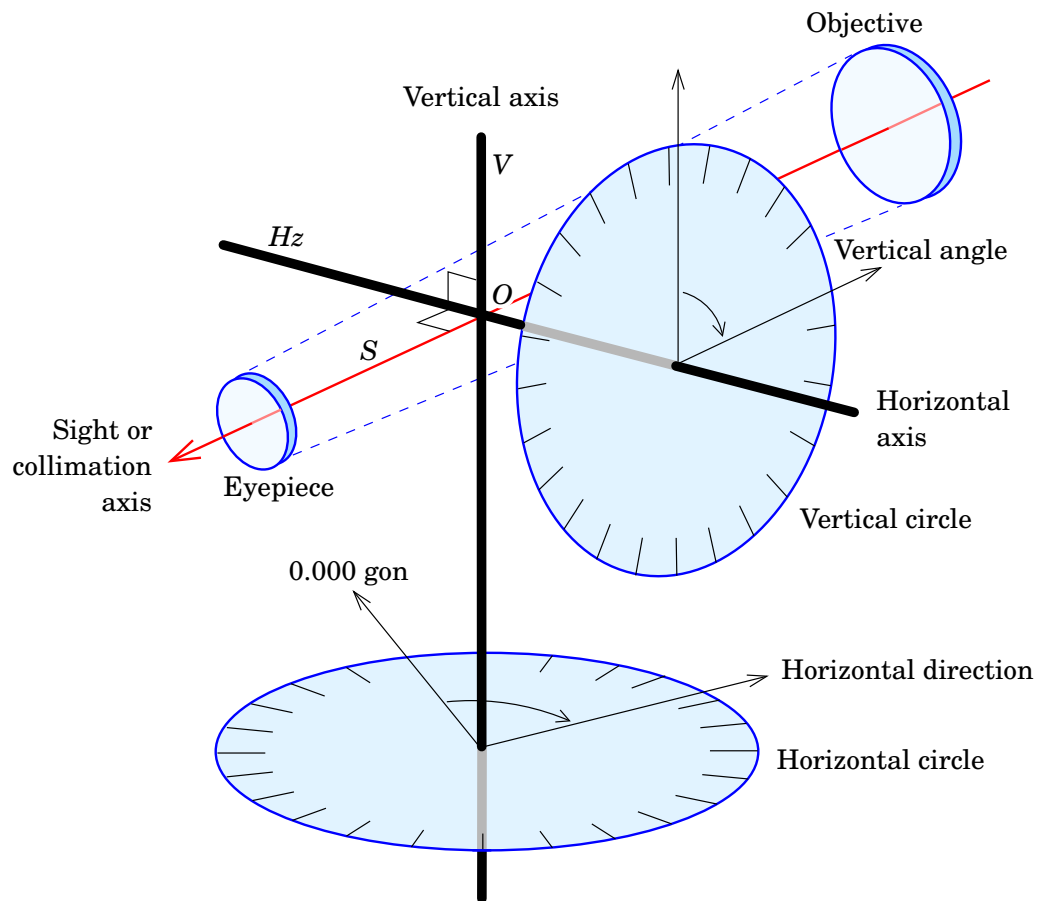


FIGURE 5.3. The axes and circles of a theodolite.

error.

- H_z and V pass through the centres of the vertical and horizontal circles. If not, we speak of *circle eccentricity*.

jakokehän
epäkeskisyyss



5.3 Construction of a theodolite

See figure 5.4.



5.3.1 Measuring telescope

The measuring telescope of a theodolite is in principle the same as that of a levelling instrument (figure 4.8): a complicated assembly of lenses, prisms and mirrors. The telescope turns around the horizontal axis and is joined in its movement by either the vertical circle itself, or by the index and reading microscope of the vertical circle, depending on the instrument type.

pystykehä



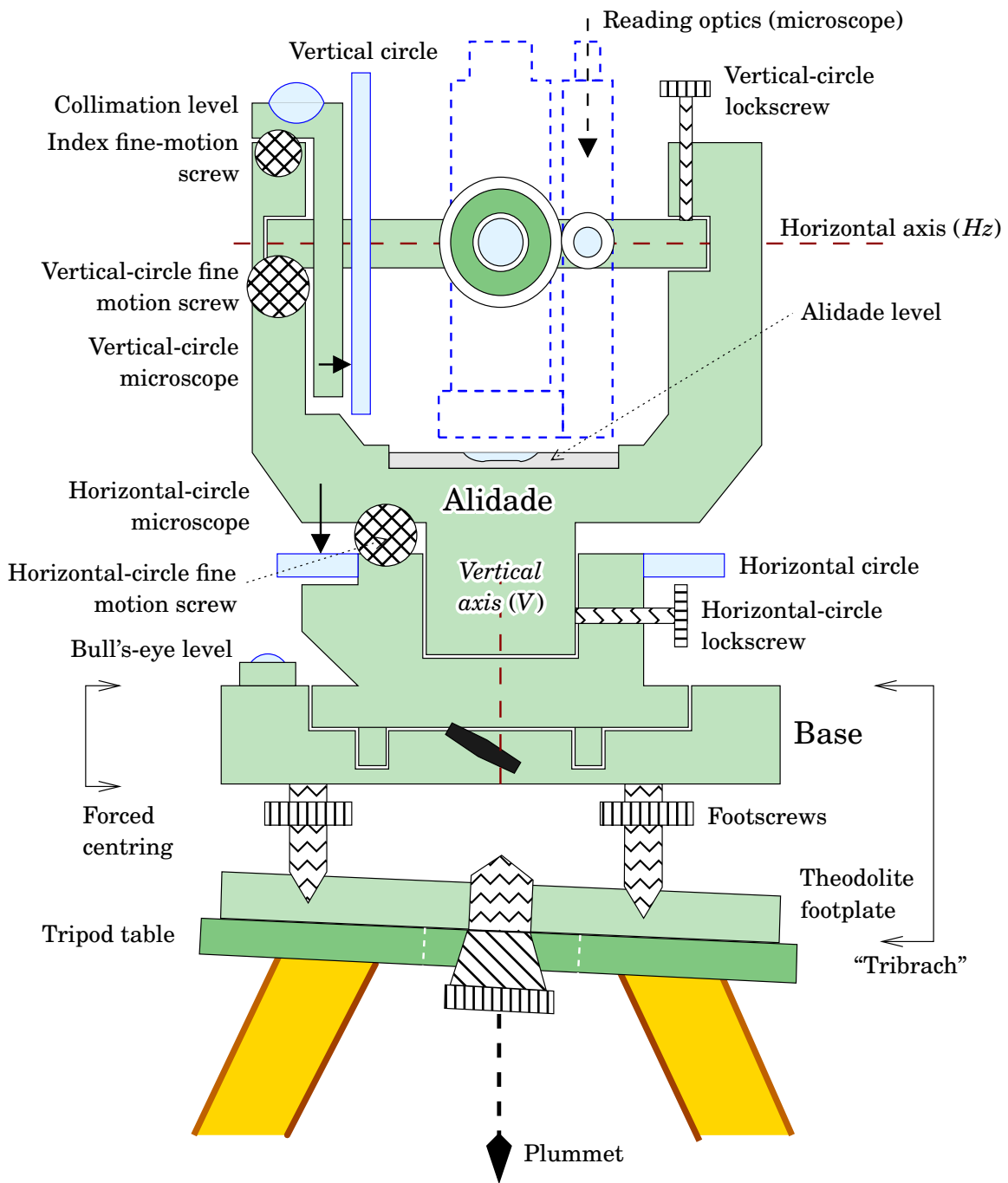


FIGURE 5.4. Theodolite construction.



5.3.2 Reading devices



The theodolite contains reading devices for the horizontal and vertical circles:

- Most often, the images are led by prisms to a microscope eyepiece beside the eyepiece of the measuring telescope, to facilitate the work of the observer.



- The scale microscope and the optical micrometer are means to obtaining a greater reading accuracy, see [Kahmen and Faig \(1988\)](#) pages 66–67.
- In a coincidence microscope, readings are taken from two opposite places on the horizontal or vertical circle. In this case, an optical micrometer is also always used ([Kahmen and Faig, 1988](#), pages 72–76).



5.3.3 Alidade

The alidade (Arabic *al-idhâdah*, ruler) is the central part of the theodolite, which turns around the vertical axis, carrying the telescope with it. It contains a tubular level, the *alidade level*, to precisely level the theodolite, see subsection [5.4.4](#).



5.3.4 Base

runko The *base* of the theodolite is the fixed part to which the horizontal circle is attached, and on which the alidade rests in its bearing.

**pakkokeskistys-
laite**

- The base contains a forced-centring or footscrew device called a *tribrach*.
- The upper part of the base sits *forcibly centred* in this device: it may be detached from and reattached to precisely the same place.
- The forced-centring device is attached to the tripod head using a large screw.² In the middle of the tripod head there is a large round hole and a mechanism to allow enough horizontal motion for *centring*.

2

keskistys

tasaus

- The footscrews are for *levelling* the theodolite: the vertical axis of the theodolite is oriented along the local plumb line.

rasiatasain

- A bull's-eye level is used for crude levelling, while the alidade level helps in precise levelling.

riippuluoti

- The attachment for a string plummet, and an optical or laser plummet, are for centring.



5.4 Theodolite handling in the field

As an expensive fine mechanical and optical instrument, a theodolite should always be treated with due respect:

²The screw conforms to the geodetic standard of $\frac{5}{8}$ inch, 11 threads per inch.





FIGURE 5.5. Forced-centring device or plate.

- Transport, especially over longer distances, is always done in the carrying case.
- The parasol is for the instrument, not the observer. It also offers some protection against rain.
- The instrument is never aimed directly at the Sun: the glass plate with the engraved crosshairs would crack, after which the instrument would need to be repaired and re-calibrated. Further damage occurs if the instrument contains a range-finder.
- *Careful book-keeping* is important: write up anything that might be relevant, like weather conditions (*metadata*).

etäisyysmittari

5.4.1 Monuments and point descriptions

Unlike a levelling instrument, a theodolite must be placed precisely over the point to be measured, so that the theodolite measurement refers to the point.

In figure 5.7 the point Q has been monumented in the terrain, a measurement point with a central mark. Figure 5.6 shows some examples of monument types in use.

maastomerkki

The choice of monument type must be made such that

- The point is clearly and uniquely defined. If one intends to do both precise horizontal positioning and heighting, it is best to use a steel



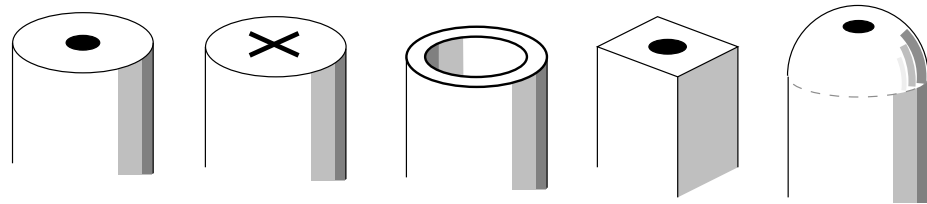


FIGURE 5.6. Various monument types.



bolt with a round head with a small hole in the middle.

- The monument withstands the impacts of the weather and environment — frost heaving! A bedrock point is the best. Iron monuments should be protection painted (rust-protection paint; “red lead” is no longer recommended for toxicity reasons).
- The monument can be easily found — point description!
- The point number is marked on the monument or painted (and chiselled!) next to it.

routa
peruskallio
lyijymönjä

pistekortti When a point is monumented for future use, a *point description* must always be drafted, which helps to find the monument even after decades. The point description may contain the following information:

- The distances of the point from at least three features in the nearby landscape which are believed to be permanent — trees, building corners, etc. — measured by tape. In the point description, a sketch is included of the relative positions of the point and the reference features.
- An approach map including kilometre counts for car drivers, road signs, description of the landscape, and other useful details.
- Precise co-ordinates for hand-held GNSS.
- Well-chosen photographs.



5.4.2 Centring and levelling

In order to ensure that angles measured from the point are correct, two actions need to be performed: *centring* and *levelling*.

keskistys **Centring** In figure 5.7, points Q and O , the monument and the intersection point of the theodolite’s axes, must be on the same plumb line.

tasaus **Levelling** The vertical axis V of the theodolite must be aligned with the plumb line. Then, when the theodolite is turned around the



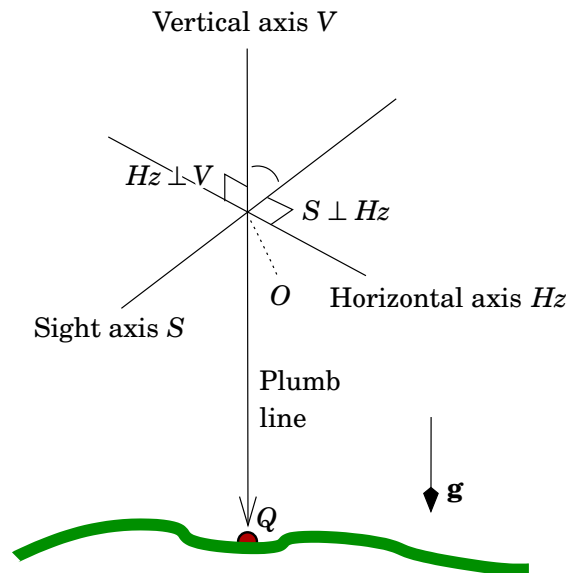


FIGURE 5.7. Theodolite axes. Q is a point monumented in the terrain.

vertical axis, the horizontal axis H_z will maintain a constant angle with the plumb line: ideally, without trunnion-axis tilt, 100° . The horizontal axis will thus always be in the horizontal plane.

Centring and levelling are commonly done alternatingly, until the desired end result is achieved.

tasaus ja keskistys



5.4.3 Crude centring

- The tripod is placed on the right spot, judging by eye and by changing the lengths of the legs.
- It is also judged by eye that the tripod head is horizontal.



5.4.4 Precise levelling

The precise levelling of the theodolite using the alidade level is done in the following steps, see figure 5.8:

1. Crude levelling is done with the bull's-eye level.
2. Level first in the direction of footscrews 1–2, by turning the screws simultaneously in opposite directions. Take into account, if necessary, the *zero error* from step 4.
3. Turn the alidade 100° , and also level in this direction using footscrew 3.
4. Turn the alidade 200° . If the alidade level is adjusted correctly, the bubble must again be in the middle. If not, move the bubble back



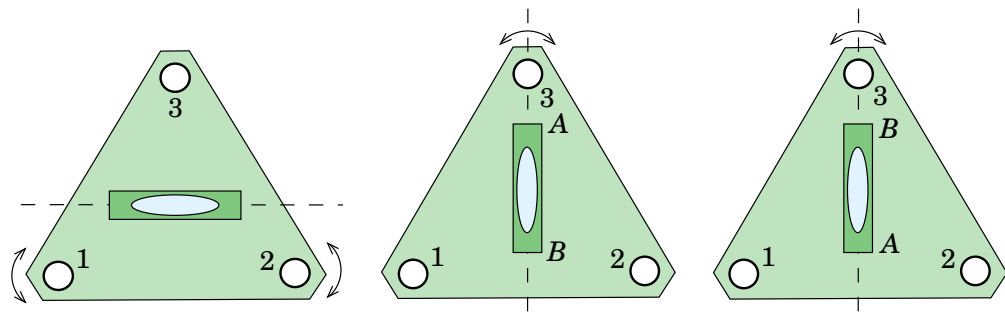


FIGURE 5.8. Precise levelling of a theodolite using the alidade level.

half of the difference, that is to the equilibrium position, using the footscrews. The *zero error* of the level is half the shift of the bubble between steps 3 and 4.

5. Repeat 2–4 until the levelling no longer changes.

Make sure that the bubble can move freely and its end does not attach to the edge of the level.

If the difference found in step 4 is large, then the alidade level is in need of adjustment.



5.4.5 Precise centring

The method of centring will depend on the type of plummet used. The most common is the *optical plummet*.

No plummet is exact. Moreover, measuring the height of the instrument over the monumented point is subject to error. Because of this, in special measurements, when the precision requirement is sub-millimetre, one uses instrument pillars permanently anchored in the bedrock or on a deep support (frost heaving!) instead of tripods.

riippuluoti **String plummet**

- Traditional.
- Attached in such a way, that levelling does not change the centring.
- Centring: *shift* the forced-centring device of the theodolite over the surface of the tripod head in such a way that the tip of the plummet points at the monumented point. Tighten the attachment screw of the theodolite. The central hole in the tripod head gives room to play.
- Levelling with the theodolite's footscrews.



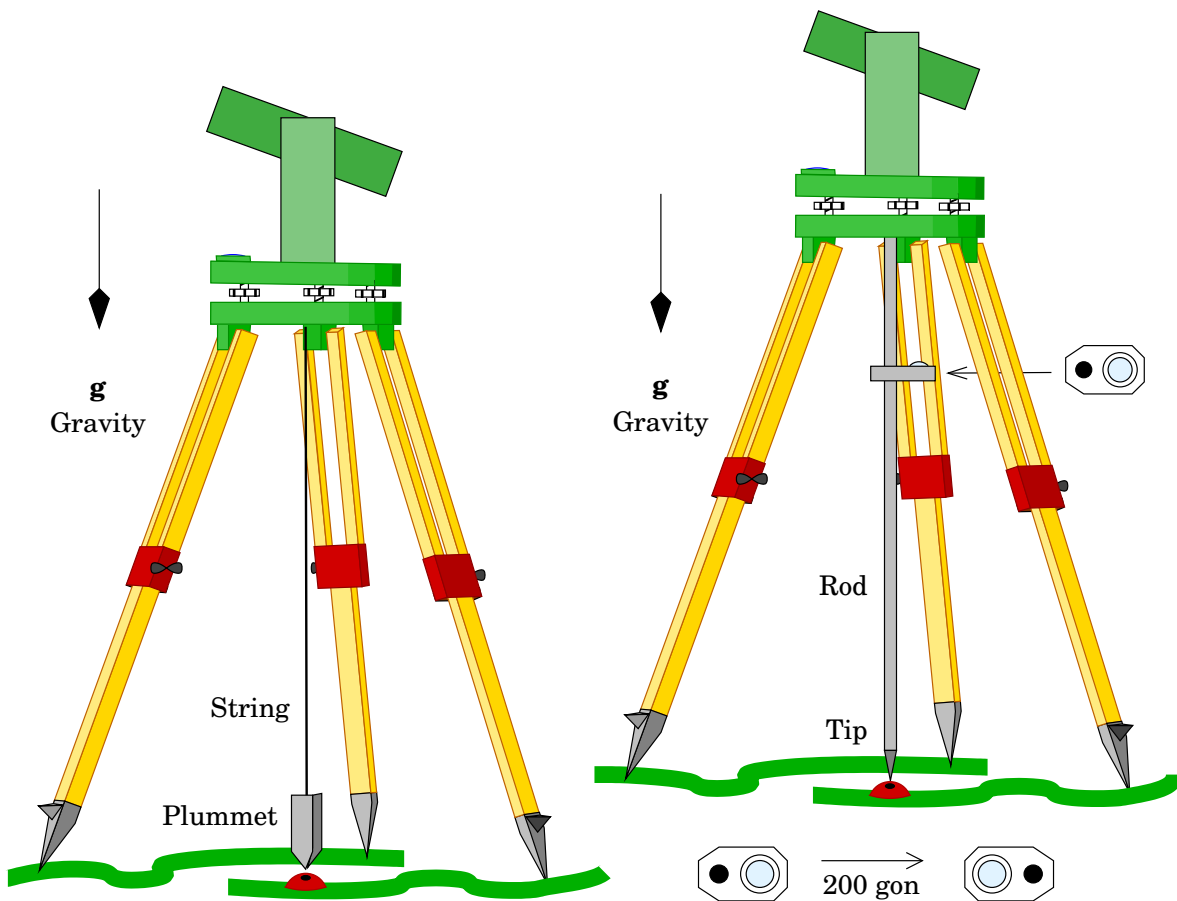


FIGURE 5.9. String plummet and rod plummet.

Problem: sensitive to the wind.

Rod plummet

sauvaluoti

- A telescoping tube, the upper end of which is attached to the forced-centring device (through the central hole in the tripod head), with the lower end placed precisely on the central mark.
- The rod has a bull's-eye level.
- Centring:
 1. Move the forced-centring device so that the bull's-eye level on the rod is in the middle.
 2. Turn the level 200^g to the opposite side of the rod: if the bubble remains in the middle, centring has been achieved.
 3. If not, move the forced-centring device so that the bubble moves half way back to the middle, that is to the



equilibrium position.

The rod plummet comes as part of certain types of tripods (Kern), which are often used in accurate engineering surveying measurements.

optinen luoti **Optical plummet**

- In the forced-centring device's side (or in the side of the theodolite) there is a small telescope with a prism, which looks straight down. In the focal plane of the telescope's eyepiece there are crosshairs or something similar, called an *index*. Katso kuva 5.10.
- The forced-centring device must have been levelled using the bull's-eye level.
- Centring and levelling procedure using an optical plummet (figure 5.12):
 1. Using the footscrews of the forced-centring device, place the index initially on the point marker — meaning that levelling is lost!
 2. By changing the lengths of two of the tripod's legs, get the bubble of the bull's-eye level in the middle again. The forced-centring device describes a *circular movement* around the tip of the third leg, and the place of the index on the image of the point changes only little.
 3. By a parallel shift — loosen the big screw of the forced-centring device just a little — get the index of the optical plummet back on the point marker.
Note! Shift, *do not turn*, because then the levelling will change!
- Because in practice both the levelling 3 and the place of the index 2 will nevertheless always change a little, *repeat* procedure 1–3 until the desired end result is achieved. Fortunately the procedure converges rapidly.

5.4.6 Problem situations

Normally, the centring and levelling procedure as described above will quickly produce a satisfactory result. In real-life field work, however, pathological situations are guaranteed to arise, as in figure 5.13. We



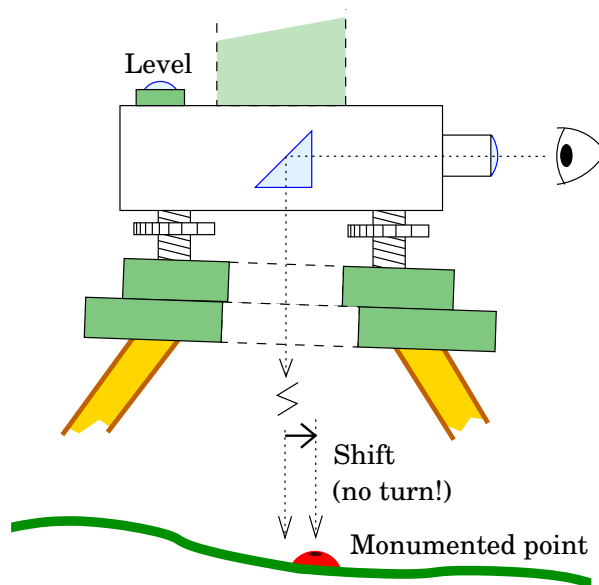


FIGURE 5.10. Optical plummet.

leave these as an exercise for the reader.

5.4.7 Forced centring

In precise measurements, a forced-centring device or tribrach is used: instrument and signal fit into the same device, figure 5.14. In this way *tähys*

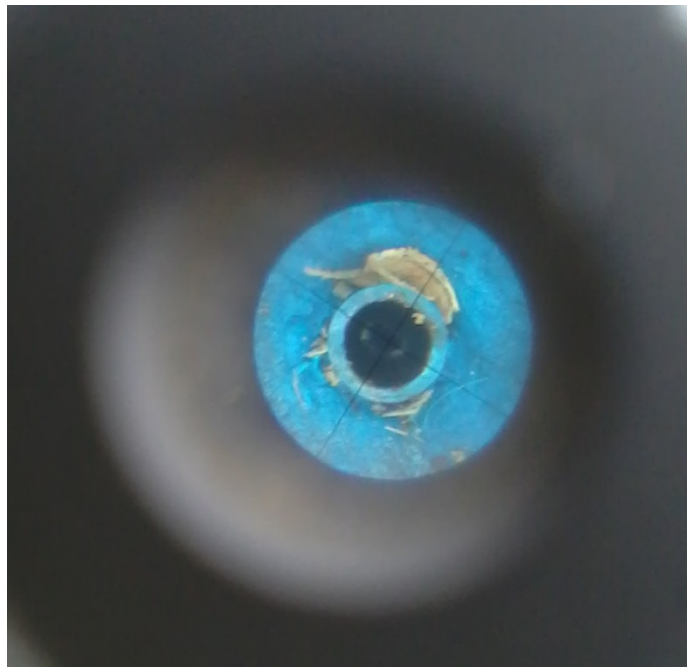


FIGURE 5.11. A benchmark (tube monument) seen through an optical plummet. After focusing, both the target and the crosshairs are sharp.



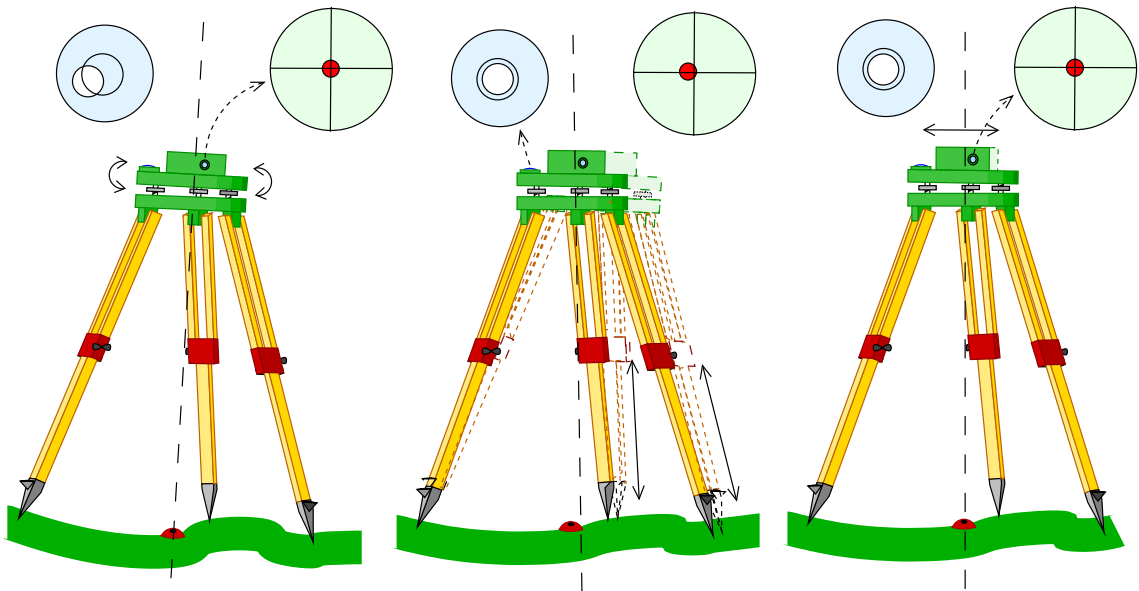


FIGURE 5.12. An optical plummet and a bull's-eye level are used at the same time to achieve centring and levelling.



FIGURE 5.13. Problem situation.



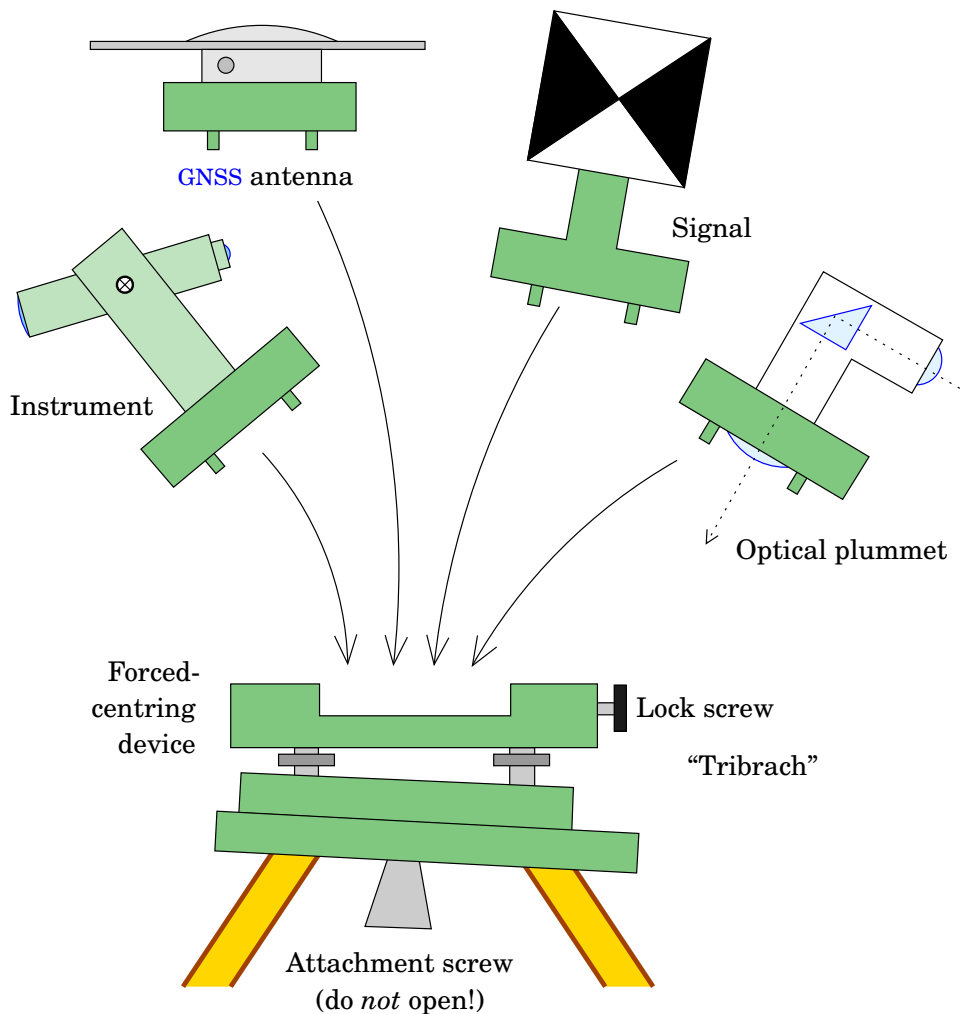


FIGURE 5.14. Principle of forced centring. The instrument and the signal fit into the same device.

one eliminates many of the errors arising from the centring of a tripod.

In really precise measurements, like in engineering geodetic measurements, we use a *separate optical plummet*. Over short distances, centring may be the largest error source.

In the measurement, care is taken that every side is measured in both directions: in a network of three points, the procedure may be according to figure 5.15. From every instrument station, measure to all (nearby) points to be equipped with signals.

Forced centring is also useful in situations where we measure on the same point with both the satellite technique (GNSS) and a terrestrial instrument. The antenna is then attached, through an adapter, to the forced-centring device, figure 5.14.

pakkokeskistys



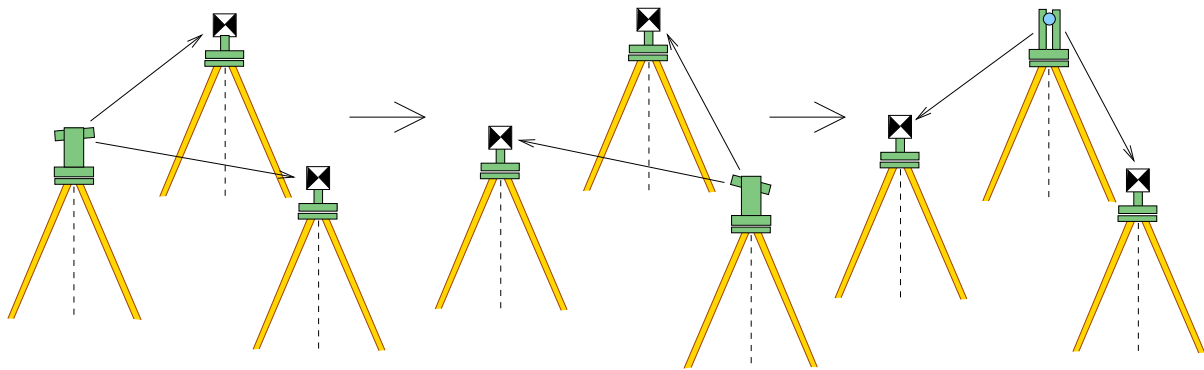


FIGURE 5.15. Measuring a network using forced centring.



5.4.8 Checking an optical plummet

The most common plummet is the optical plummet — although the laser plummet is gaining ground. A laser plummet works practically in the same way as an optical plummet, only the light travels in the opposite direction (Crawford, 2009).

The weak spot of the optical plummet is that it must be adjusted so that it looks really straight down when the bubble of the level is in the middle — the same problem as with a levelling instrument's sight axis and tube-level horizon having to be parallel, figure 4.12. This property, achieved by adjustment, can easily be lost in the handling of the device. For this reason it must be *checked* at regular intervals.

One procedure for checking is described in Kahmen and Faig (1988), on pages 95–96:

1. Mount the forced-centring device — or the whole theodolite, if the optical plummet is built-in — onto a tripod, level it, and mark the point on the floor seen in the plummet's eyepiece on paper taped to the floor.
2. Mark the outline of the forced-centring plate with pencil or chalk on the tripod head.
3. Loosen the forced-centring device from the tripod, rotate it 120° , and place it carefully back into the drawn outline. Level, and mark on the floor again the point under the crosshairs of the plummet's eyepiece.
4. Repeat 3.
5. If the points drawn in steps 1, 3 and 4 on the floor are identical, then the plummet is in good adjustment. If not, shift the crosshairs of the eyepiece, using its adjustment screws, to the centre of mass

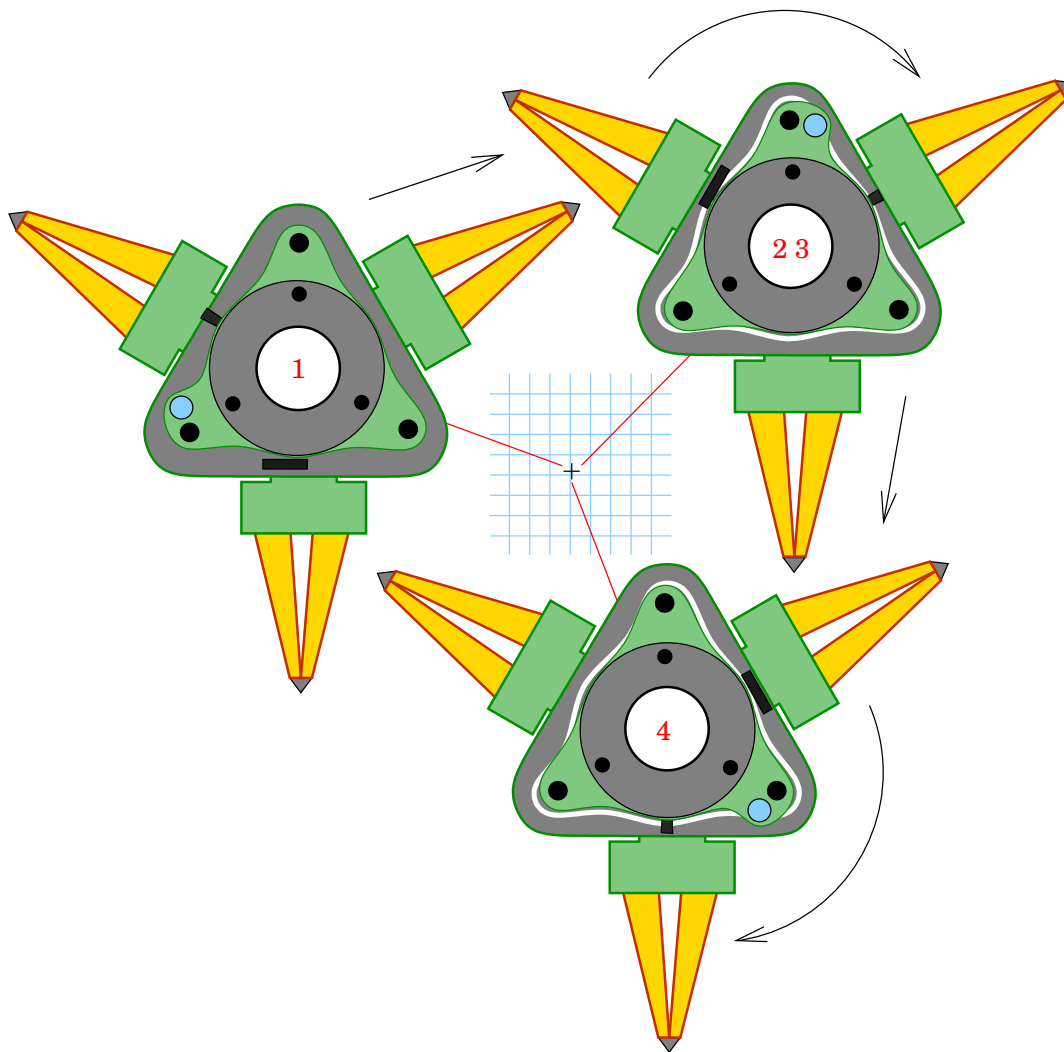


FIGURE 5.16. Checking an optical plummet. The plummet should be adjusted so that it always points to the centre of mass of the three points, the black cross.



of the three points on the floor.

A non-adjusted optical plummet makes the observational material collected with it worthless. A regular check of the state of adjustment, for example before and after field work, is just as important as measuring the theodolite instrument height over the point.



5.4.9 Sighting and targeting

- As with a levelling instrument, always do the focusing carefully.
- At the beginning of the observation work, focus the crosshairs by adjusting the eyepiece. If you wear ordinary eyeglasses, leave these off, as explained in section 4.5.



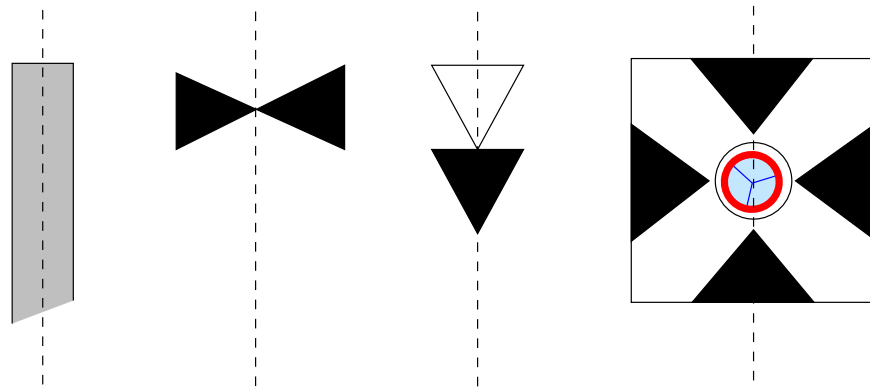


FIGURE 5.17. Good signals for horizontal angles.

sarjahavainto-
menetelmä

- For every target, focus carefully on the signal using the focusing screw.
- During precise measurements using the method of complete sets, one is not allowed to re-focus during measurement, as then the collimation will change a little. This presupposes that all targets are sufficiently far away.

Problem situations:

- The signal is *thin*, for example very far away, and partly covered by the crosshair line. This is not a problem if the crosshairs look like figure 5.18a.
- One side of the signal is in shadow as sunlight comes from the side. The asymmetry causes an error called the *phase error*. See figure 5.18b.

Black-and-white signals are always best.



5.5 Taking readings

In demanding theodolite measurements it is good practice to measure ³ always *in both faces*, ³ face left and face right. The instrument has two face positions because it has two axes: turning it around both axes by an amount of 200^g will make the measurement telescope point at the same object again. This redundancy ⁴ allows the elimination of a number ⁴ of systematic errors. The first position, “face left”, meaning the vertical

kojeasento I

³The inventor of the theodolite with two faces, or “transit”, was presumably William J. Young of Philadelphia, USA, in 1831.

⁴See for theoretical background [Wikipedia, 3D rotation group, Topology](#).



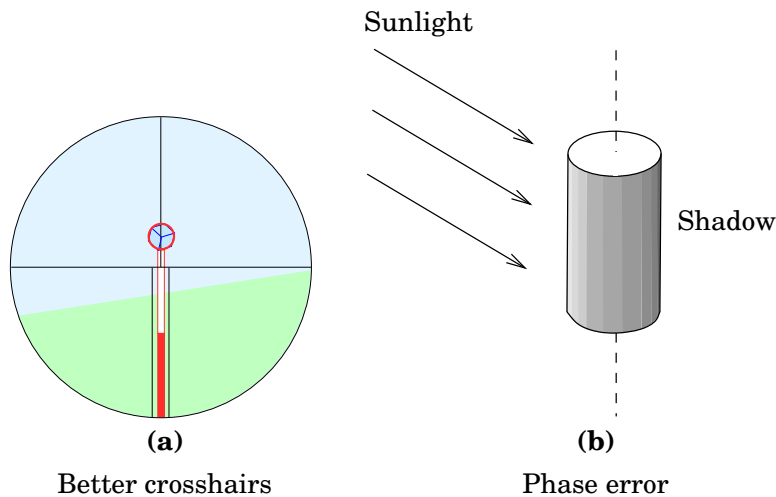


FIGURE 5.18. Targeting. On the left, better crosshairs. Here, both a single and a double line are available for use. On the right, the effect of asymmetric lightfall (“phase error”).

circle to the left of the telescope as seen by the observer, yields vertical angles near 100° if the sighting direction is close to the horizontal. The second position, “face right”, yields values close to 300° .

The traditional method of complete sets, which further reduces a number of instrument errors, is no longer used:

- The measurement of first-order or national base networks has fully transitioned to using satellite positioning.
- Digital angle measurement techniques automatically carry out a procedure equivalent to the method of complete sets, without observer intervention.

5.5.1 Graduation circles and the classification of theodolites

The graduation circles of a theodolite are generally made of glass. The diameter is 60–100 mm, for the most precise instruments 250 mm. The interval for the main scale is 1, 0.5, 0.2, or 0.1 gon, depending on the precision class of the instrument. The traditional method draws a dense line pattern using a “graduation machine” into a layer of wax covering the glass, the lines are etched into the glass using acid, and filled with dye. This was for a long time a carefully protected Swiss business secret (Penry and Ingram, 2013).

Theodolites are classified, traditionally and somewhat unofficially, into “one-minute theodolites”, “one-second theodolites”, and precision



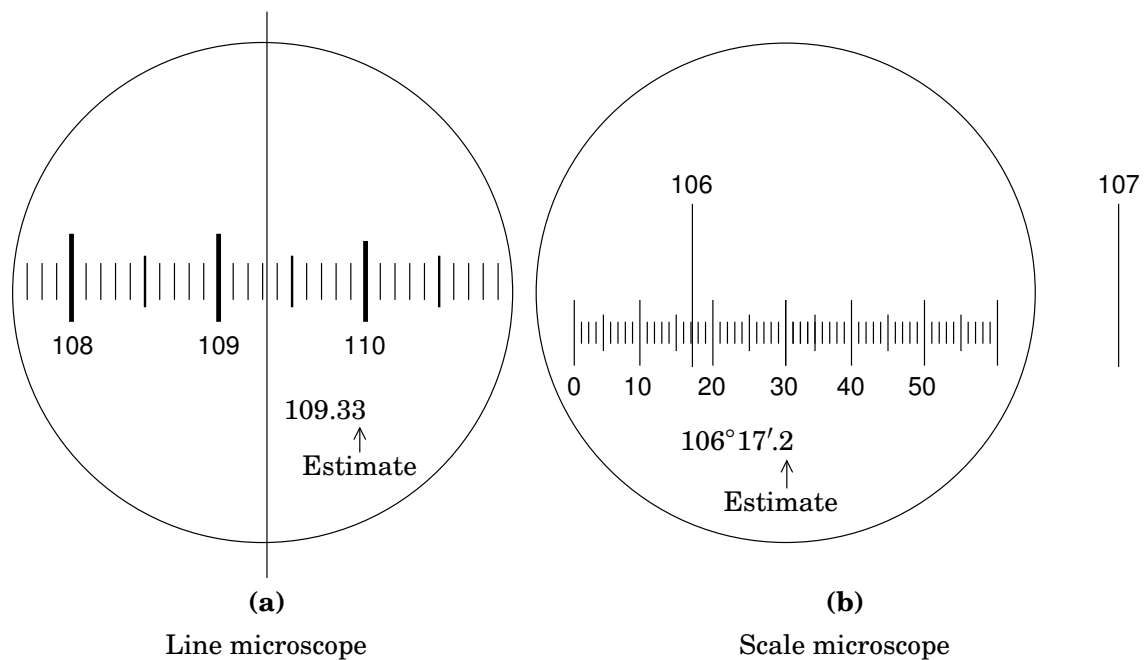


FIGURE 5.19. Various types of reading microscopes.

theodolites. The boundaries separating these classes are not well defined. One source (Simonen, 2012) gives > 1 mgon for one-minute theodolites, $0.5 - 1.0$ mgon for one-second theodolites, and < 0.5 mgon for precision theodolites. See also Anon. (1971).



5.5.2 Reading devices and fine-reading methods

The reading device *magnifies* the image of the scale on the graduation circle through a microscope. With it, a fine reading is *interpolated* between the numbers on the main scale.

Reading microscopes are divided into two types (figure 5.19):

- Line microscope:
 - Magnification of the main scale and an index line.
 - The fine readings are estimated visually, precision is about one tenth of a graduation interval.
- Scale microscope:
 - An additional scale, the length of which is the same as the graduation interval of the main scale: generally, main interval is 1 gon, the interval of the additional scale 0.01 gon, 100 graduation lines.

jakoväli

jakoviiva



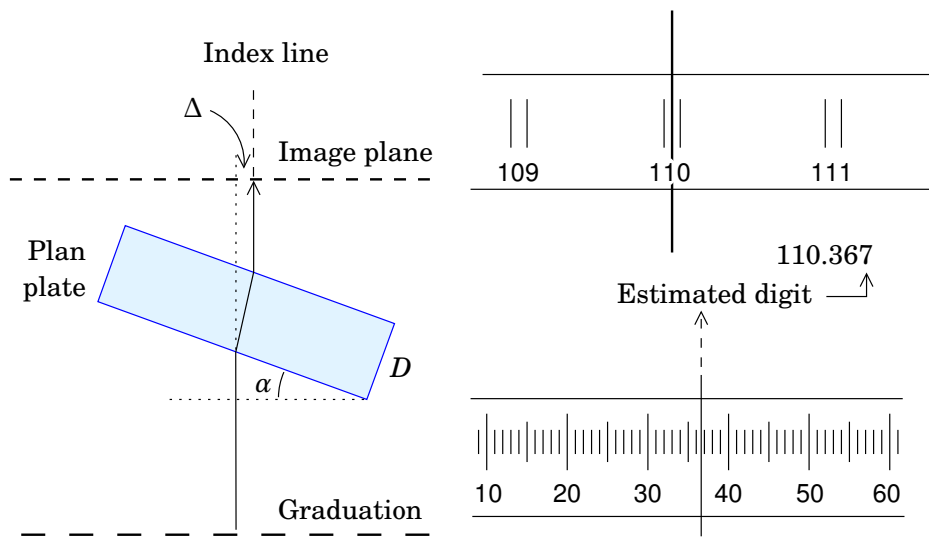


FIGURE 5.20. Optical micrometer and its reading.

- The graduation line of the main scale is used as the index of the additional scale.

The line microscope may be equipped with an *optical micrometer* in order to achieve a better reading precision, figure 5.20.

When the glass plate is turned by angle α (assumed small), the light ray shifts in a parallel fashion by an amount⁵

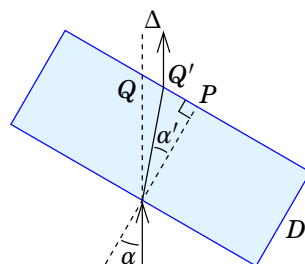
$$\Delta = D \sin \alpha \cos \alpha \left(1 - \frac{1}{n}\right) \approx \alpha D \left(1 - \frac{1}{n}\right),$$

in which n is the index of refraction of the glass. The mechanism for turning the glass plate has a scale showing the value Δ in the angular units of the main scale.

The graduation lines of the main scale are double lines, so aligning with them is easy.

⁵Because

$$\left. \begin{aligned} PQ &= D \sin \alpha, \\ PQ' &= D \sin \alpha' = D \frac{\sin \alpha}{n} \end{aligned} \right\} \Rightarrow Q'Q = D \sin \alpha \left(1 - \frac{1}{n}\right),$$



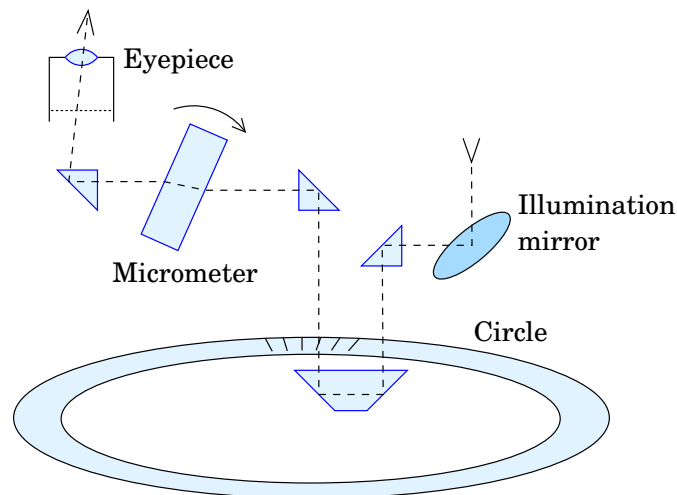


FIGURE 5.21. Reading the graduation circle. One circle location.

The *nonius* or *vernier* is no longer used in theodolites. In [Heiskanen \(1943\)](#) it is told that the Portuguese Pedro Núñez (Petrus Nonius) invented a precursor to the nonius as early as in 1542, and the Frenchman Pierre Vernier today's nonius in 1631. See also [Kahmen and Faig \(1988\)](#) page 65.



5.5.3 Reading the circle

- One circle location is read, figure 5.21.
- In *precision theodolites*, two opposite circle locations are simultaneously read. In this way the eccentricity error of the circle is cancelled out. See figure 5.22. The crude reading is taken from the circle, the fine reading from the micrometer: $244 + 0.4 + 0.0617$ (in which 7 is estimated) or 244.4617.

A single circle reading is used in one-minute theodolites, whereas the method using two opposite readings, using a coincidence microscope, is usually found in one-second and precision theodolites.

The same methods are used for reading the horizontal and the vertical circles. Generally the diameter of the vertical circle is less than that of the horizontal circle, so the reading precision is correspondingly less.

The reading device includes furthermore

it follows that

$$\Delta = Q'Q \cos \alpha = D \sin \alpha \cos \alpha \left(1 - \frac{1}{n}\right).$$



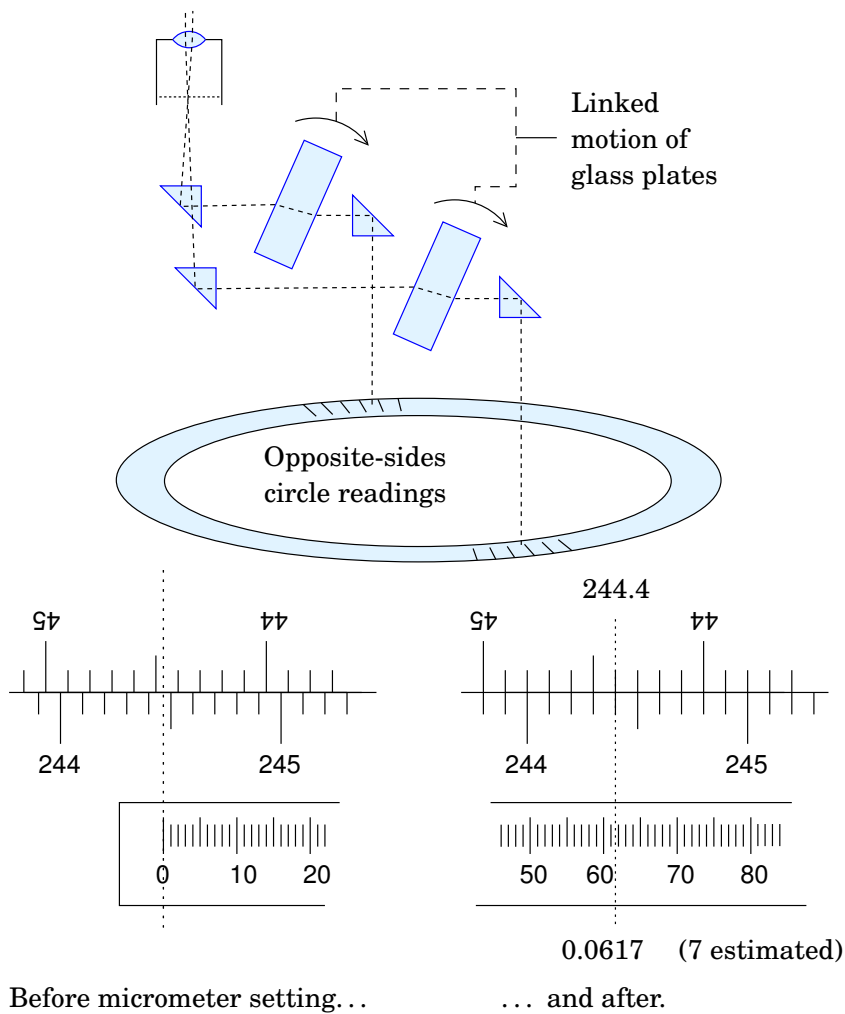


FIGURE 5.22. Reading the graduation circle. Two opposite circle locations.

- A *reading microscope*, to which all the readings of the optics are guided. In the field of view of the microscope, usually both horizontal and vertical readings can be seen. One has to be careful to choose the correct and wanted numbers. Often, the vertical reading is marked with the letter *V* and the horizontal reading with *H* or *H_z*.
- An *illumination system*:
 - a mirror that can be turned to guide light to the reading optics
 - a light source with a battery to be mounted in the place of the mirror or in its own socket.



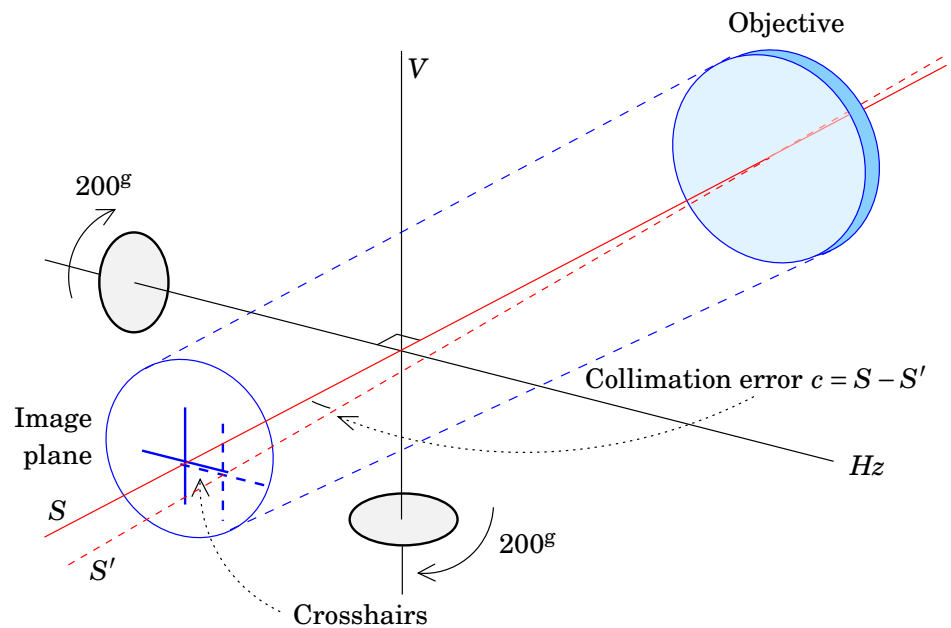


FIGURE 5.23. Adjusting the sight axis by shifting the crosshairs.



5.6 Instrumental errors of a theodolite

The instrumental errors of a theodolite are divided into the following groups:

- Axis errors:
 - The axes are not perpendicular to each other: *collimation error* and *trunnion-axis tilt*, subsections 5.6.1 and 5.6.2.
 - The axes do not intersect at the same point.
- Eccentricity errors:
 - eccentricity of the graduation circles
 - eccentricity of the measuring telescope.
- graduation errors of the circles.

jaotusvirhe

In modern theodolites the eccentricity and graduation errors are small.

The instrumental errors cannot in general be corrected, but they can be determined, in *calibration*.



5.6.1 Collimation error

The most consequential error is the *collimation error*, which is easy to determine and correct. Collimation error means that the angle between the sight axis S and the horizontal axis H_z is not a right angle: $S \not\perp H_z$.

The sight axis S is realised by the crosshairs in the measuring telescope



(more precisely, S is the straight line going through the optical centre of the objective and the crosshairs, figure 5.23). This is why by shifting the crosshairs one adjusts to $S \perp Hz$.

- If there is no collimation error ($c = 0$), the readings of the same target A in face left and face right are a_1 and a_2 , where $a_1 = a_2 \pm 200^g$.
- If there is collimation error ($c \neq 0$) and the telescope is turned through from face left to face right (200^g around both the horizontal and the vertical axis), then A will not show under the crosshairs until the telescope is turned an amount $2c$ more.

Assume that object A is approximately in the horizontal plane. Take, in face left, reading a_1 , and, in face right, reading a_2 .

The correct readings, without collimation error, *would* be A_1 and A_2 , and we would have exactly $A_1 = A_2 \pm 200^g$.

In reality we obtain $a_1 = A_1 + c$ and $a_2 = A_2 - c$. Their difference is

$$a_1 - a_2 = 2c \pm 200^g,$$

from which one obtains

$$c = \frac{1}{2}(a_1 - a_2 \pm 200^g). \quad (5.1)$$

In this way one may determine c , usually a small number.

Because c is so small, only fractions of a gon are of interest, not the integer number of gons. Therefore we use the following notation:

$[a]$ signifies the rounding residue of a , the difference between the precise value and the value rounded to an integer. So $[127.4531] = 0.4531$, $[16.9850] = -0.0150$, and so on. The outcome of the operation is always between the values -0.5 gon and $+0.5$ gon.

Then

$$c = \frac{1}{2}[a_1 - a_2].$$

This can be calculated from the observation notebook, if the same objects have been measured in both faces.

$$[c] = \frac{1}{2} \frac{[\sum a_1 - \sum a_2]}{n} = \frac{\sum [a_1 - a_2]}{2n}.$$

a_1 reading taken in face left



a_2 corresponding reading in face right (correspondence means both the same set and the same direction)

n number of readings (sets \times directions).

Correcting collimation error Carry out, in face left, a turning of the instrument so that the reading is the precomputed “true value” $A_1 = a_1 - c$. ⁶Shift the crosshairs (focal plate) to be on top of the image of the target A . Verify, in face right, that the second “true value” $A_2 = a_2 + c$ is obtained when the crosshairs are on the target.

A calculation example from an observation notebook (all values in gon):

Face	1	2	3	4	$\sum_{i=1}^4$
Left	40.5223	88.6932	119.7601	205.7571	454.7327
Right	240.5217	288.6924	319.7591	5.7564	854.7296
$a_1 - a_2$	-199.9994	-199.9992	-199.9990	200.0007	-399.9969
$[a_1 - a_2]$	0.0006	0.0008	0.0010	0.0007	0.0031

The number of observations per face $n = 4$, so

$$[c] = \frac{3.1 \text{ mgon}}{8} = 0.3875 \text{ mgon} = 3^{\text{cc}}.875.$$

This is a good *field check*. In this example, there is hardly any collimation error.



5.6.2 Trunnion-axis tilt

tappikaltevuus The second axis error is the *trunnion-axis tilt* t . This error means that the horizontal axis, or trunnion ([Wikipedia, Trunnion](#)) axis, H_z , is not actually horizontal even after levelling the instrument: it and the vertical axis V are not perpendicular, $V \not\perp H_z$.

The trunnion-axis tilt may be determined by using a target far away from the horizontal ($\zeta \neq 100^{\text{g}}$). Then it holds that

$$a_1 = A_1 + c + t \cos \zeta, \quad a_2 = A_2 - (c + t \cos \zeta),$$

from which the combined term $c + t \cos \zeta$ can be determined in the same way as explained above for c . Separating the errors c and t requires measurements at two different vertical angles ζ . If one of these is $\zeta = 100^{\text{g}}$ ($\cos \zeta = 0$), we get back the original formula 5.1 for calculating the collimation error.

⁶This is a task for maintenance: the screws are small and hidden.



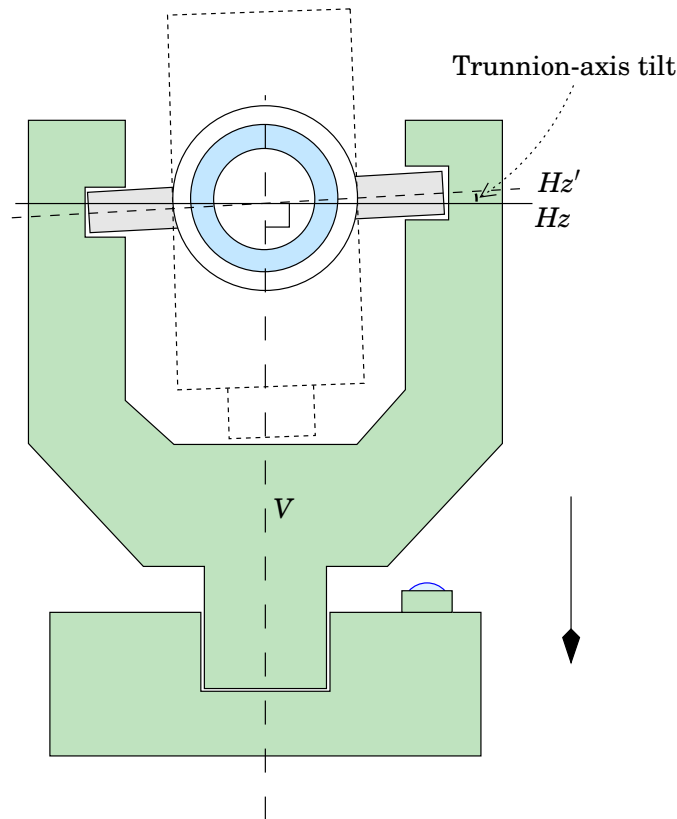


FIGURE 5.24. Trunnion-axis tilt.



5.6.3 Zenith-angle measurement and index error

Measuring a zenith angle with a theodolite requires that the index of the vertical circle — the place where the values for the zenith angle are being read — is in the horizontal plane. For this purpose a theodolite has its own fine-motion screw (figure 5.25) with which the frame on which the index or indices of the vertical circle and the vertical-index level (collimation level) are mounted together is moved. The frame can be turned slightly around the horizontal axis. The measuring telescope and the vertical circle are similarly connected to each other.

pystykehän
indeksi

Before every zenith-angle measurement it must be ascertained, using the fine-motion screw of the level and index, that the vertical-circle indices are really in the horizontal plane. Of course one may not assume that the vertical-index level has been adjusted so that the index really gives exactly 100^g and 300^g precisely when the optical axis of the measuring telescope is horizontal. This error is called the *index error* (i). It can be eliminated by measuring *in both faces*, left and right. One moves between faces by “plunging the instrument through”: turn the telescope 200^g around the vertical axis, and 200^g around the horizontal axis. The



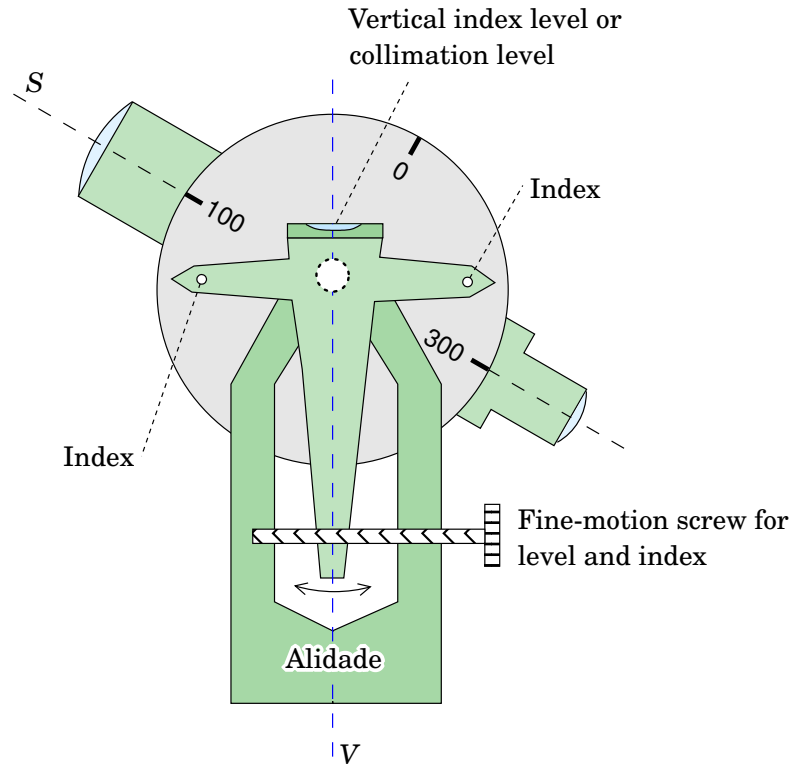


FIGURE 5.25. Observing a zenith angle.

measurement values obtained in both faces are added together:

$$\zeta_1 + \zeta_2 = 400^g + 2i,$$

from which the index error i follows.

Both observations ζ_1 and ζ_2 are corrected by an amount $-i$:

$$\zeta'_1 = \zeta_1 - i, \quad \zeta'_2 = \zeta_2 - i.$$

After this, the condition

$$\zeta'_1 + \zeta'_2 = 400^g$$

holds exactly.

See figure 5.26, which depicts (unlike figure 5.25) a theodolite of which the vertical circle is read only in one place. In the figure, the sight axis is in the horizontal plane, so $\zeta = 100^g$. The formulas given below, however, apply generally.

In the left image, the angle ζ is measured and the reading $\zeta_1 = \zeta + i$ is obtained, in which i is the index error. In the right image, the same angle ζ is measured, but in face right, and the reading obtained is $\zeta_2 = (400^g - \zeta) + i$. We obtain:

$$\zeta_1 + 400^g - \zeta_2 = \zeta + i + \zeta - i = 2\zeta.$$



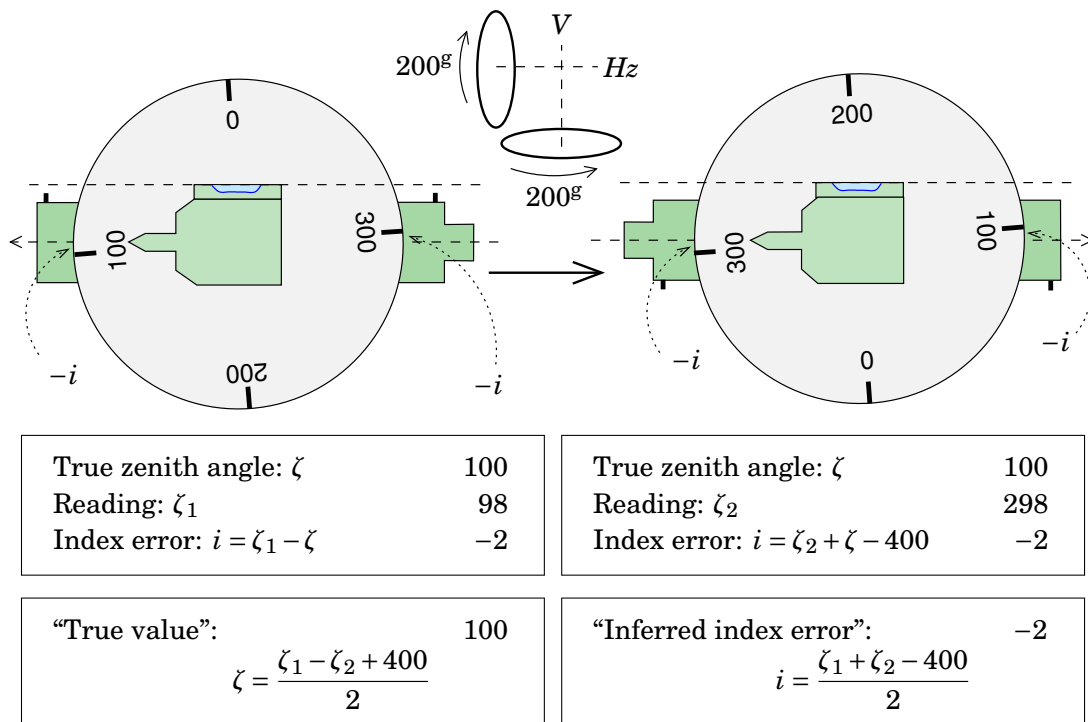


FIGURE 5.26. Index error.

The angle ζ :

$$\zeta = \frac{1}{2} (\zeta_1 - \zeta_2 + 400^g).$$

The angle i :

$$\zeta_1 + \zeta_2 - 400^g = \zeta + i - \zeta + i \implies i = \frac{1}{2} (\zeta_1 + \zeta_2 - 400^g).$$

The corrected reading is

$$\zeta = \zeta_1 - i = 400^g - \zeta_2 + i.$$

Removing the index error Assume the construction in figure 5.25.

1. Aim at a target of which the “true angle” ζ has been calculated.
2. Turn the index and collimation level together, using the fine-motion screw, to reading ζ .
3. The bubble of the level will move, adjust the level with its adjustment screw(s)⁷ until the bubble is in the middle again.⁷

⁷The adjustment screws are small and may be somewhat hidden. Adjustment is a maintenance job.



Check

1. Observe the target in face left (ζ_1) and immediately after, the same target in face right (ζ_2).
2. Calculate the sum $\zeta_1 + \zeta_2$. If the index error $i = 0$, $\zeta_1 + \zeta_2 = 400^g$. This is also a good field check.

In figure 5.25 there is a manual vertical-circle index (height index). An automatic height index is more common. The principle is similar to that used in a self-levelling levelling instrument, and like it, the index must already be approximately level:

- A pendulum compensator.
- A liquid compensator using silicone oil. The light to the reading microscope is either refracted through the liquid or reflected from its surface. See [Kahmen and Faig \(1988\)](#) pages 394, 395.
- In electronic theodolites, this compensation mechanism has been implemented digitally using a digital tilt meter. So, the index is not adjusted but rather, the reading is corrected computationally based on the reading from the tilt meter.

In zenith-angle measurements, instrumental errors are largely the same as for horizontal angles. Some are eliminated by measuring in both faces. Measuring complete sets will certainly not diminish the impact of vertical circle graduation errors, because the vertical circle, unlike the horizontal circle, cannot be loosened and turned, see section 6.3. It does allow however for a check on the correctness of the measurements.

The observations in face left and face right should be made *as quickly as possible in succession*. For this reason, vertical angles should always be observed separately, never together with horizontal angles.



5.7 Electronic theodolites

In *electronic* theodolites all measurements are obtained in *numerical* form, facilitating automatic storage, correctness check in connection with the measurements, and forwarding. The monetary savings in the measurement activity can be substantial, not just in measurement time saved, but also in *quality gain*, when the error-prone reading and manual writing down of observations is eliminated.

However, though electronic theodolites record the observations on their own — typically raw data is not even recorded but already pre-processed



0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0
0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0
0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0



FIGURE 5.27. A four-bit Gray code.

data — the information to go with the data (“metadata”) must still be carefully recorded.

In electronic theodolites, the edges of the graduation circles have been imprinted with different line patterns.



5.7.1 Absolute encoding circles

The *Gray code*⁸ is often used, which consists of bit strings in which, at every step, only one bit changes. In figure 5.27 we see an example of a four-bit Gray code, which has 16 different values.

In realistic applications, more bits are used, for example if one wants a horizontal-angle resolution of 0.1 gon, there must be $400 \cdot 10 = 4000$ different values. This already requires 12 bits.⁹

The code is a black-and-white pattern imprinted on the circle, which a row of light sensitive diodes, photodiodes, scans, figure 5.28 left.

The advantage of the Gray code is that one knows constantly, unambiguously, which place on the circle one is reading; in other words, there is a *zero direction marked* on the circle. This is why the encoding is called absolute. In the example case, four photodiode rows side by side are needed, one for each bit field.

The patterns on the circles are observed electro-optically, in our example using a row of photodiodes, nowadays an integrated microelectronic circuit, like a CCD sensor. The light detector observes the edges between the black-and-white fields on the circle. Crude, absolute values originate from the edges of the pattern on the circle; more decimal places are obtained from the interpolating property of the diode array — or imaging light sensor.

⁸Frank Gray (1887–1969) was an American physicist and electronicist, a developer of television technology. The Gray code is part of a patent awarded to him in 1953.

⁹... because $2^{12} = 4096 \gtrsim 4000$.



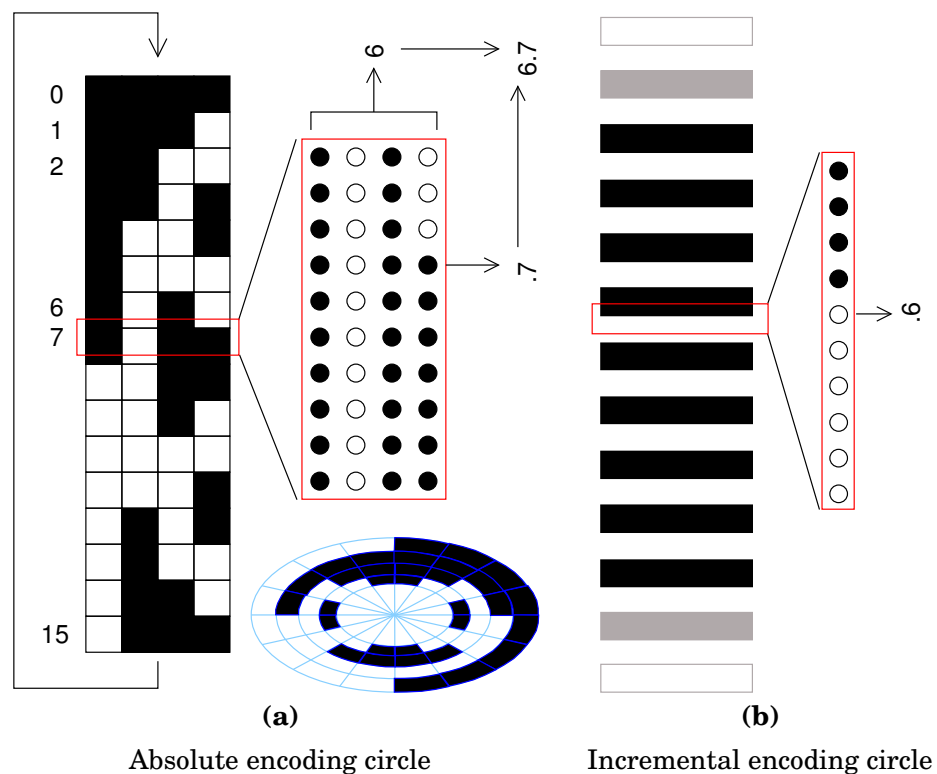


FIGURE 5.28. An absolute and an incremental encoding circle.

5.7.2 Incremental encoding circles

koodauskehä There is *no* zero direction marked on the encoding circle. We may only follow *changes* in the instrument orientation by counting lines, figure 5.28 right. In the incremental solution, at least two diode detectors are needed to establish the turning direction: in the depicted example there are ten. The incremental method of course can only measure direction *differences*, that is angles. Crude readings come from counting lines on the circle, more decimal places are again obtained by interpolation.

5.7.3 Modern automatic instruments

In electronic theodolites, commonly an **ATR**, *automatic target recognition*, system is used. For this purpose, a **CCD** image sensor has been placed in the image plane of the measuring telescope. The telescope is aimed visually at the object. The uncorrected reading comes, as usual, from the digital encoding-circle system. To the reading is added programmatically, by means of image processing, a *correction* from the **CCD** image.

With electronically readable circles one uses many read-out detectors of which each reads the code from a different place on the circle, figure 5.29.



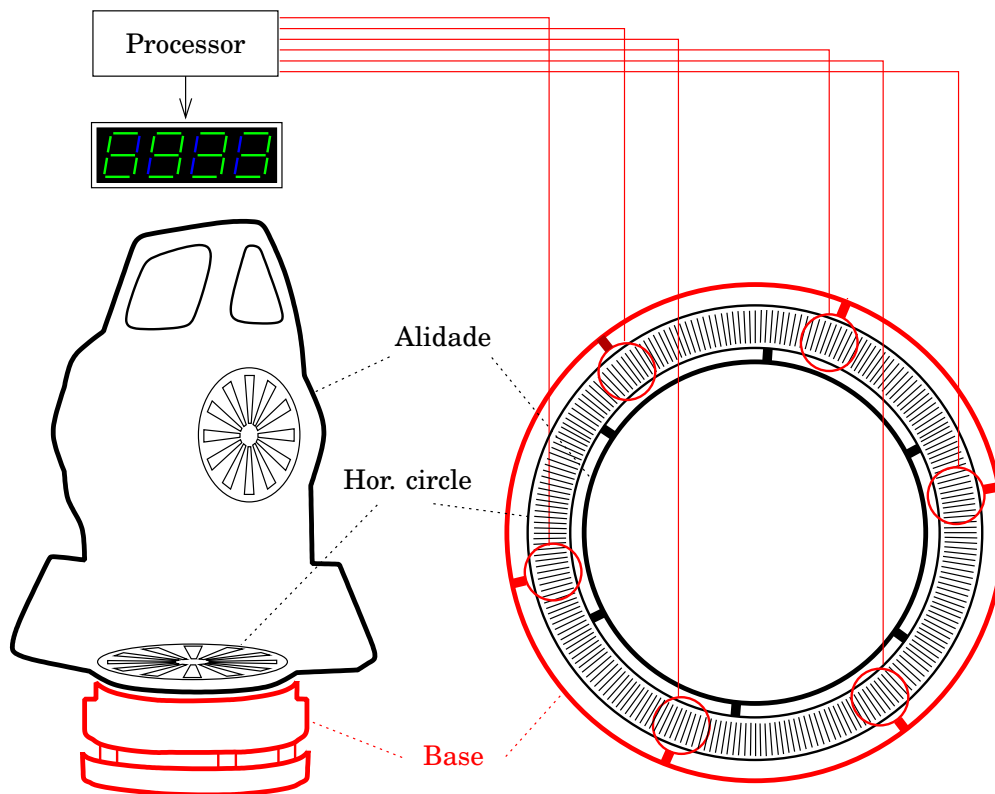


FIGURE 5.29. Electronic readout of the horizontal circle: readings are taken from all parts of the circle. This replaces the method of complete sets known from optical theodolites.



All readings are collected in a processor, which averages them. In the average, the periodic and random errors of the circle are eliminated or essentially reduced. Compare this with the method of complete sets used with optical theodolites, section 6.3.

With electronic theodolites, locking and unlocking the horizontal circle, and turning it — and thus applying the method of complete sets by hand — is unnecessary and impossible. Turning over the telescope and measuring in both faces is however still necessary, for the same reasons as with optical instruments (section 5.5).

vaakakehä

In some instruments, a *rapidly spinning circle* is used, drawing on the same technology used in spinning computer disc drives. In this dynamic solution, the light detectors transform the line pattern on the circle into a block signal, figure 5.30. The time shift between two block signals together with the circle's spinning velocity ω gives directly the angle $\alpha(t)$ between the detectors as a function of time — plus an unknown constant, the “ambiguity”,¹⁰ which does not depend on time. For this reason, the

kokonaisluku-
tuntematon
10



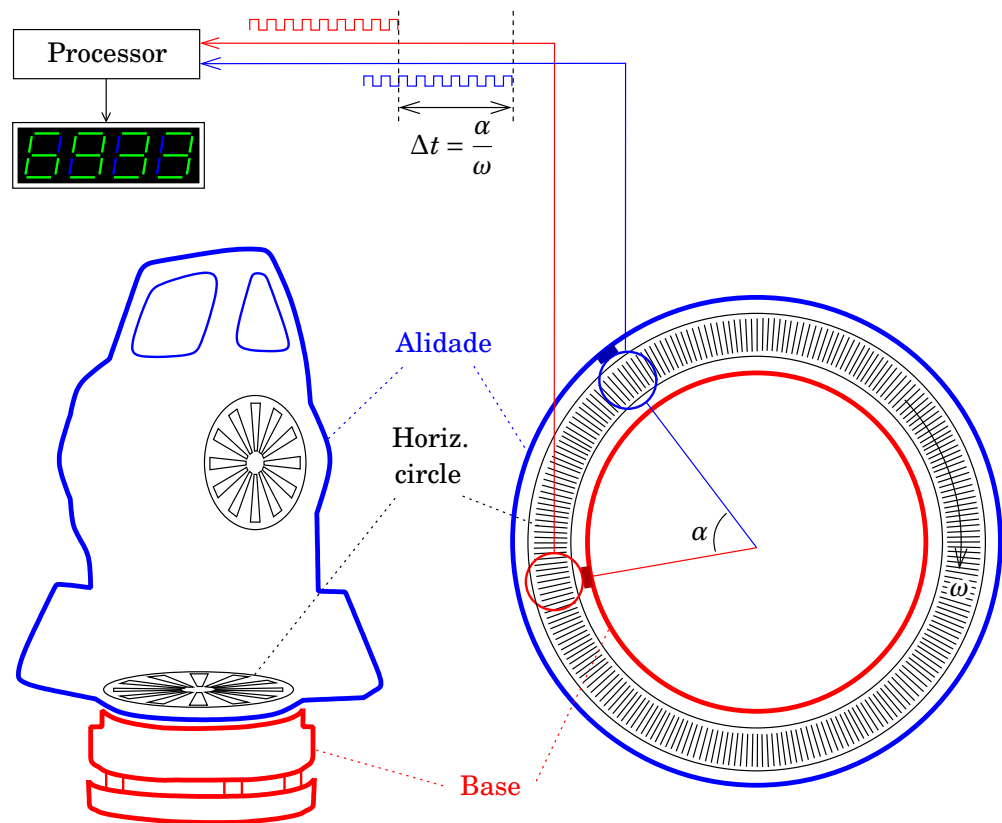


FIGURE 5.30. A spinning circle converts the measurement of angles into one of time differences.

spinning circle method is also *incremental*: it can only measure *changes* in the angle.

The advantages of the method are:

- It converts angle measurement into electronic time-difference measurement, which can be extremely precise.
- The precision with which the graduation lines on the circle are produced is not critical; the even spinning speed of the circle is.

The range of measurement systems and technical solutions for electronic theodolites is broad and rapidly developing. Therefore we only take a closer look at one case.

5.8 Case: Leica robotic tacheometer TCA2003

TCA2003 (Leica, 1997) is a good example of a fairly modern electronic theodolite or *tacheometer*. The instrument measures horizontal angles

¹⁰ Obviously because to the detector, every line on the circle looks like every other.

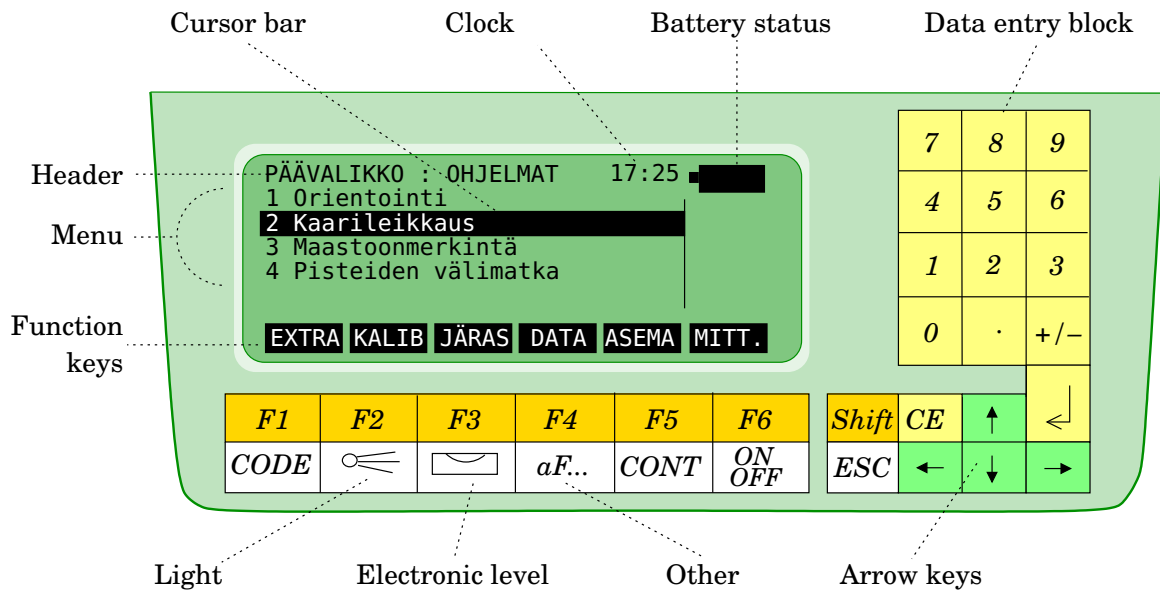


FIGURE 5.31. Leica TCA2003 robot tacheometer control panel (in Finnish).

and zenith angles as well as slant ranges. The calculation capabilities of *vinoetäisyys* its built-in software are quite versatile.



FIGURE 5.32. Leica TCA2003 robot tacheometer.



Measurement precision as stated by the manufacturer amounts to $0''.5$ or 0.15 mgon in horizontal and vertical angle measurement, and $1\text{ mm} + 1\text{ ppm}$ in distance measurement. The longest measurement distance under normal conditions is 3.5 km .

The instrument has co-axial [ATR](#) (automatic target recognition, subsection [5.7.3](#)), which also takes part in angle measurement. The [CCD](#) sensor measures the deviations in the horizontal and vertical directions of the laser beam reflected by the prism, and guides the motors so that the crosshairs are placed almost over the prism. The small remaining deviation is measured in the [CCD](#) image and the angle readings are corrected correspondingly.

The [ATR](#) system can also be programmed to follow a moving target, or to systematically seek — scan for — a target if it is not in the expected place.

The instrument has *self-calibration facilities*, a measurement programme for determining the vertical-circle index error (subsection [5.6.3](#)), the collimation error (sight axis not perpendicular to horizontal axis, subsection [5.6.1](#)), the trunnion-axis tilt (horizontal and vertical axes not perpendicular to each other, subsection [5.6.2](#)) as well as the *zero point* or *constant error* of the distance measurement device or range-finder (subsection [7.4.4](#)).

pystykehän
indeksivirhe

Many “corrections”, or rather *reductions*, of distance measurement, like weather corrections, reduction to the horizontal plane, and even reduction to the map projection plane (for example using the Gauss–Krüger or [UTM](#) projections) can be done within the instrument itself.

laserluoti

The instrument has a laser plummet, which works in the same way as an optical plummet — the precision is also similar, a little better than $\pm 1\text{ mm}$ — but the light moves in the other direction, from the theodolite down to the ground. Levelling and centring procedures are otherwise the same as with an optical plummet.

The instrument has, in addition to a bull’s-eye level, an accurate electronic level. The display is an [LCD](#) and looks the same as a real bull’s-eye level. It is self-calibrating, and it is not necessary to turn the instrument by 100° or 200° every time it is levelled (figure [5.8](#)).

The instrument is highly automated and offers a pre-programmed monitoring measurement of 50 points. This is useful especially in deformation measurements in industry and construction. Two programming languages are on offer: [GSI \(Leica\)](#) for simple use, and [GeoCOM](#) for



advanced use. In addition, the GeoBasic environment allows the development of more applications in the PC environment, and their upload to the instrument.

The exchange of data between the instrument and a computer can be done in two ways:

- Through a serial interface (RS232). This technique is obsolete: modern devices use a USB port or Bluetooth.
- Using a non-volatile PCMCIA memory card. Storage capacity may be 512kB–4MB. The format of the card is the MS-DOS file system FAT. This solution has also been obsolesced by USB memory sticks.

The format of the observations themselves is the Leica-designed GSI (Leica) (Geo Serial Interface), documented in the manual.



Self-test questions

1. Describe the three axes of a theodolite. Which of them are mutually perpendicular?
2. Before measuring from a known point, a theodolite must be *centred* and *levelled*. Describe the stages in which this is done, for a theodolite equipped with an optical or laser plummet.
3. Explain the idea of forced centring. Why is it useful?
4. Reading the circles. What is a scale microscope, what is a coincidence microscope? Make a drawing!
5. Explain Gray codes.
6. What is the difference between absolute and incremental encodings?
7. What is ATR, automatic target recognition?
8. Describe collimation error, trunnion-axis tilt and index error. How are they determined?
9. A theodolite should never be aimed directly at the Sun. Are there exceptions to this rule? Google “Roelofs prism”.





Angle measurement

6

[...] When preparing a third American edition for the press, Blunt decided that Bowditch had revised Moore’s work to such an extent that Bowditch should be named as author.

The title was changed to *The New American Practical Navigator* and the book was published in 1802 as a first edition. Bowditch vowed while writing this edition to “put down in the book nothing I can’t teach the crew,” and it is said that every member of his crew including the cook could take a lunar observation and plot the ship’s position.

Nathaniel Bowditch biography, in *The American Practical Navigator*, *Bowditch*, 2017 edition



6.1 Horizontal angle measurement

In traditional topographic surveying, horizontal angles are measured to be used in three different contexts: maastomittaus

1. One carries out a *triangulation*, where horizontal angles and distances, or side lengths, are measured between points of the triangle network. The triangle sides may be as long as tens of kilometres. From the observations, the geodetic co-ordinates of the triangle points are computed, hierarchically in densification stages of the network. In this way, the first-order triangulation network measured by the Finnish Geodetic Institute has been densified by the Finnish National Land Survey. kolmiomittaus
2. As a further densification of the triangle network, *traverse measurements* are carried out. In this method, side lengths and bending angles are measured between successive points of a traverse (figure monikulmiojono)

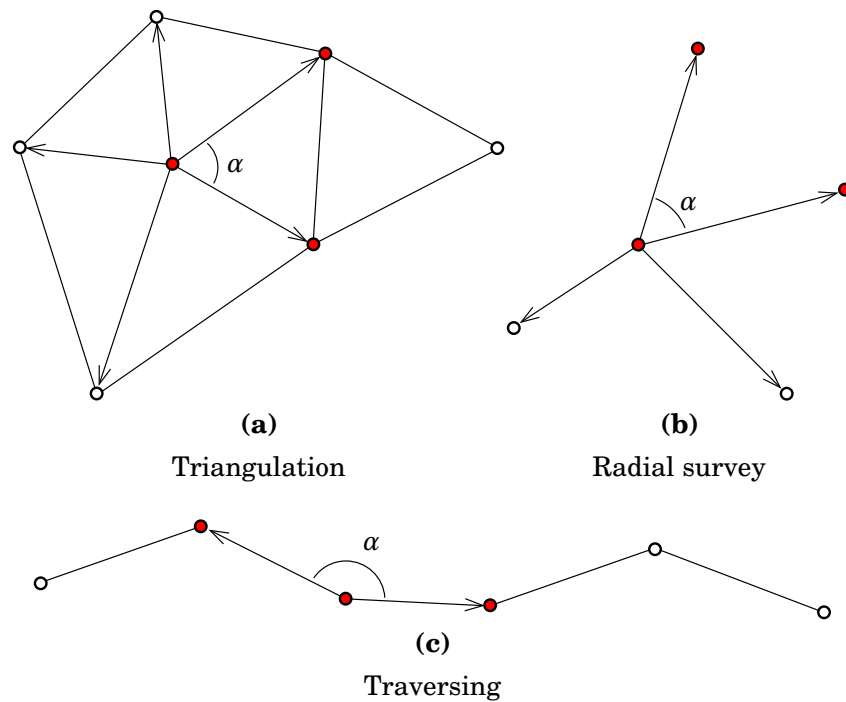


FIGURE 6.1. Use situations for horizontal angle measurement.

6.1c) in order to determine traverse point co-ordinates. To make calculation possible, the starting and closing points of the traverse need to be known, for example from a higher-order triangulation.

3. From the points of a traverse, *radial surveys*, for example, may be conducted, figures 6.1b and 8.10.

Many of these observation techniques have been replaced by GNSS measurement. Triangulation is no longer done anywhere, while traverse measurements are done only in situations where GNSS is not useable, such as underground — tunnels — or in high-rise urban landscapes.



6.2 Intersection and resection

- ¹ Geometries that are commonly used are *intersection* and *resection*,¹ figure 6.2.

Intersection works as follows: let the known points A and B have co-ordinates (x_A, y_A) and (x_B, y_B) . Let the perpendicular projection point between these two points of the unknown point C be P , and

¹In Finnish *eteen- ja taaksepäin leikkaus*, in Swedish *avskärning och inskärning*, in Dutch *voorwaartse en achterwaartse insnijding*, in German *Vorwärtsschnitt und Rückwärtsschnitt*.



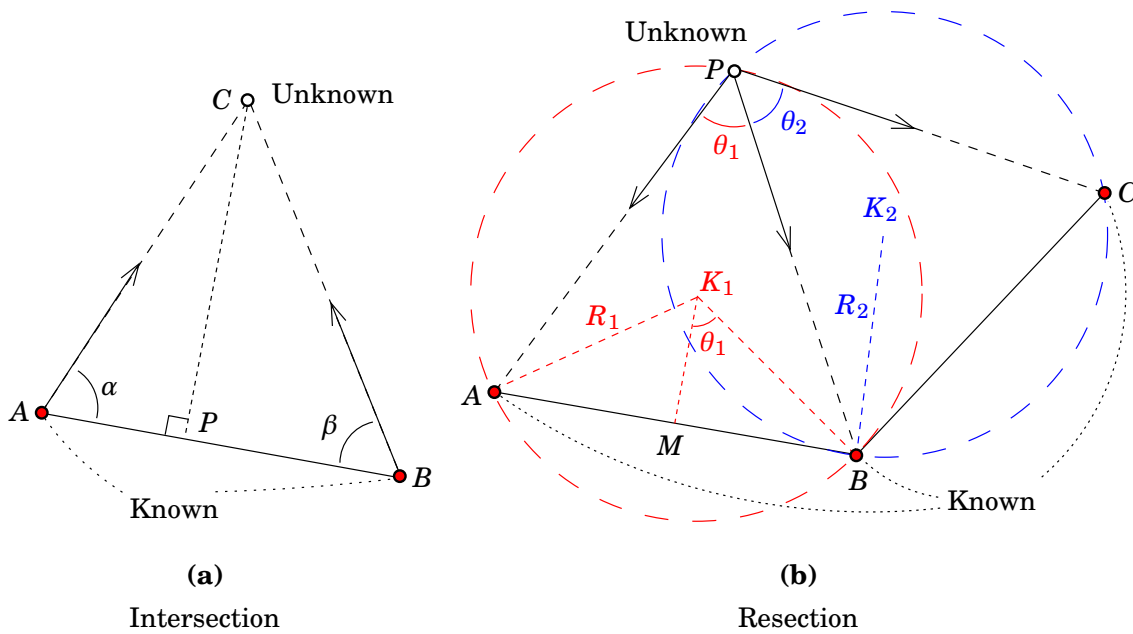


FIGURE 6.2. Intersection and resection.

its co-ordinates (x_P, y_P) . Then

$$PC = AP \tan \alpha = PB \tan \beta,$$

so

$$AB = AP + PB = PC \cot \alpha + PC \cot \beta = PC (\cot \alpha + \cot \beta).$$

From this

$$PC = \frac{AB}{\cot \alpha + \cot \beta}$$

and

$$AP = \frac{AB \cot \alpha}{\cot \alpha + \cot \beta}, \quad PB = \frac{AB \cot \beta}{\cot \alpha + \cot \beta}.$$

Now we use these distances, or equivalently, the coefficients $\cot \alpha$ and $\cot \beta$, as *weights* in calculating the co-ordinates of point P as weighted averages of the co-ordinates of points A and B :

$$x_P = \frac{x_A \cot \beta + x_B \cot \alpha}{\cot \alpha + \cot \beta}, \quad y_P = \frac{y_A \cot \beta + y_B \cot \alpha}{\cot \alpha + \cot \beta}.$$

After this, the direct calculation formulas for the co-ordinates of intersection point C are

$$x_C = x_P + \tan \alpha (y_P - y_A) = x_P + \frac{y_B - y_A}{\cot \alpha + \cot \beta},$$

$$y_C = y_P - \tan \alpha (x_P - x_A) = y_P - \frac{x_B - x_A}{\cot \alpha + \cot \beta}.$$



Resection is inherently more difficult, because it is an *inverse problem*.

The problem was studied as early as the 17th century by Snellius and Laurent Pothenot² and is often named after them.

Let (figure 6.2b) M be the midpoint of points A and B . If from the unknown point P the angle θ_1 between points A and B has been measured, then this point is on the circumference of a circle that passes through points A and B , and seen from the centre point K_1 of which the angle between points A and B is $2\theta_1$.

The line segment MK_1 stands perpendicularly on AB : the co-ordinates (x_M, y_M) of the midpoint M are calculated using the formulas derived above for the special case $\alpha = \beta$. The co-ordinates of point K_1 are calculated in the following way — remember that in the role of α we now have $90^\circ - \theta_1$:

$$\begin{aligned}x_{K_1} &= x_M + \frac{1}{2} \cot \theta_1 (y_B - y_A) = \frac{1}{2} (x_A + x_B) + \frac{1}{2} \cot \theta_1 (y_B - y_A), \\y_{K_1} &= y_M - \frac{1}{2} \cot \theta_1 (x_B - x_A) = \frac{1}{2} (y_A + y_B) - \frac{1}{2} \cot \theta_1 (x_B - x_A).\end{aligned}$$

The radius of the circle is obtained by the Pythagoras theorem:

$$R_1 = \frac{1}{2 \sin \theta_1} \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}.$$

Now, because there is also another point pair B and C through which runs a circle of its own, with centre point K_2 and radius R_2 , we have a pair of quadratic equations describing two circles:

$$(x - x_{K_1})^2 + (y - y_{K_1})^2 = R_1^2, \quad (x - x_{K_2})^2 + (y - y_{K_2})^2 = R_2^2.$$

From this, the co-ordinates (x, y) of the unknown resection point P can be solved — at least in principle. There are many ways of solving this, for example, linearisation with respect to a pair of approximate values (x_0, y_0) with an iterative solution.

These two circles have *two intersection points*, from which the right solution must be chosen, and it is not point B ...

Singularity The precision of resection depends on the geometry of the points. If points A , B , C , and the unknown point lie on the same circle, the solution is even impossible: we speak of a *singularity*. Then, the circles are identical, and any point on that circle serves as a solution. We also speak of the *dangerous circle*: already close to this geometry, precision deteriorates ominously.

²Laurent Pothenot (1650–1732) was a French mathematician and geodesist.



Note also that the situation is conceptually similar if two *distances* $R_1 = PK_1$ and $R_2 = PK_2$ are measured to the now *known* points K_1 and K_2 . Here, P is the unknown point, co-ordinates (x, y) , and a singularity occurs when K_1 , K_2 , and P are on a straight line.

When using distance measurement, we do not speak of intersection and resection, but they are similar situations.



6.3 Observation method of complete sets

The graduation of the horizontal circle of an optical theodolite has been manufactured with great care, but is always imprecise. Every gon of the graduation should be the same size, precisely one gon, but in reality is not. However, the sum of all the gons of the graduation will always be exactly 400 gon. This is why the effect of graduation errors can be diminished by measuring the same angle *using different sectors of the horizontal circle*, and taking the average of those. This principle is implemented in the *observation method of complete sets*.

vaakakehä

jaotusvirhe

sarjahavainto-
menetelmä
runkoverkko

In this method, traditionally used in the measurement of base networks, the same fan of directions is measured multiple times — complete sets — such that between the sets, the horizontal circle is *unlocked*³ and *turned* (by roughly $200^g/n$, in which n is the number of complete sets). In this way one makes use of the different sectors of the horizontal circle, and systematic errors, for example due to graduation errors of the horizontal circle, are minimised as one takes the average of the sets.

3

In this observation method, the instrument's telescope is also always turned over — both axes by 200^g — so one measures in both faces, face left and face right.

kojeasennot I
ja II

The following errors are eliminated by observing in two faces: collimation error and trunnion-axis tilt, axes intersection error, eccentricities of horizontal circle and measuring telescope. In the method of complete sets, the impact of horizontal-circle graduation errors is essentially diminished. Levelling or centring errors are *not* eliminated!

tasaus ja
keskistys

Rejection of observations:

- If, in the method of complete sets, the sets do not agree with each other, they must *all* be measured again. One may not reject only a

³The locking screw of the horizontal circle is protected by a lid. In that way, the circle is not turned by accident during measurement.



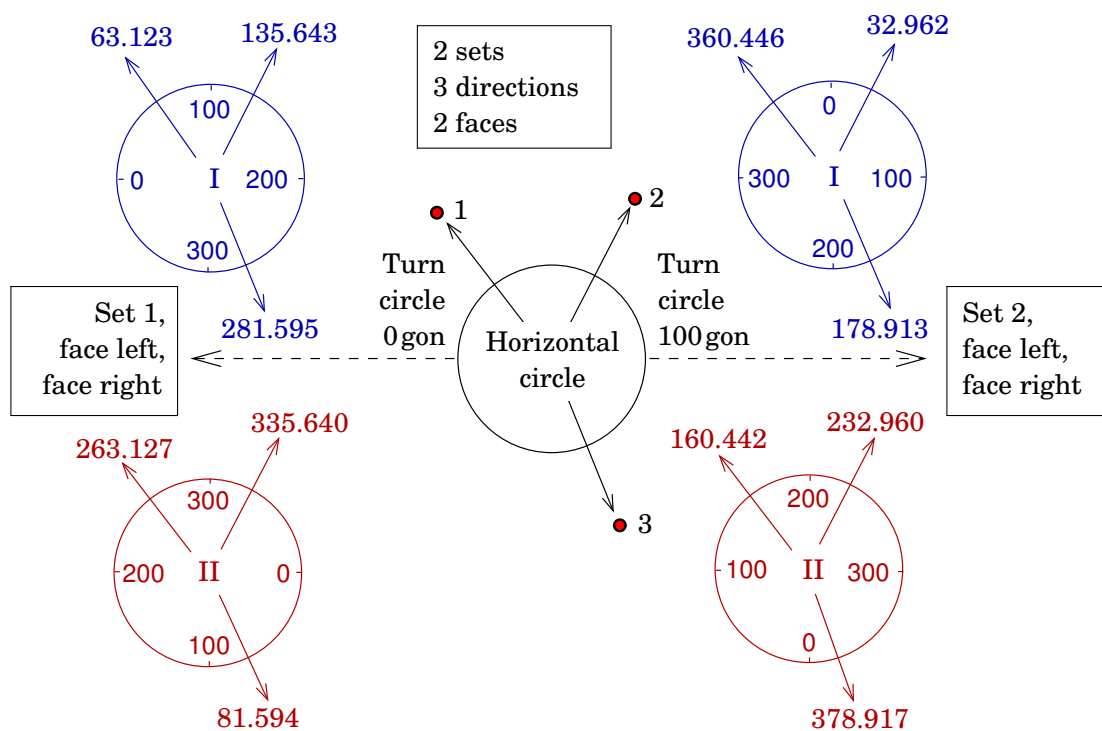


FIGURE 6.3. Observation method of complete sets. In this case a measurement of two complete sets of three directions is made in both instrument faces.



single set, because that would distort the statistical properties of the observational material.

- An individual sighting direction may be rejected from all sets.

One does not apply the method of complete sets with an electronic theodolite — or more precisely, the instrument does so itself internally, see section 5.7.



6.4 Station adjustment of horizontal angles

asematasoitus Station adjustment, the merging of the several complete sets of horizontal direction measurements into one optimal solution or measurement set, is an *adjustment problem*, although a simple one. Ordinarily it was already solved at the observation-notebook stage.

Modern electronic tacheometers or total stations do not need station adjustment as they do not collect multiple observation sets in the way described in section 6.3. We describe it here nevertheless briefly for historical and methodological interest. There will be a more thorough discussion of adjustment calculus in chapters 14 and 15.





6.4.1 Observation equations

In the following calculation example, four complete sets to three sighting directions have been measured. The two instrument faces for each series, face left and face right, have already been merged. In the example, the number of complete sets is $s = 4$, the number of directions $r = 3$, so the number of observations is $n = rs = 12$. The number of unknown directions in the complete-set average is $r - 1 = 2$. In each complete set there is one orientation unknown⁴ also to be estimated. The *number of excess observations* (redundancy, number of degrees of freedom) is thus

$$b = rs - (r - 1 + s) = (r - 1)(s - 1).$$

The observation equations are

$$\underline{\theta}_{ij} + \underline{v}_{ij} = \hat{\alpha}_j - \hat{\Omega}_i, \quad (6.1)$$

in which $i = 1, \dots, 4$ — generally $i = 1, \dots, s$ — is the number of the complete set, $j = 1, \dots, 3$ — generally $j = 1, \dots, r$ — is the number of the sighting direction, Ω_i the orientation unknown of the horizontal circle for the complete set, α_j the direction unknown of the target (the azimuth of the measured direction), and $\underline{\theta}_{ij}$ the (raw) direction reading.

We use the “hat” notation for estimators as is standard in statistics, and we underscore stochastic quantities.

If the vector of observations is

$$\underline{\ell} = \left[\underline{\theta}_{11} \quad \underline{\theta}_{12} \quad \underline{\theta}_{13} \mid \underline{\theta}_{21} \quad \underline{\theta}_{22} \quad \underline{\theta}_{23} \mid \underline{\theta}_{31} \quad \underline{\theta}_{32} \quad \underline{\theta}_{33} \mid \underline{\theta}_{41} \quad \underline{\theta}_{42} \quad \underline{\theta}_{43} \right]^T$$

and the vector of unknowns

$$\hat{\mathbf{x}} = \left[\hat{\alpha}_1 \quad \hat{\alpha}_2 \quad \hat{\alpha}_3 \mid \hat{\Omega}_1 \quad \hat{\Omega}_2 \quad \hat{\Omega}_3 \quad \hat{\Omega}_4 \right]^T,$$

then the *design matrix* is, for the case of four complete sets and three rakennematriisi

⁴The orientation unknown captures the reality that a theodolite *cannot measure absolute horizontal directions or azimuths*. It can measure only direction *differences*, that is, angles. Therefore the unknown azimuth of the zero mark on the instrument’s horizontal circle for every measured series is added as an unknown to the adjustment problem.

The situation is similar to that in GPS measurement. There, too, all observations from a receiver contain a common clock-error unknown ΔT , which is why the observable is called pseudo-range and not range. Similarly we could talk here about “pseudo-direction observations”. These uninteresting additional unknowns are called *nuisance parameters*.



sighting directions,

$$A = \left[\begin{array}{ccc|ccc} 1 & & & -1 & & \\ & 1 & & -1 & & \\ & & 1 & -1 & & \\ \hline 1 & & & & -1 & \\ & 1 & & & -1 & \\ & & 1 & & -1 & \\ \hline 1 & & & & & -1 \\ & 1 & & & & -1 \\ & & 1 & & & -1 \end{array} \right],$$

bringing the observation equations into the standard form

$$\underline{\ell} + \underline{v} = A\hat{x},$$

jäännös-
virheiden
vektori

with \underline{v} the vector of residuals.

In order to get a unique solution, one must also fix one direction, or linear combination of directions — one cannot solve absolute directions from relative angle measurements alone. So, in addition is required, say,

$$\alpha_1 = 0. \quad (6.2)$$

This is totally arbitrary. One could choose just as well $\alpha_2 = 0$, $\alpha_3 = 0$, or why not $\alpha_1 + \alpha_2 + \alpha_3 = 0$, which would at least be “democratic”.

In the following we present, without formal proof, how one may calculate the result of a station adjustment, and how the calculation can be arranged in a simple template. The procedure is statistically optimal.



6.4.2 Angle transformation

First, we carry out the *angle transformation*: we subtract from every observation $\underline{\theta}_{ij}$ which is not the first of a complete set, precisely that first observation of that set $\underline{\theta}_{i1}$:

$$\underline{\theta}'_{ij} \stackrel{\text{def}}{=} \underline{\theta}_{ij} - \underline{\theta}_{i1}, \quad i = 1, \dots, s, \quad j = 2, \dots, r.$$

Subtract the rows of the system of observation equations 6.1 from each other, yielding

$$\underline{\ell}' + \underline{v}' = A'\hat{x}$$

in which

$$\underline{\ell}' = \left[\begin{array}{cc|cc|cc|cc} \theta'_{12} & \theta'_{13} & \theta'_{22} & \theta'_{23} & \theta'_{32} & \theta'_{33} & \theta'_{42} & \theta'_{43} \end{array} \right]^T$$



and

$$A' = \left[\begin{array}{cc|c} -1 & 1 & 0 \\ -1 & & 1 & 0 \\ \hline -1 & 1 & & 0 \\ -1 & & 1 & 0 \\ \hline -1 & 1 & & 0 \\ -1 & & 1 & 0 \\ \hline -1 & 1 & & 0 \\ -1 & & 1 & 0 \end{array} \right] \implies A' = \left[\begin{array}{cc} -1 & 1 \\ -1 & 1 \\ \hline -1 & 1 \\ -1 & 1 \\ \hline -1 & 1 \\ -1 & 1 \\ \hline -1 & 1 \\ -1 & 1 \end{array} \right],$$

if we shorten the vector of unknowns accordingly:

$$\hat{x} = [\hat{\alpha}_1 \quad \hat{\alpha}_2 \quad \hat{\alpha}_3 \mid \hat{\Omega}_1 \quad \hat{\Omega}_2 \quad \hat{\Omega}_3 \quad \hat{\Omega}_4]^\top \implies \hat{x} = [\hat{\alpha}_1 \quad \hat{\alpha}_2 \quad \hat{\alpha}_3]^\top.$$

Now we may define the following implicit unknowns:

$$\hat{\alpha}'_j \stackrel{\text{def}}{=} \hat{\alpha}_j - \hat{\alpha}_1, \quad j = 2, \dots, r,$$

which corresponds to further simplifying the design matrix as follows:

$$A' = \left[\begin{array}{cc} -1 & 1 \\ -1 & 1 \\ \hline -1 & 1 \\ -1 & 1 \\ \hline -1 & 1 \\ -1 & 1 \\ \hline -1 & 1 \\ -1 & 1 \end{array} \right] \implies A'' = \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \\ \hline 1 & 1 \\ 1 & 1 \\ \hline 1 & 1 \\ 1 & 1 \end{array} \right]$$

if the new unknowns are

$$\hat{x}' = [\hat{\alpha}'_2 \quad \hat{\alpha}'_3]^\top.$$

These new unknowns are now in relation to the first direction, they are thus *angle-transformed* direction unknowns.

The observation equations are now

$$\begin{bmatrix} \theta'_{12} \\ \theta'_{13} \\ \theta'_{22} \\ \theta'_{23} \\ \theta'_{32} \\ \theta'_{33} \\ \theta'_{42} \\ \theta'_{43} \end{bmatrix} + \begin{bmatrix} v'_{12} \\ v'_{13} \\ v'_{22} \\ v'_{23} \\ v'_{32} \\ v'_{33} \\ v'_{42} \\ v'_{43} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \hat{\alpha}'_2 \\ \hat{\alpha}'_3 \end{bmatrix}$$



or

$$\begin{bmatrix} \underline{\theta}'_{i2} \\ \underline{\theta}'_{i3} \end{bmatrix} + \begin{bmatrix} \underline{v}'_{i2} \\ \underline{v}'_{i3} \end{bmatrix} = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} \begin{bmatrix} \hat{\alpha}'_2 \\ \hat{\alpha}'_3 \end{bmatrix} = \begin{bmatrix} \hat{\alpha}'_2 \\ \hat{\alpha}'_3 \end{bmatrix}, \quad i = 1, \dots, 4. \quad (6.3)$$

The optimal solution is the *average*, see equation 14.6. For the general case

$$\hat{\alpha}'_j = \frac{1}{s} \sum_{i=1}^s \theta'_{ij}, \quad j = 2, \dots, r.$$

The adjusted angle transformation is the average over complete sets of the angle transformations.

Furthermore we may define consistently

$$\hat{\alpha}'_1 \stackrel{\text{def}}{=} \frac{1}{s} \sum_{i=1}^s \theta'_{i1} \stackrel{\text{def}}{=} \frac{1}{s} \sum_{i=1}^s (\theta_{i1} - \underline{\theta}_{i1}) = 0.$$

tasoitus The α' values are adjusted but still relative directions, not absolute azimuths.

We have eliminated the orientation unknowns Ω_i from the observation equations, and good riddance. We are however interested in the residuals.



6.4.3 Residuals and degrees of freedom

pienimmän
neliösumman
menetelmä
jäännösvirhe

Now that the unknowns $\hat{\alpha}'_j$ have been calculated, more precisely *estimated*, with the least-squares method, we can also calculate the *residuals*. Equation 6.3 gives us residuals of sorts:

$$\underline{v}'_{ij} = \hat{\alpha}'_j - \theta'_{ij}, \quad i = 1, \dots, s, \quad j = 2, \dots, r.$$

These have the problem that direction 1 is not included: if it were, its residual would always be

$$v'_{i1} = \hat{\alpha}'_1 - \theta'_{i1} = 0 - 0 = 0.$$

The proper residuals are obtained by

$$\underline{v}_{ij} = \underline{v}'_{ij} - \frac{1}{r} \sum_{k=1}^r \underline{v}'_{ik}, \quad i = 1, \dots, s, \quad j = 1, \dots, r, \quad (6.4)$$

of which the average over directions, including direction 1, for each set vanishes. Calculation 6.4 shows what happens to the residuals when



every complete set $i = 1, \dots, s$ is assigned its own orientation unknown Ω_i , minimising the sum of squared residuals within that set.

Let the mean error σ of the direction observations $\underline{\ell}_{ij}$ be the same for all, and let the direction observations be uncorrelated. In that case the *variance matrix* of the observations is

$$\Sigma_{\ell\ell} = \text{Var}\{\underline{\ell}\} = \sigma^2 I_{s \times r},$$

in which $I_{s \times r}$ is the unit matrix of size $(s \times r) \times (s \times r)$.

In this case, the quantity (“shifting variate”, [Baarda, 1968](#))

siirtosuure

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \underline{v}^T \Sigma_{\ell\ell}^{-1} \underline{v} = \frac{1}{\sigma^2} \sum_{i=1}^s \sum_{j=1}^r v_{ij}^2 = \frac{1}{\sigma^2} \sum_{i,j} v_{ij}^2$$

is distributed according to the χ_b^2 *distribution*, the expectancy of which is b , the number of *degrees of freedom*, or “excess” observations. See section [15.4](#).

odotusarvo

Earlier we saw that $b = (r-1)(s-1)$, the number of angle transformations (independent direction measurements) multiplied by the number of *excess* complete sets. Thus, $\underline{\mathcal{E}}/\sigma^2$ is distributed according to $\chi_{(r-1)(s-1)}^2$ and its expectancy is $(r-1)(s-1)$. So:

$$E\left\{\frac{1}{\sigma^2} \sum_{i,j} v_{ij}^2\right\} = (r-1)(s-1) \implies \sigma^2 = E\left\{\frac{\sum_{i,j} v_{ij}^2}{(r-1)(s-1)}\right\},$$

and we see that

$$\widehat{\sigma^2} \stackrel{\text{def}}{=} \frac{1}{(r-1)(s-1)} \sum_{i=1}^s \sum_{j=1}^r v_{ij}^2$$

is an unbiased *estimator* of the variance σ^2 of a single observation:

harhaton
estimaattori

$$E\{\widehat{\sigma^2}\} = \sigma^2.$$

For the quantity σ we use the name “*mean error of unit weight*”. It represents the uncertainty, the mean error, of a typical observation, in this case, a single, raw direction reading.

painoyksikön
keskivirhe

We shall return to the subject of adjustment in chapter [14](#).

6.4.4 Calculation table for station adjustment

A station adjustment consists of the following steps:

asematasoitus



sarjakeskiarvo

1. We fix the arbitrariness in the definition of the directions, like by fixing the first direction to zero, $\theta'_{i1} = 0$; this is called the *angle transformation*. We do it by subtracting the first observed direction reading from the other direction readings in the same complete set.
2. We compute the *average over all complete sets* (“complete-set average”)

$$\frac{1}{s} \sum_{i=1}^s \theta'_{ij}, \quad j = 1, \dots, r$$

for every sighting direction j .

3. We compute the residuals of the observations relative to these *complete-set averages*:

$$\underline{v}'_{ij} = \theta'_{ij} - \frac{1}{s} \sum_{k=1}^s \theta'_{kj}, \quad i = 1, \dots, s, \quad j = 1, \dots, r.$$

4. We compute the averages of the “first residuals” \underline{v}'_{ij} over all directions within every complete set:

$$\frac{1}{r} \sum_{j=1}^r \underline{v}'_{ij} = \frac{1}{r} \sum_{j=1}^r \left(\theta'_{ij} - \frac{1}{s} \sum_{k=1}^s \theta'_{kj} \right), \quad i = 1, \dots, s.$$

5. We compute the final residuals

$$\underline{v}_{ij} = \underline{v}'_{ij} - \frac{1}{r} \sum_{k=1}^r \underline{v}'_{ik}, \quad i = 1, \dots, s, \quad j = 1, \dots, r,$$

which corresponds precisely to those residuals we would obtain *if* we gave each complete set its own orientation unknown $\widehat{\Omega}_i$ in addition to the direction unknowns \widehat{a}_j , like in the original system of observation equations.

The end result is the *calculation template* presented in table 6.1, which contains these operations and ought to be self evident.

After filling in the template — in the order: average over complete sets, v' , sum per series, v — the mean error of a single direction, or *unit weight*, is calculated by first calculating the quadratic sum of the residuals over all series and directions:

$$\underline{v}^T \underline{v} = \sum_{i=1}^s \sum_{j=1}^r \underline{v}_{ij}^2 = [vv]$$





TABLE 6.1. Calculation template table for station adjustment.

	Angle transformation (gon)	v' (cc)	v (cc)
1. Complete set	0.0000	0	-3.7
	68.8430	+4	+0.3
	209.1880	+7	+3.3
	Complete set sum, average	+11	+3.7
2. Complete set	0.0000	0	-1.3
	68.8425	-1	-2.3
	209.1878	+5	+3.7
	Complete set sum, average	+4	+1.3
3. Complete set	0.0000	0	+2.3
	68.8424	-2	+0.3
	209.1868	-5	-2.7
	Complete set sum, average	-7	-2.3
4. Complete set	0.0000	0	+2.7
	68.8425	-1	+1.7
	209.1866	-7	-4.3
	Complete set sum, average	-8	-2.7
Average over complete sets	0.0000		
("complete-set average")	68.8426		
	209.1873		

(the latter notation was invented by Gauss and is sometimes still used).

Then, we calculate, or *estimate*⁵

5

$$\hat{\sigma} = \sqrt{\widehat{\sigma^2}} = \sqrt{\frac{\underline{v}^T \underline{v}}{rs - (r+s) + 1}} = \sqrt{\frac{\underline{v}^T \underline{v}}{(r-1)(s-1)}} = 3.8^{cc}.$$

Here

- rs is the number of observations (r directions in s complete sets).
- $r + s$ is the number of unknowns (r direction unknowns α_j , s orientation unknowns Ω_i).
- The number $+1$ represents the *defect* of the problem, the dimension of the solution space: the circumstance mentioned above, that

⁵Note that $E\{\widehat{\sigma^2}\} = \sigma^2$, however $E\{\hat{\sigma}\} \neq \sigma$! So $\hat{\sigma}$ is not unbiased although $\widehat{\sigma^2}$ is. The reason for this is that the E operator and the square root may not be interchanged: in general, $E\{\sqrt{\underline{x}}\} \neq \sqrt{E\{\underline{x}\}}$.

Counter example: if the probability distribution of \underline{x} is $p(1) = 0.5$, $p(4) = 0.5$, and elsewhere $p(x) = 0$, we obtain $E\{\sqrt{\underline{x}}\} = 1.5$ but $\sqrt{E\{\underline{x}\}} = \sqrt{2.5} \approx 1.58$.



adding an amount Δ to both the directions α_j and the orientations Ω_i does not affect the observations in any way that can be observed, and that thus, the solution is not unique without *one* extra condition,⁶ for example, fixing α_1 to some value, such as zero.

The number $b = rs - (r + s) + 1 = (r - 1)(s - 1)$ is called the *number of degrees of freedom*.

After this, the estimate of the mean error of the average taken over all complete sets is $\hat{\sigma}_s = \hat{\sigma} / \sqrt{s} = 1.9^{\text{cc}}$.



6.5 Traverse measurement and computation

jonomittaus
tosiaikainen

A traditional method for densifying lower-order base networks has been traverse measurement or *traversing*. In spite of the existence of satellite positioning, and especially real-time kinematic positioning (RTK), there continue to be situations where traversing is the best, or even the only, method of base-network measurement. Such situations include indoors measurements, underground — mine or tunnel — measurements, and measurements in high-rise urban landscapes, “*urban canyons*”, where GNSS measurement has problems due to the blocking of signal by buildings, and spurious signal reflections off buildings, the *multipath* problem.

monitie

⁷ Going back in history, then, in 1807 Nathaniel Bowditch⁷ won a competition for the most appropriate way to compute a traverse. The prize was ten US dollars (Cooper, 1982, pages 147–150).

The variant of the Bowditch method presented here, a traditional separate adjustment of directions and co-ordinates, is suboptimal. Yet it is worth looking at as an example of geodetic computation.

monikulmiojono

A *traverse* is a line of measurement stations, where at every station one measures directions and distances to the previous and next stations. The instrument used is generally a *total station* or tacheometer. The measurement is usually the last stage in bringing co-ordinates (x, y) to reference points in the immediate vicinity of objects to be measured.

There are two types of traverses: *open* and *closed*. See figure 6.4. We

⁶This extra condition can also be understood as a “pseudo-observation”: then, the number of observations is $rs + 1$.

⁷Nathaniel Bowditch (1773–1838) was an American mathematician, student of navigation at sea, and scientific translator and textbook author. One of his writings was “*The American Practical Navigator*”, published in 1802. The modern edition of the book, “*The Bowditch*”, can be found on the Internet: Bowditch (2017).



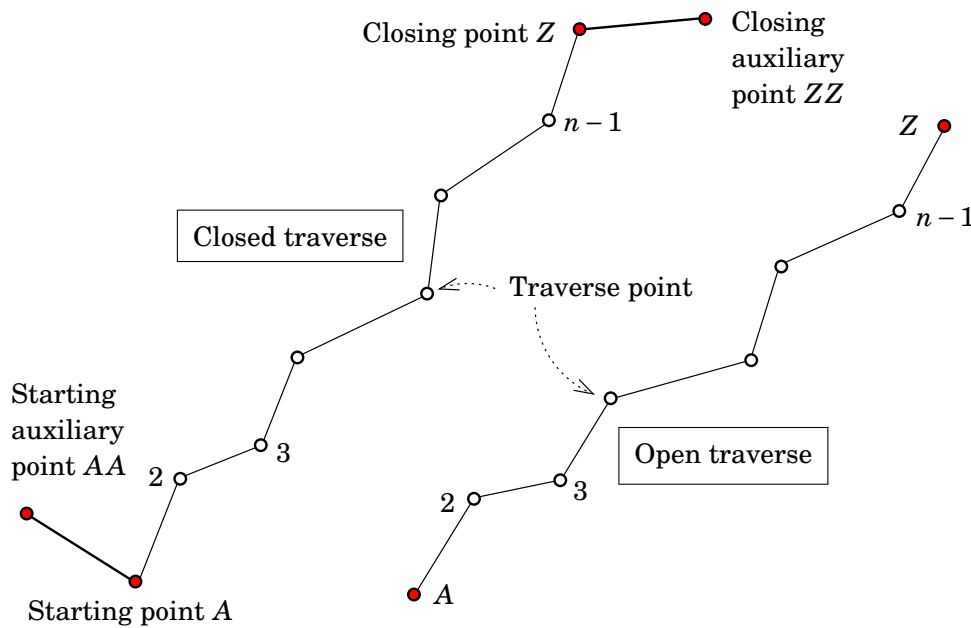


FIGURE 6.4. An open and a closed traverse.

always know the co-ordinates (x, y) of the starting and closing points. The difference between an open and a closed traverse is in the use of *auxiliary points* in the latter. These points, the co-ordinates of which are known, help to orient the traverse correctly.

liitospiste

vallinaisesti suljettu

6.6 Open traverse

Generally, we try to measure a closed traverse, because it gives the possibility to check and adjust both the measured angles and distances. There are, however, situations where this is impossible or difficult and where one has to use an *incompletely closed* traverse.

The common case is where we know neither the starting nor the closing direction. See figure 6.5.

We know

- (x_A, y_A) , starting point
- (x_Z, y_Z) , closing point.

We need to compute, or *estimate*,⁸ co-ordinates for the traverse points (\hat{x}_i, \hat{y}_i) , $i = 2, \dots, n - 1$.

8

⁸Using again the “hat” notation for estimators, as well as underscoring stochastic quantities.



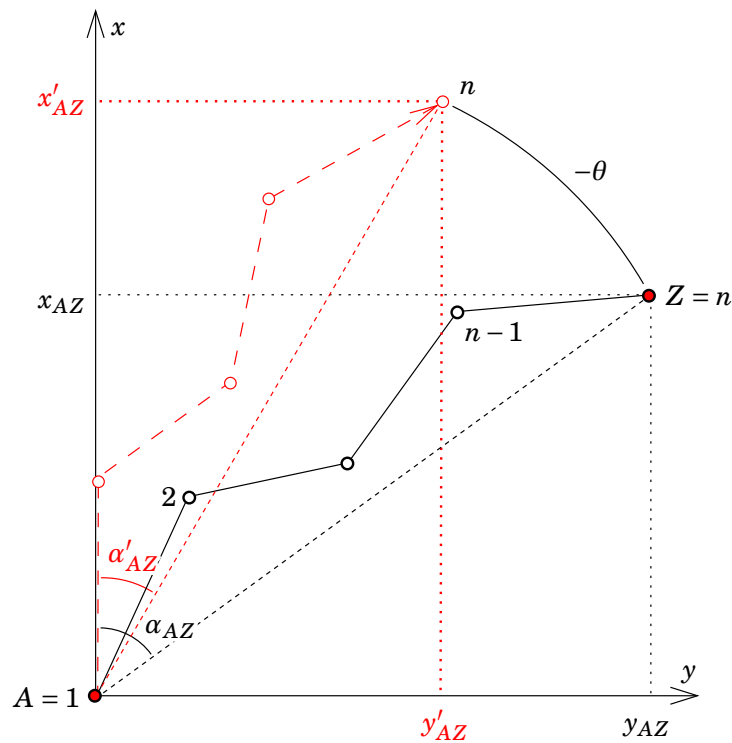


FIGURE 6.5. An open traverse.



6.6.1 Starting direction

Because we do not know the starting or closing directions, we cannot adjust directions. Computing the direction α_{12} is not possible as we lack a starting direction. We may obtain one in two different ways:

- by setting the direction angle of the first side $\alpha'_{12} = 0$, figure 6.5, or
- (preferred) obtaining somehow, for example by measuring in a map, an *approximate value* α'_{12} for α_{12} .

likiarvo



6.6.2 Computing the traverse

We carry out the calculation of traverse directions in the familiar way:

$$\begin{aligned}\alpha_{12} &= \alpha'_{12}, \\ \alpha_{23} &= \alpha_{12} - 200^g + \angle \alpha_2, \\ &\vdots \\ \alpha_{i,i+1} &= \alpha_{i-1,i} - 200^g + \angle \alpha_i, \\ &\vdots \\ \alpha_{n-1,n} &= \alpha_{n-2,n-1} - 200^g + \angle \alpha_{n-1}.\end{aligned}$$

geodeettinen
päätehtävä

Using these, we calculate co-ordinates for the points using the forward



geodetic problem recursively:

$$\underline{x}'_2 = x_A + \underline{s}_{12} \cos \underline{\alpha}_{12}, \quad \underline{y}'_2 = y_A + \underline{s}_{12} \sin \underline{\alpha}_{12},$$

and so forth:

$$\underline{x}'_i = \underline{x}'_{i-1} + \underline{s}_{i-1,i} \cos \underline{\alpha}_{i-1,i}, \quad \underline{y}'_i = \underline{y}'_{i-1} + \underline{s}_{i-1,i} \sin \underline{\alpha}_{i-1,i}.$$

Finally

$$\underline{x}'_n = \underline{x}'_{n-1} + \underline{s}_{n-1,n} \cos \underline{\alpha}_{n-1,n}, \quad \underline{y}'_n = \underline{y}'_{n-1} + \underline{s}_{n-1,n} \sin \underline{\alpha}_{n-1,n}.$$

This calculation has been carried out in the correct way, but *in a wrongly oriented co-ordinate frame*. The closing errors

$$\underline{w}_x \stackrel{\text{def}}{=} \underline{x}'_n - x_Z, \quad \underline{w}_y \stackrel{\text{def}}{=} \underline{y}'_n - y_Z$$

do not tell us anything about measurement errors, but rather about the approximateness of the assumed starting direction α'_{12} . For this reason, it is not permissible to eliminate the closing errors by adjustment as is done in a closed traverse.

The whole traverse is *rotated* by an angle amount θ , and should be rotated back by a Helmert or similarity transformation.

Because starting point A is the common turning point between both co-ordinate systems, we may simply calculate the scale ratio \underline{K} and rotation angle $\underline{\theta}$ between them: kiertokulma

$$\underline{K} = \frac{\sqrt{\Delta x_{AZ}^2 + \Delta y_{AZ}^2}}{\sqrt{(\Delta \underline{x}'_{AZ})^2 + (\Delta \underline{y}'_{AZ})^2}},$$

$$\underline{\theta} = \arctan \frac{\Delta \underline{y}'_{AZ}}{\Delta \underline{x}'_{AZ}} - \arctan \frac{\Delta y_{AZ}}{\Delta x_{AZ}}$$

— or with the half-angle formula, with the definition

$$y//x \stackrel{\text{def}}{=} \frac{y}{x + \sqrt{x^2 + y^2}} :$$

$$\underline{\theta} = 2 \left(\arctan(\Delta \underline{y}'_{AZ} // \Delta \underline{x}'_{AZ}) - \arctan(\Delta y_{AZ} // \Delta x_{AZ}) \right).$$

Also used is the notation

$$\begin{aligned} \Delta x_{AZ} &= x_Z - x_A, & \Delta y_{AZ} &= y_Z - y_A, \\ \Delta \underline{x}'_{AZ} &= \underline{x}'_n - x_A, & \Delta \underline{y}'_{AZ} &= \underline{y}'_n - y_A. \end{aligned}$$



Let us construct a transformation:

$$\underline{c} = \underline{K} \cos \theta, \quad \underline{s} = \underline{K} \sin \theta,$$

with the help of which we obtain as the co-ordinates of the traverse points $i = 1, \dots, n$:

$$\begin{aligned} \widehat{x}_i &= x_A + \underline{c}(x'_i - x_A) + \underline{s}(y'_i - y_A), \\ \widehat{y}_i &= y_A - \underline{s}(x'_i - x_A) + \underline{c}(y'_i - y_A). \end{aligned}$$

In matrix form

$$\begin{bmatrix} \widehat{x}_i \\ \widehat{y}_i \end{bmatrix} = \begin{bmatrix} x_A \\ y_A \end{bmatrix} + \begin{bmatrix} \underline{c} & \underline{s} \\ -\underline{s} & \underline{c} \end{bmatrix} \begin{bmatrix} x'_i - x_A \\ y'_i - y_A \end{bmatrix}.$$

Note that in the matrix

$$\underline{M} = \begin{bmatrix} \underline{c} & \underline{s} \\ -\underline{s} & \underline{c} \end{bmatrix} = \underline{K} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

the direction correction θ is dominated by the assumed starting direction, when again the scale correction \underline{K} contains only a correction for the closing error caused by the imprecision of the measurements:⁹ even if the starting direction α_{12} were guessed exactly correctly, nevertheless $(\widehat{x}'_n, \widehat{y}'_n)$ would not necessarily coincide precisely with point (x_Z, y_Z) . In the sideways direction however, perpendicular to the traverse, there will remain observation error, which is not adjusted away, but more precisely is “absorbed” into θ .

A useful final sanity check is

$$x_Z \stackrel{?}{=} \widehat{x}_n, \quad y_Z \stackrel{?}{=} \widehat{y}_n.$$

In addition to this, one should check that \underline{K} is realistically valued, meaning close enough to the value 1 in view of the known precision of distance measurement.



6.7 Closed traverse

We know

alkuliitospiste

- (x_A, y_A) and (x_{AA}, y_{AA}) : starting point A and auxiliary starting point AA,

loppuliitospiste

- (x_Z, y_Z) and (x_{ZZ}, y_{ZZ}) : closing point Z and auxiliary closing point ZZ.

⁹... in other words, it is a primitive *adjustment*.



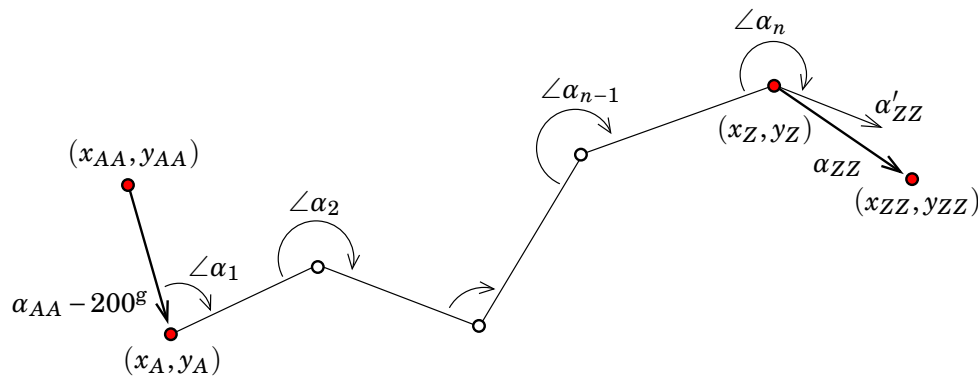


FIGURE 6.6. The geometry of a closed traverse. Only angles and starting and closing directions are marked.

See figure 6.6.

We observe

- n bending angles $\angle \alpha_1, \dots, \angle \alpha_n$,
- $n - 1$ distances or side lengths $s_{12}, s_{23}, \dots, s_{n-1,n}$.

We must compute, or estimate

the co-ordinates (\hat{x}_i, \hat{y}_i) for the traverse points $i = 2, \dots, n - 1$. jonopiste

There are $n - 2$ new points.

6.7.1 Computing the auxiliary starting and closing directions

The *auxiliary starting and closing directions* are obtained by solving the inverse geodetic problem:

alku- ja loppu-
liitossuunta
geodeettinen
käänteistehtävä

$$\begin{aligned} \alpha_{AA} &= \arctan \frac{y_{AA} - y_A}{x_{AA} - x_A} + k \cdot 200^g, \\ \alpha_{ZZ} &= \arctan \frac{y_{ZZ} - y_Z}{x_{ZZ} - x_Z} + k \cdot 200^g, \end{aligned} \tag{6.5}$$

where $k \in \{0, 1\}$ needs to be chosen so the result is in the correct quadrant.

The approximate Bowditch-like method that we shall use has two stages. First we process the *directions*, then the *co-ordinates*.

6.7.2 Adjustment of directions

Compute the directions or *azimuths* $\alpha_{AA}, \alpha_{12}, \dots, \alpha_{n-1,n}, \alpha_{n,n+1}$, in which α_{AA} is known, and $\alpha_{n,n+1}$, which is computed from the observations, corresponds to the above computed auxiliary closing direction α_{ZZ} . As follows.¹⁰

10



$$\begin{aligned}
\underline{\alpha}_{12} &= \alpha_{AA} + \angle \underline{\alpha}_1, \\
\underline{\alpha}_{23} &= \underline{\alpha}_{12} - 200^g + \angle \underline{\alpha}_2, \\
&\vdots \\
\underline{\alpha}_{i,i+1} &= \underline{\alpha}_{i-1,i} - 200^g + \angle \underline{\alpha}_i, \\
&\vdots \\
\underline{\alpha}_{n-1,n} &= \underline{\alpha}_{n-2,n-1} - 200^g + \angle \underline{\alpha}_{n-1}, \\
\underline{\alpha}'_{ZZ} &\stackrel{\text{def}}{=} \underline{\alpha}_{n,n+1} = \underline{\alpha}_{n-1,n} - 200^g + \angle \underline{\alpha}_n.
\end{aligned} \tag{6.6}$$

At the end we thus obtain the value $\underline{\alpha}'_{ZZ}$, which, if all the angle observations $\underline{\alpha}_i$ were errorless, would be equal to α_{ZZ} computed from co-ordinates by equation 6.5. In reality though, the *direction closing error* amounts to

$$\underline{w}_\alpha \stackrel{\text{def}}{=} \underline{\alpha}'_{ZZ} - \alpha_{ZZ}.$$

This closing error is removed, or *adjusted*, by dividing it evenly among all the bending angles, so to every measured bending angle we apply a *correction*:

$$\delta \underline{\alpha} = -\frac{\underline{w}_\alpha}{n},$$

after which we again carry out the calculation of angles, producing the *adjusted directions*:

$$\begin{aligned}
\hat{\alpha}_{12} &= \alpha_{AA} + (\angle \underline{\alpha}_1 + \delta \underline{\alpha}), \\
\hat{\alpha}_{23} &= \underline{\alpha}_{12} - 200^g + (\angle \underline{\alpha}_2 + \delta \underline{\alpha}), \\
&\vdots \\
\hat{\alpha}_{i,i+1} &= \underline{\alpha}_{i-1,i} - 200^g + (\angle \underline{\alpha}_i + \delta \underline{\alpha}), \\
&\vdots \\
\hat{\alpha}_{n-1,n} &= \underline{\alpha}_{n-2,n-1} - 200^g + (\angle \underline{\alpha}_{n-1} + \delta \underline{\alpha}), \\
\hat{\alpha}'_{ZZ} &= \underline{\alpha}_{n-1,n} - 200^g + (\angle \underline{\alpha}_n + \delta \underline{\alpha}).
\end{aligned}$$

6.7.3 Co-ordinate adjustment

Using the thusly adjusted directions we compute, for the points $2, \dots, n-1$, co-ordinates by applying the forward geodetic problem:

- Point 2:

$$\Delta x_{12} = s_{12} \cos \hat{\alpha}_{12}, \quad \Delta y_{12} = s_{12} \sin \hat{\alpha}_{12},$$

¹⁰For the first bending angle $\angle \underline{\alpha}_1$ we do *not* subtract 200 gon! And to keep the directions $\underline{\alpha}_{i,i+1}$ in the interval $[0, 400^g)$, one may sometimes have to add or subtract 400 gon.



with the aid of which

$$\underline{x}_2 = x_A + \Delta \underline{x}_{12}, \quad \underline{y}_2 = y_A + \Delta \underline{y}_{12}.$$

- General point i (and also closing point $i \rightarrow n$):

$$\Delta \underline{x}_{i-1,i} = \underline{s}_{i-1,i} \cos \widehat{\alpha}_{i-1,i}, \quad \Delta \underline{y}_{i-1,i} = \underline{s}_{i-1,i} \sin \widehat{\alpha}_{i-1,i},$$

with the aid of which

$$\underline{x}_i = \underline{x}_{i-1} + \Delta \underline{x}_{i-1,i} = \underline{x}_{i-1} + \underline{s}_{i-1,i} \cos \widehat{\alpha}_{i-1,i},$$

$$\underline{y}_i = \underline{y}_{i-1} + \Delta \underline{y}_{i-1,i} = \underline{y}_{i-1} + \underline{s}_{i-1,i} \sin \widehat{\alpha}_{i-1,i}.$$

- Generally — remember $(x_A, y_A) = (x_1, y_1)$:

$$\underline{x}_k = x_A + \sum_{i=2}^k \underline{s}_{i-1,i} \cos \widehat{\alpha}_{i-1,i}, \quad \underline{y}_k = y_A + \sum_{i=2}^k \underline{s}_{i-1,i} \sin \widehat{\alpha}_{i-1,i}.$$

By substitution of $k \rightarrow n$ we obtain the equations and co-ordinates $(\underline{x}_n, \underline{y}_n)$ of the closing point. If the observations were errorless, we would have $\underline{x}_n = x_Z$ and $\underline{y}_n = y_Z$, but they are not. The *co-ordinate closing errors* are

$$\underline{w}_x \stackrel{\text{def}}{=} \underline{x}_n - x_Z, \quad \underline{w}_y \stackrel{\text{def}}{=} \underline{y}_n - y_Z.$$

Closing errors are adjusted by giving a *weight coefficient* $q_{i-1,i}$ for each point interval, and the closing errors \underline{w}_x and \underline{w}_y are distributed over the point intervals in proportion to these weight coefficients.

Small weight coefficient \longleftrightarrow large weight,

large weight coefficient \longleftrightarrow small weight!

We compute the *standard correction* corresponding to the sum of weight coefficients, separately for the x and y co-ordinates:

$$\delta \underline{x} = -\frac{\underline{w}_x}{\sum_{i=2}^n q_{i-1,i}}, \quad \delta \underline{y} = -\frac{\underline{w}_y}{\sum_{i=2}^n q_{i-1,i}}.$$

The adjusted co-ordinates become

$$\widehat{x}_i = \widehat{x}_{i-1} + \underline{s}_{i-1,i} \cos \widehat{\alpha}_{i-1,i} + q_{i-1,i} \delta \underline{x},$$


$$\widehat{y}_i = \widehat{y}_{i-1} + \underline{s}_{i-1,i} \sin \widehat{\alpha}_{i-1,i} + q_{i-1,i} \delta \underline{y},$$

or, counting from starting point A ,

$$\widehat{x}_k = \widehat{x}_A + \sum_{i=2}^k (\underline{s}_{i-1,i} \cos \widehat{\alpha}_{i-1,i} + q_{i-1,i} \delta \underline{x}),$$

$$\widehat{y}_k = \widehat{y}_A + \sum_{i=2}^k (\underline{s}_{i-1,i} \sin \widehat{\alpha}_{i-1,i} + q_{i-1,i} \delta \underline{y}).$$



 TABLEAU 6.2. Traverse calculation template, separate adjustment of angles and co-ordinates (“Bowditch method”).

i	$\angle \alpha_i$	$\underline{\alpha}_{i,i+1}$	$\delta \alpha_{i,i+1}$	$\hat{\alpha}_{i,i+1}$						
A (= 1)		<u>345.3750</u>								
1	212.2345	157.6095	+7	157.6102						
2	151.4565	109.0660	+15	109.0675						
3	221.9823	131.0483	+22	131.0505						
4	175.9831	107.0314	+29	107.0343						
5	165.3467	72.3781	+37	72.3818						
Z (= 5)		<u>72.3818</u>								
Closing error \underline{w}_α		-0.0037	↑							

i	$\hat{\alpha}_{i,i+1}$	$\underline{s}_{i,i+1}$	$\underline{s} \cos \hat{\alpha}$	$\underline{s} \sin \hat{\alpha}$	\underline{x}_i	\underline{y}_i	δx_i	δy_i	\hat{x}_i	\hat{y}_i
1	157.6102	502.345	-395.038	+310.315	<u>1000.235</u>	<u>256.256</u>				
2	109.0675	487.241	-69.164	+482.307	605.197	566.571	+9	-37	.206	.534
3	131.0505	445.981	-209.001	+393.977	536.033	1048.878	+18	-73	.051	.805
4	107.0343	512.125	-56.472	+509.002	327.032	1442.855	+26	-110	.058	.745
5	72.3818				270.560	1951.857	+35	-146	.595	.711
Z (= 5)					<u>270.595</u>	<u>1951.711</u>				
Closing errors $\underline{w}_x, \underline{w}_y$					-0.035	+0.146	↑	↑		

6.7.4 Calculation template

The whole calculation may be carried out using the template in tableau 6.2, which is readily automated. Let it be given that $\alpha_{AA} = 345.3750$ gon and $\alpha_{ZZ} = 72.3818$ gon, as well as the co-ordinates of starting point A and closing point Z. The number of stations $n = 5$. Weighting used: uniform. The values for the starting and closing points are underlined, observations are in black, computed values in red, closing errors and adjustment corrections in blue. This kind of diagram is good enough for practical work in local base-network measurement.

Some remarks:

- The adjustment method described above is *approximate*, meaning sub-optimal. In a proper least-squares adjustment, the direction and co-ordinate corrections are computed simultaneously, as they *correlate* with each other. The results will then be slightly different. Nevertheless, if the traverse is relatively straight — the sides are

pienimmän
neliösumman
tasoitus



roughly in the same direction — the suboptimality will be small, and a proper adjustment would only lead to small changes in the final co-ordinates.

On the other hand, it is not desirable that the sides are in *precisely* the same direction: that would make finding gross errors more difficult. karkea virhe

- Here, the calculation of the traverse was done *in the plane*. The measurements however have been obtained three-dimensionally, in space. Of course it is assumed that instrument and signal are precisely and correctly centred and that the instrument is levelled, and that all relevant reductions to the observations have been done. tähy This includes especially the slope reduction, equation 7.9. Also, depending on size and topography of the measurement area, the height reduction (same equation 7.9) may be necessary.
- In computing plane co-ordinates, directions are always referred to the *map north*, and distances are also reduced to the map plane, meaning that the *scale distortion* caused by the map projection used, equation 7.10, is accounted for.
- Only then can it be said that
 - The horizontal angles are plane angles.
 - The reduced distances are horizontal distances in the map plane.

6.8 Zenith angles and refraction

In the measurement of zenith angles, because of the stratification of the atmosphere, the overwhelmingly greatest source of error is *refraction*. kerrostuneisuus On must always be careful concerning weather and the landscape when choosing time and place, and when processing observations. For example, hot *asphalt* in summer can be treacherous.

The effect of refraction does not show in the check $\zeta_1 + \zeta_2 = 400$, subsection 5.6.3: measuring in both faces does not help. kojeasennot

6.8.1 Refraction coefficient

In geodesy, the habit has formed to characterise refraction by a quantity named the *refraction coefficient*, symbol k . This quantity characterises the refraktiokerroin curvature of the measurement ray or path in the atmosphere compared



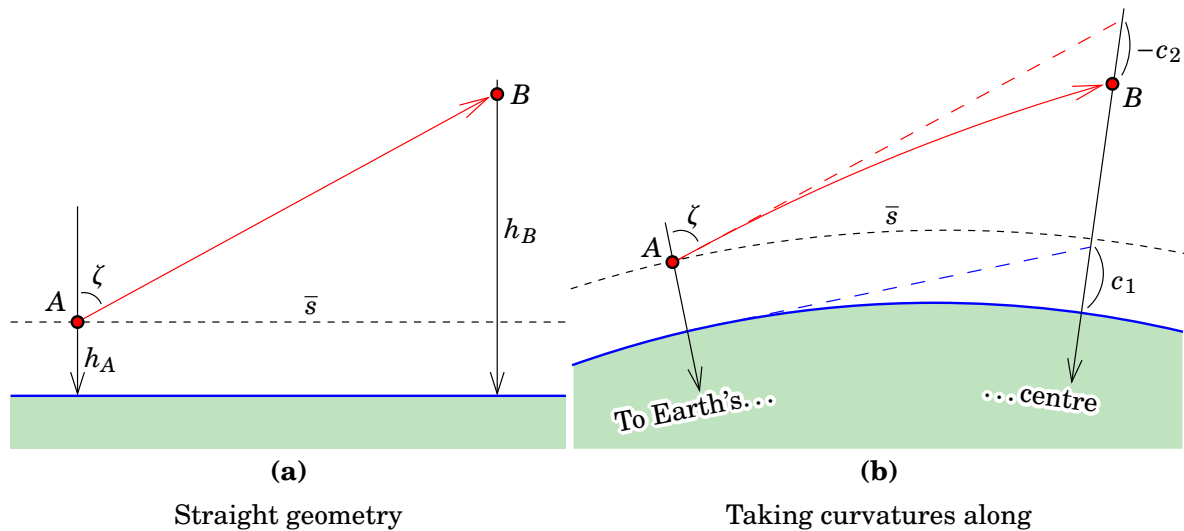


FIGURE 6.7. The effects of refraction and Earth curvature on zenith-angle measurement.

to the curvature of the Earth:

$$k = \frac{\text{ray curvature}}{\text{Earth curvature}} = \frac{R}{\rho}, \quad (6.7)$$

in which

R radius of curvature of the Earth. $1/R$ is the curvature of the Earth's surface. The curvature is the inverse of the *radius* of curvature!

ρ radius of curvature of the measurement ray. $1/\rho$ is the curvature of the measurement ray.

Typical k values in the atmosphere are $k_L = 0.13$ for visible light, and $k_M = 0.25$ for microwaves, see [Kahmen and Faig \(1988\)](#) page 167. The curvature of the measurement ray is thus 4–8 times weaker, and the radius of curvature 4–8 times larger, than the curvature, respectively radius of curvature, of the Earth's surface. However, during an atmospheric *inversion*, exceptionally large k values may occur, even 0.3–0.4, see [Grafarend et al. \(1987\)](#).

6.8.2 Trigonometric heighting

In figure 6.7 one sees on the left-hand side, how one may determine the height difference between two points A and B using zenith-angle measurement. The applicable trigonometric equation is

$$h_B = h_A + \bar{s} \cot \zeta,$$



in which \bar{s} is the *horizontal distance* — the slant range projected onto the horizontal plane — and ζ the measured zenith angle. vinoetäisyys

When, in reality, both the surface of the Earth and the measurement ray's path in the atmosphere are curved, in practice the right-hand side figure applies, in which, however, the curvatures have been exaggerated. Based on the figure, one should add to the above equation *two correction terms*:

$$c_1 \approx \frac{\bar{s}^2}{2R},$$

the correction due to the Earth's curvature, and

$$c_2 \approx -\frac{\bar{s}^2}{2\rho} = -k \frac{\bar{s}^2}{2R},$$

the correction due to the curvature of the measurement ray — assuming that, unlike in the figure, the height difference between the points is small compared to their distance.

Everything together:

$$h_B = h_A + \bar{s} \cot \zeta + (1 - k) \frac{\bar{s}^2}{2R}.$$

If we still take along the height of the theodolite, or instrument, i over point A and the height of the signal or target t over marker B underneath it, we obtain

$$h_B = h_A + \bar{s} \cot \zeta + (1 - k) \frac{\bar{s}^2}{2R} + i - t, \quad (6.8)$$

the *fundamental equation of trigonometric heighting*.



6.8.3 Simultaneous measurement in opposite directions

If we carry out zenith-angle measurements simultaneously at points A and B , we obtain

$$\begin{aligned} h_B &= h_A + \bar{s} \cot \zeta_A + (1 - k) \frac{\bar{s}^2}{2R} + i_A - i_B, \\ h_A &= h_B + \bar{s} \cot \zeta_B + (1 - k) \frac{\bar{s}^2}{2R} + i_B - i_A. \end{aligned}$$

Here, for the measurement done in A , the instrument height is i_A and the signal height i_B , and for the measurement done in B , the instrument height is i_B and the signal height i_A . So, we assume the heights of instrument and signal in the opposite-direction measurements to be the same at the same point — easy to do with *forced centring*, subsection pakkokeskistys
5.4.7.



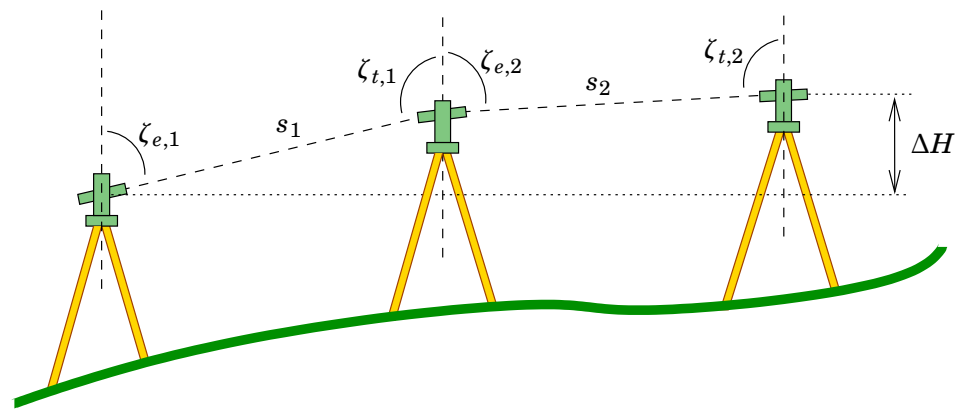


FIGURE 6.8. Trigonometric levelling traverse.

Rearranging the terms and subtraction yields

$$h_B = h_A + \frac{1}{2}\bar{s}(\cot\zeta_A - \cot\zeta_B) + i_A - i_B, \quad (6.9)$$

from which the *term describing atmospheric refraction has vanished*.

This method for determining height differences between points has been found to be good and has also been much-used over long distances. It assumes the distance \bar{s} to be known with sufficient precision.

takymetri This method is used in trigonometric levelling (figure 6.8). The method can replace traditional levelling, for example, in terrain with great height variations, where the staff distance of levelling would become short and the work laborious. In the method, two total stations and two signal-reflector assemblies are used, and the observation data is transmitted by radio modems from one instrument to the other for processing, error control and storage. Movement to the next point is by car where the terrain allows (Takalo, 1995).

6.8.4 Ölander's method

- ¹¹ As early as in the 1930s, V. R. Ölander¹¹ used in the Finnish primary triangulation a method for refraction modelling which was based on approximately simultaneous measurements *from each triangulation point to all neighbouring points*. This is thus a method essentially different from that of simultaneous measurements *in opposite directions*.

verkkotasoitus Ölander assigned every triangulation point in the network its own refraction coefficient as an unknown, which were all solved for by means of network adjustment (Ölander, 1932, see also Grafarend et al., 1987).

¹¹Victor Rafael Ölander (1897–1973) was a Finnish geodesist and astronomer who played a central role in the Finnish national primary triangulation.

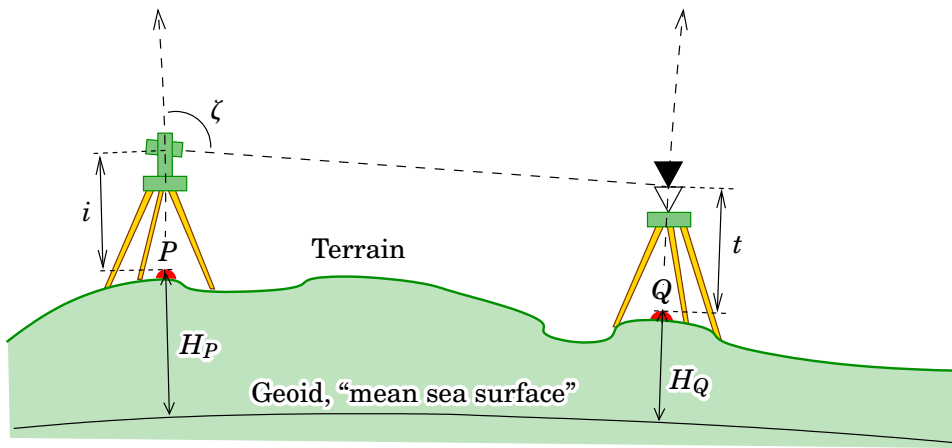


FIGURE 6.9. Heights of instrument and signal.



6.9 Heights of instrument and signal

Even though the instrument is levelled and centred on the point, it is nevertheless always *eccentrically set up in the height direction*.

The zenith angle to be measured is the angle between the plumb line and the sight axis of the telescope, figure 6.9. Thus, one must measure the height of the instrument i . Generally the aiming is at a signal on a tripod, not the point marker (monumented point) itself, so the signal or target height t must also be measured. So, one must always measure

luotiviiva
tähtäysakseli

- at the instrument, the height difference between the marker or monumented point, and the horizontal instrument axis
- at the signal, the height difference between the marker (monument) and the signal target point.

In the depicted situation (figure 6.9), the aiming point is the upper edge of the white triangle, and it is the height t of that point that must be measured.

In the height direction, both the instrument and the signal are *eccentrically set up*.

The figure also includes the heights of the monumented points from a computational reference surface (“sea level”), which the measurements aim to determine:

- height of the instrument from the reference level: $H_P + i$
- height of the signal or target from the reference level: $H_Q + t$.



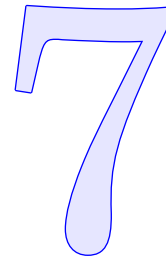


Self-test questions

1. What three basic topographic surveying measurement types use horizontal angle measurement?
2. What is station adjustment? What is an orientation unknown?
3. What is the difference between an open and a closed traverse? How many closing errors (redundancy, degrees of freedom) are there in each, which can be used to check for mistakes, or gross errors?
4. Why is traversing not entirely obsolete in spite of the existence of GNSS?
5. Why is atmospheric refraction a particular concern when measuring vertical angles, and not so much when measuring horizontal angles? Could you think of situations where horizontal-angle measurement could also suffer from refraction?
6. What is the refraction coefficient k ?
7. What is trigonometric levelling, and in what situations is it used? How is the effect of atmospheric refraction eliminated?
8. What was Ölander's refraction modelling method like?



Distance measurement



BERGSTRAND, ERIK Ö, observator, Danderyd, f i
 Upps 3/7/04 av prof Östen B o Anna Ericsson.
 Fil kand 35, fil dr 50, observator Rikets allm kartverk sed
 56. – Delt i solförmörkelseexp t Afrika 47. – *Utg* drsavh om
 ljushastigh, utv ny metod f avst:bestämn (geodimetern).
Gift 38 m folkskollär Lisa Torpson f 99, dtr t sem:rektor
 Nils T o Klara Svensson.

Erik Östen Bergstrand (1904 – 1987), inventor of the
 Geodimeter, *Who is Who*, [Harnesk and Davidsson \(1962\)](#)

7.1 Mechanical distance measurement

The SI unit of distance, or rather, of *length*, is the metre, see section 2.1. *Traceability* to the metre standard is of central importance in precise jäljitettävyys distance measurement.

Although today, even short distances are measured electronically or electro-optically, it is good to understand the oldest¹ and technically¹ simplest length measurement method, tape measurement. It continues to be used in local measurements, when distances are short and the precision afforded by tape measurement suffices.² And of course the²

¹Back in the day, the scale of the Finnish primary triangulation was transferred from the Nummela baseline to the baselines of the twelve principal sides of the network using invar wires. Invar is an iron-nickel alloy (64% iron, 36% nickel) which has a very small coefficient of thermal expansion. In the Lapland grade measurement by Maupertuis, the scale of the triangle network was also obtained mechanically from a baseline — during winter 1736 on the ice of the Torne river!

²*Careful* tape measurement is surprisingly precise!

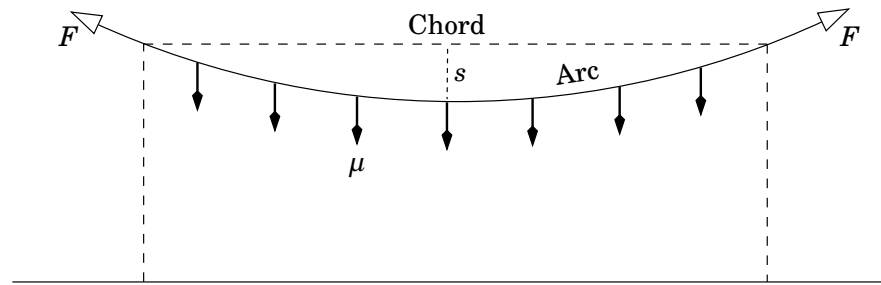


FIGURE 7.1. Sag correction of measuring tape.

equipment required is inexpensive!

In tape measurement, *four corrections* must be taken into account.

jäljitettävyy-
ketju

The tape correction is a tape-specific reduction. It is determined by *calibration* in a *comparator*, the true length of which is known precisely — thanks to comparison, through the traceability chain, with the standard metre. The tape correction $\Delta\ell_0$ is now the difference between the comparator's true length $\ell_0 + \Delta\ell_c$ and its length $\ell_0 + \Delta\ell_m$ measured by the tape to be calibrated:³

3

$$\Delta\ell_0 = \Delta\ell_c - \Delta\ell_m.$$

Here, ℓ_0 is the nominal length of the tape, for example $\ell_0 = 30\text{ m}$.

The temperature correction is caused by the thermal expansion of steel, and thus requires the temperature of the tape to be measured. The temperature correction is assumed zero for a standard temperature of $t_0 = 20^\circ\text{C}$. If the coefficient of thermal expansion of steel is α , the temperature correction is

$$\Delta\ell_t = \alpha\ell_0(t - t_0).$$

Here, α is expressed in micrometres per metre and degree. For example, the coefficient of thermal expansion of a certain steel alloy is $11.34\ \mu\text{m}/\text{m}^\circ\text{C}$, or $\alpha = 11.34 \cdot 10^{-6} (\text{C})^{-1}$, because $\mu\text{m}/\text{m} = 10^{-6}$. If a measuring tape is 30 metres long and the temperature is 28°C , it follows that

$$\Delta\ell_t = 2.7\text{ mm}.$$

painumis-
korjaus

Sag correction During measurement, the tape is tensioned with a

³Question: why do we use the algebraic sign like this? Why not

$$\Delta\ell_0 = \Delta\ell_m - \Delta\ell_c?$$



known force F . No matter how strong this force is, the tape will always settle into a *catenary*, in fact a *cosinus hyperbolicus* (cosh function) figure. We thus need to measure the *tensioning force* of the tape, figure 7.1. The theory of the phenomenon is surprisingly complicated. The end result, the sag correction, the difference in length between the chord and arc, is

$$\Delta\ell_s = -\frac{\mu^2}{24F^2}\ell^3,$$

proportional to the length of the tape to the third power. In this equation, μ is the weight of the tape per metre, and F the tensioning force of the tape. Alternatively one can measure the sag s in the middle and use the equation

$$\Delta\ell_s = -\frac{8s^2}{3\ell}.$$

The slope correction is actually not a correction but rather a *reduction*, which is necessary if, instead of the slope distance or *slant range*, one wishes to obtain the horizontal distance between the points, the projection of the slant range onto the horizontal plane.

In routine tape measurements, the tape correction, temperature correction and sag correction are often ignored. Their detailed description with formulas is found in the literature, for example Kahmen and Faig (1988) pages 122–130. The *slope reduction* can however be significant: if it is given that the height difference of the end points of a distance ℓ is Δh , then the horizontal distance is, according to Pythagoras' theorem,

$$\ell_{\perp} = \sqrt{\ell^2 - \Delta h^2}. \quad (7.1)$$

If we only know the slope angle α , $\sin \alpha = \Delta h / \ell$, between end points, we obtain the horizontal distance as follows, figure 7.2:

$$\ell_{\perp} = \sqrt{\ell^2 - \ell^2 \sin^2 \alpha} = \ell \cos \alpha.$$

Equation 7.1 may be written with sufficient precision for small slopes

$$\ell_{\perp} = \ell \sqrt{1 - \left(\frac{\Delta h}{\ell}\right)^2} \approx \ell \sqrt{1 - \left(\frac{\kappa}{100}\right)^2} \approx \ell - \frac{1}{2} \frac{\kappa^2}{10000} \ell \stackrel{\text{def}}{=} \ell + \Delta\ell_{\perp},$$

in which the slope reduction

$$\Delta\ell_{\perp} = -\frac{\ell\kappa^2}{20000} \quad (7.2)$$

is presented as a function of the *slope percentage* $\kappa \stackrel{\text{def}}{=} 100 \Delta h / \ell_{\perp} \approx 100 \Delta h / \ell$.



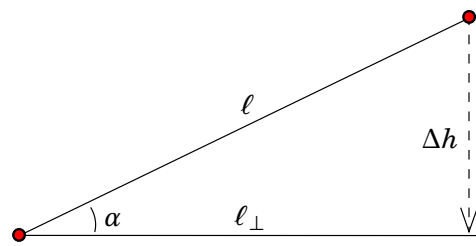


FIGURE 7.2. slope reduction of slant ranges.

7.2 Electromagnetic radiation

Light, radio waves and many other forms of radiation are examples of *electromagnetic radiation*.

In physics, it has been long considered whether visible light is a wave motion (Huygens) or a stream of particles (Newton). The finding of interference phenomena resolved the controversy in favour of the wave-motion theory. This was also decisively helped by the development, by James Clerk Maxwell,⁴ of the field theory of electromagnetism, which, using partial differential equations, presents *electromagnetic waves* as a natural wave phenomenon occurring in this field. Maxwell even succeeded in calculating theoretically the propagation speed c , which was close to the already observed speed of light. . .

vaihe Being a wave motion, electromagnetic radiation has a *phase* ϕ . When we describe a wave motion as the projection of a uniform circular motion onto one dimension⁵ (figure 7.3), then ϕ is the angle at the centre of

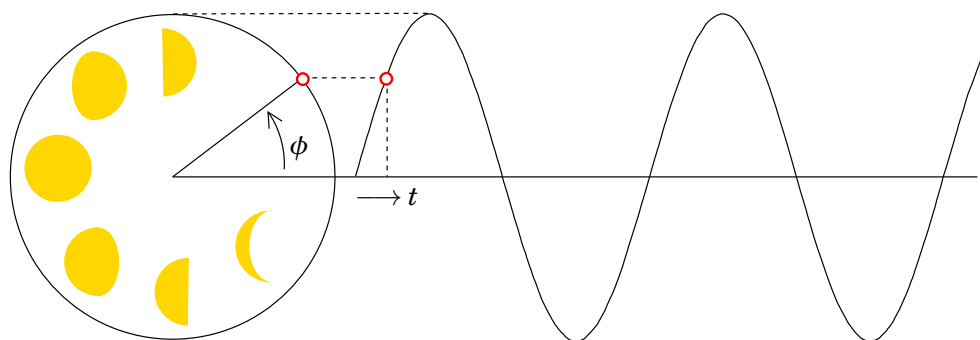


FIGURE 7.3. The phase ϕ of a wave motion as a function of time t . The name “phase” probably originates from the Moon’s phases.

⁴James Clerk Maxwell [FRS](#) [FRSE](#) (1831–1879) was a Scottish mathematical physicist who set the theory of the electromagnetic field on a mathematical footing, and also contributed essentially to theoretical thermodynamics.

⁵Equivalently: as the real part of the complex wave function $\exp(i\phi) = \cos\phi + i\sin\phi$.

that circle that measures this uniform motion. The relationship between phase ϕ and frequency f is

$$\phi(t) = \phi(t_0) + 2\pi f(t - t_0),$$

in which t is time and t_0 reference time. Clearly, the phase is periodic and repeats after one cycle, or 2π . Therefore we may always reduce the phase angle to the interval $[0, 2\pi)$ without changing its physical meaning.

Nowadays one can measure the wavelengths and frequencies of the various forms of electromagnetic radiation very precisely; between them there is the relationship

$$\lambda f = c,$$

in which f is the frequency and λ the wavelength. The quantity c is the *speed of light* (in vacuum), which according to Einstein — or actually already Maxwell — is a constant of nature. See figure 7.4.

However, *quantum theory* has made the particle model relevant again. Light can be described as a stream of particles, *photons*, the energy of each of which is

$$E = hf,$$

in which h is *Planck's*⁶ constant. The particle model is especially fertile⁶ for high energy levels, the left side of figure 7.4.

The electromagnetic field is a *vector field*. Therefore electromagnetic radiation is a *transversal wave motion*, and may be *polarised*.⁷ Figure 7.5

poikittainen aaltoliike

⁶Max Karl Ernst Ludwig Planck (1858–1947) was a German physicist and organiser of German physics. He is remembered for his discovery that the thermal radiation spectrum of a *black body* is caused in a natural way by the quantisation of electromagnetic radiation.

⁷The correct understanding of polarisation was the achievement of Thomas Young (1773–1829) and Augustin-Jean Fresnel (1788–1827). In studying polarisation, a certain mineral, clear calcite (CaCO_3), *Iceland spar*, had a central role. The crystal is *birefringent* (“doubly refracting”) and splits light into two parts according to their direction of polarisation. Apparently the Vikings used it in navigating by the Sun in overcast weather. The Dutch Christiaan Huygens (1629–1695), the father of the wave theory of light, spent ample time on the experimental study of Iceland spar.

Another background story links polarisation to the chemistry of life. Chiral (non-mirror-symmetric, “handed”) molecules, like sugar, rotate the plane of polarisation of light passing through them: *optical activity*. Already Louis Pasteur (1822–1895) was studying the background of the optical activity of organic tartaric acid, and today we know that the phenomenon is linked to the chirality of life itself, its “handedness”, like the winding direction of the helix of the [DNA](#) molecule. The *polarimeter* is a vital tool in the medical



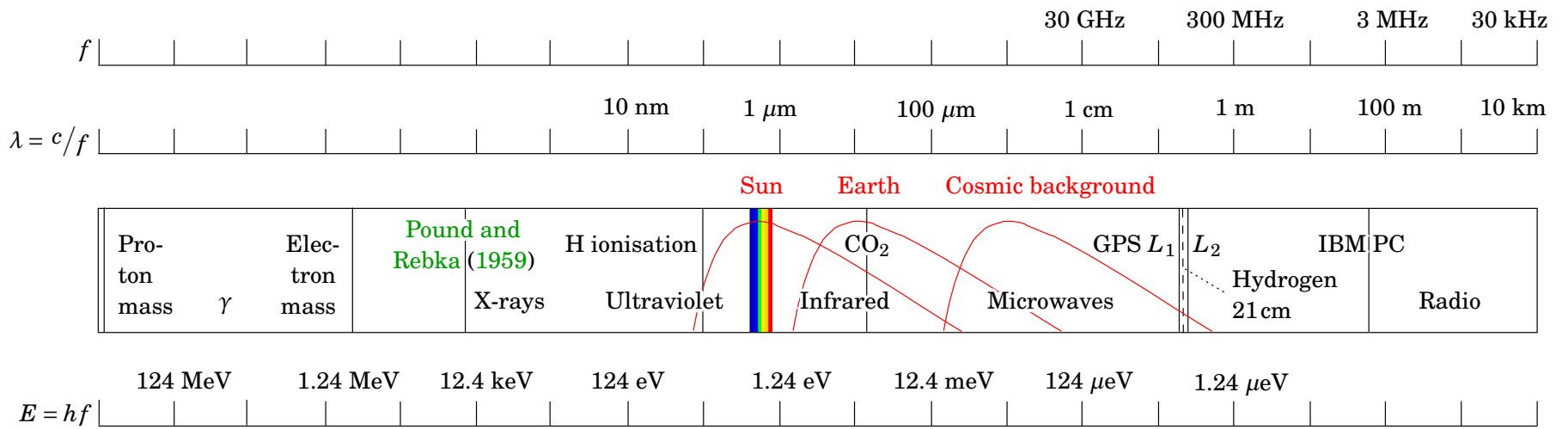


FIGURE 7.4. The electromagnetic radiation spectrum, a physics and astronomy landscape. Note the narrowness of what we can see with our own eyes.



shows both linearly and circularly polarised radiation. One can choose two independent directions of polarisation, for example up-down and left-right, of which all others can be composed by combination. For example, circularly polarised radiation is obtained by combining two mutually orthogonal, linearly polarised rays with a phase difference between them of $\pi/2$. This also works the other way around: by combining clockwise and anti-clockwise circularly polarised rays, one obtains again a linearly polarised ray.

An intermediate form between linear and circular polarisation is represented by elliptically polarised radiation, the field vector of which turns along an ellipse-shaped path.

In particle language one might say that the electromagnetic field is the quantum theoretical wave function of the photon, which thus is a vector-valued function. The photon is a *vector particle*, having an intrinsic angular momentum, or *spin*, of magnitude $h/2\pi$. This spin — which may be visualised as an angular-momentum vector — may be oriented along the flight direction, or opposite to it, which corresponds to either clockwise or anti-clockwise circular polarisation. Linearly polarised radiation may again be described as an equal mix of both spin directions ([Wikipedia, Photon polarization](#)).

pyörähdys-
momentti



7.3 Väisälä interferometry

One classical distance measurement technique which continues to be in use is Yrjö Väisälä's white-light interferometry technique, invented as early as in the 1920s, for the precise measurement of the lengths of long baselines. See figure 7.6.

perusviiva

The method works as follows. White light travels from a source to the observation instrument along two paths:

- directly, by being reflected from the far mirror
- by being reflected multiple times — in the figure, four times — back and forth between the near mirrors.

The light used is *white* and contains all the different wavelengths that make up white light. Therefore the *coherence length* of the light is very short, only $1.3\ \mu\text{m}$.

and food industries.

Iceland spar was also used in military equipment, for which reason it was long classified as a strategic material ([Cicala, 2013](#)).



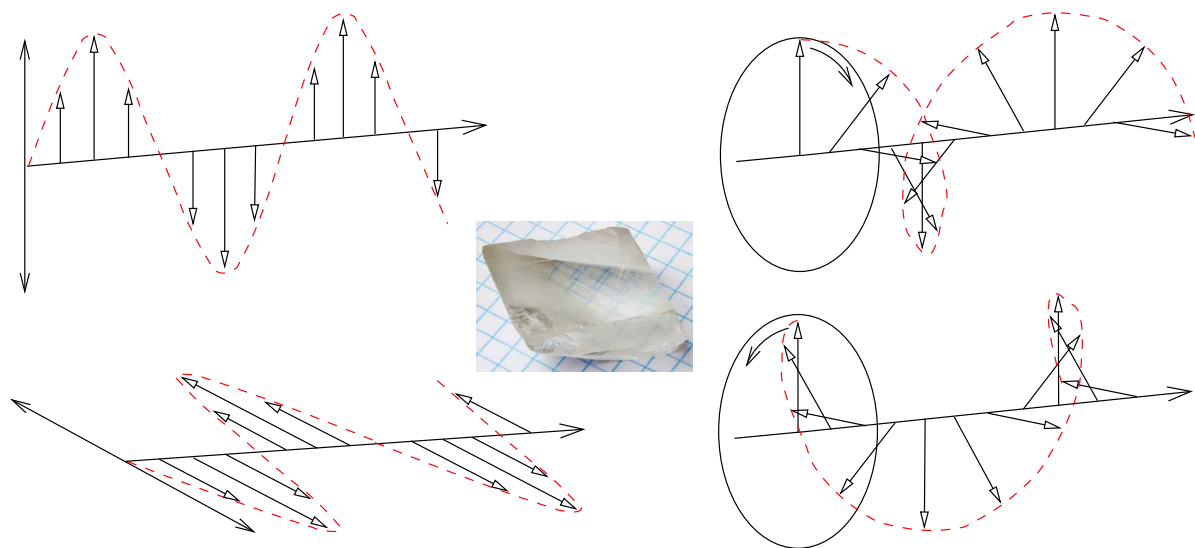


FIGURE 7.5. Polarisation of electromagnetic radiation. The arrows show the field vector \mathbf{E} . On the left, a linearly polarised wave motion; on the right, a circularly polarised wave motion. Photograph of Iceland spar [Wikimedia Commons, A calcite crystal](#).



This is why the interference fringes will only show up if both paths are, ⁸ at this accuracy, equally long:⁸ so, if the distance of the far mirror is a *multiple* of the distance between the near mirrors.

This enables the multiplication of a given distance. Say that the distance between mirrors 0 and 1 is precisely 1 m. Then we can, using interference, place the far mirror at distance 6 m — precisely. After that, we take mirror 1 away, and using the same method, using now the mirror pair 1 and 6 as near mirrors, place the far mirror at 24 m. And so on . . .

In practice, this does not quite work this way. There will always remain a small difference between the path lengths, which is eliminated and at the same time measured, using a turnable glass plate of constant thickness: a *compensator plate*.

The instrument described is actually an analogue optical *correlator*. See figure 7.7. We will go deeper into the correlation method in subsection 12.5.1, in connection with the GPS.

Realising the original distance of one metre between the 0 and 1 mirrors [kvartsimitta](#) is not quite simple, either. This uses a one metre long *quartz gauge*,⁹

⁸This is why laser light will not do! It would produce interference fringes even when the distances are unequally long.

⁹Calibrating those precisely is again a story of its own . . .

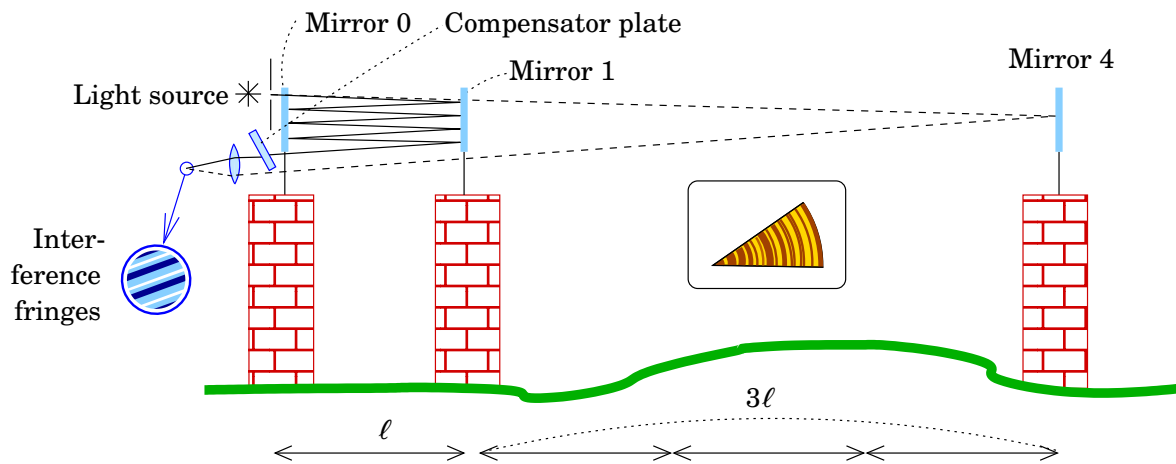


FIGURE 7.6. Väisälä's interference method.

which is allowed to touch the surface of one mirror. In the air gap between the other mirror and the end of the gauge — which has been polished to be slightly convex — Newton rings will now show, an interference pattern. By observing the rings in sodium light, the width of the air gap can be determined.

The Väisälä interferometry method is extremely time consuming. Measurement conditions are suitable only very rarely for measuring the longest distance, 864m. Moreover, setting up the mirrors, their orientation, and the transfer of their measured places to the underground permanent markers by *projection measurements*, is a complex operation demanding its own time. During measurements, the air temperature

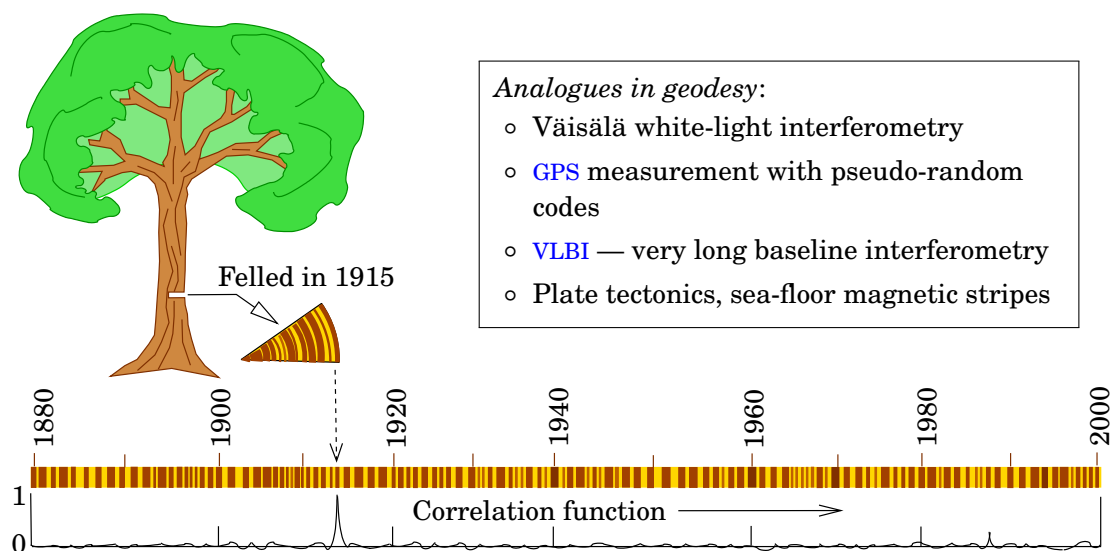


FIGURE 7.7. Dendrochronology: dating of wood using tree rings.

is continuously read from precision thermometers suspended all along the line, providing gainful employment and physical exercise to two measurement assistants.

- 10 The method has been used for measuring baselines as long as 864¹⁰ metres in Nummela, Finland, and in many places around the world. The precision achievable is at its best ± 0.02 mm. The interference measurement itself is more precise still: the bottleneck is the projection measurement, the transfer of the measurement values from the mirrors to the underground markers.



7.4 Electronic distance measurement



7.4.1 The speed of light

The speed of light in a vacuum is a constant of nature. Apparently Galileo (1564–1642) already tried to measure it in 1638 using two lamps and an assistant: a cover was taken off one lamp, and the assistant on another hill responded in the same way. Of course the result was useless: the speed of light would, according to the experiment, be very large, possibly infinite.

- etäisyysmittari The first terrestrial distance measurement devices or range-finders were developed to determine the speed of light. The prototype of the method is Fizeau's¹¹ instrument, consisting of a light source, a reflector and a rapidly spinning camwheel. If the wheel spins at the right speed, light leaving through an opening between the teeth will on return be blocked by a tooth. With a slightly greater or smaller rotation speed, however, some light will get through. In Fizeau's tests, the distance measured was 8.6 km.
- 11 hammaspyörä

Fizeau's camwheel is a primitive *modulator*. Nowadays electronic or electro-optical modulators are used, the task of which is to vary or *modulate* the intensity of outgoing light periodically at a certain frequency.

¹⁰864 = $2 \times 2 \times 3 \times 3 \times 4 \times 6$.

¹¹Armand Hippolyte Louis Fizeau [MIF FRS FRSE](#) (1819–1896) was a French physicist. He also measured the speed of light in flowing water and found in it an anomaly which only special relativity could explain. He is one of the 72 French scientists and engineers whose names were inscribed on the Eiffel Tower, outside and below the first balcony ([Eiffel Tower, 72 names](#)).

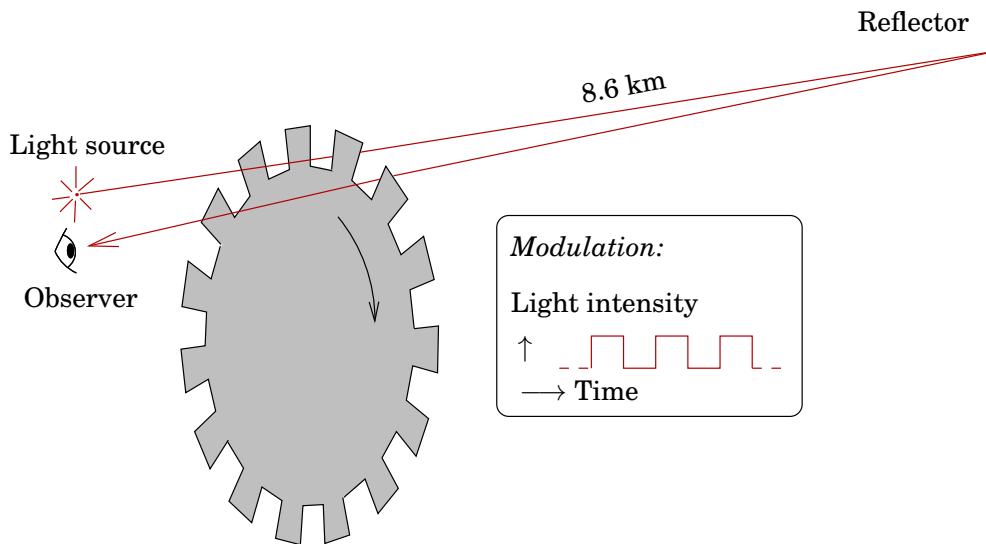


FIGURE 7.8. Fizeau's method for measuring the speed of light. The distance used for the 1849 measurement was from Montmartre to Mont Valérien, in the municipality of Suresnes, 8633 m.



7.4.2 Electronic distance measurement instruments

When the superior accuracy of the devices developed to determine the speed of light became clear, the picture changed. Today, the speed of light in a vacuum is no longer measured: instead it is a quantity derived from the definitions of the metre and the second (subsection 2.1.4) which has the conventionally agreed value of *exactly* $299\,792\,458\text{ m/s}$.

Electronic distance measurement devices or range-finders can be of three types of construction:

- separate: the device is placed into the forced-centring device. This solution has become rare due to the devices becoming ever smaller. pakkokeskistys-laite
- a separate part that is locked onto the theodolite's measuring telescope. This solution has also become impractical:
 - An instrumental tilt correction must be made.
 - For example, the classical Distomat range-finder was placed on top of the theodolite's telescope, preventing it from being plunged through to the other face. kojeasento
- integrated with the theodolite. One speaks of a *coaxial* solution: the light moves in both directions through the theodolite's measuring telescope and uses the same optics.

Electronic distance measurement is, based on the frequency range used, divided into two main types:



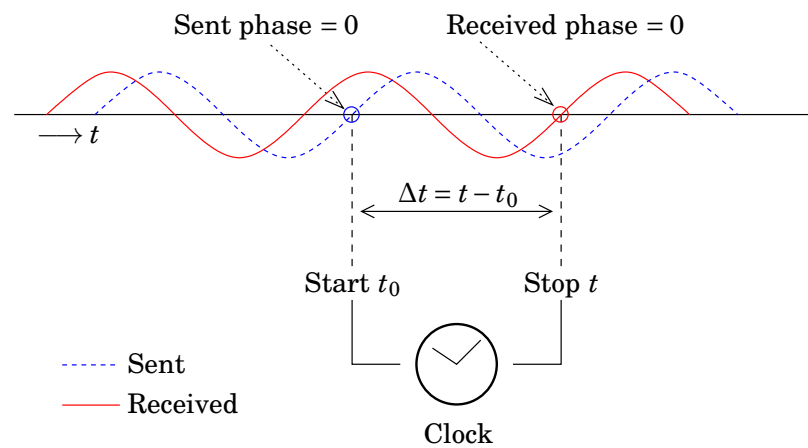


FIGURE 7.9. One method of electronic phase measurement: zero phase starts / stops an electronic counter or “clock”.

- electromagnetic, using microwaves. This is obsolete as a terrestrial method — but satellite positioning also uses microwaves!
- electro-optical:
 - visible light, white: Mekometer
 - laser light or light-emitting diode (LED), visible or near infrared. Monochromatic.

hohtodiidi

vaihe-ero

Independent of the device type, the measurement takes place by *modulating* either light (or infrared) or radio waves (microwaves) with a certain frequency, and measuring the *phase difference* between the outgoing radiation and the radiation returning from the target. Electronic phase measurement can be very precise, but does not tell us how many *whole* wavelengths fit in the distance: the *ambiguity problem*.

The travel time of the signal is

$$\overline{\Delta t} = \left(\frac{\Delta\phi}{2\pi} + N \right) \frac{1}{f},$$

kokonaisluku-
tuntematon

in which f is the frequency, $\Delta\phi$ the measured phase difference in radians, and the integer N the unknown *ambiguity*. Because the measured phase difference is assumed to lie in the interval $[0, 2\pi)$, the measurement is not yet enough to fix $\overline{\Delta t}$. We can certainly calculate a *possible* travel time

$$\Delta t = \frac{\Delta\phi}{2\pi} \frac{1}{f},$$

but it would be a coincidence if it were the true travel time.

Phase measurement can be done through the measurement of *time differences*: when the reference signal passes through zero in the positive



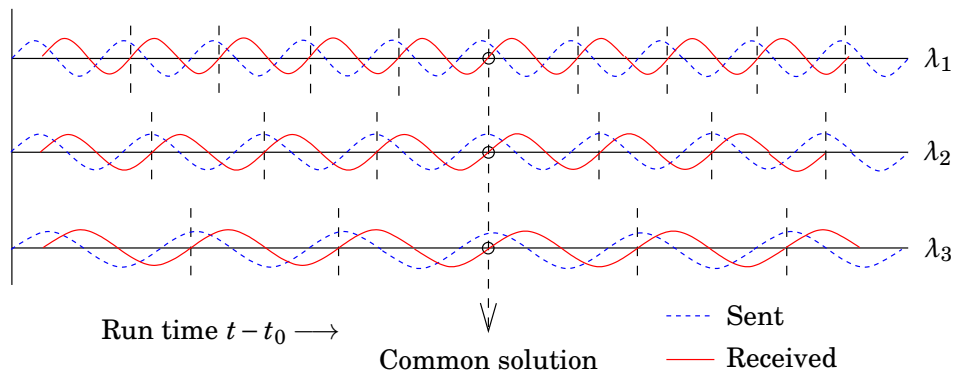


FIGURE 7.10. Ambiguities, or integer unknowns, are resolved by using several wavelengths. The figure shows signals on three different wavelengths and assumes that the phase angles of the transmitted signals are all zero at the moment of transmission t_0 . The only possible travel time is the one for which also the calculated phase angles of all three received signals are identical to their measured values, here also assumed to be zero. See also figure 13.4.

direction, it starts a counter, and when the returning measurement signal does the same, the counter stops and the value is read out. See figure 7.9.

From the phase measurement $\Delta\phi$ the *distance* is computed:

$$s = \frac{1}{2}c \overline{\Delta t} = \frac{1}{2}(c \Delta t + N\lambda) = \frac{1}{2} \left(\frac{\Delta\phi}{2\pi} + N \right) \lambda,$$

in which $\lambda = c/f$ is the wavelength, and $N \in \mathbb{Z}$ an unknown number of whole wavelengths. Determining the integer unknowns or *ambiguities* N is a similar problem as seen in connection with GPS carrier-phase measurements. In a range-finder, several different modulation frequencies f_i (or, equivalently, wavelengths $\lambda_i = c/f_i$), are used, which have been chosen so that only one distance s , and corresponding travel time $\overline{\Delta t}$, is compatible with integer-valued unknowns N_i for all wavelengths. See figure 7.10.

kantaaallon
vaihe

Nowadays there exist small and inexpensive hand-held range-finders, which work either based on an infrared beam or an acoustic (ultrasound) beam. They are handy in construction projects, and even real-estate brokers use them.

7.4.3 Reflectors

In order for electro-optical devices to work, they need a *reflector* placed at the target. Typically, a so-called *corner-cube prism* is used, figure 7.11. The principle of operation of a corner-cube prism is based on using

kuutioprisma



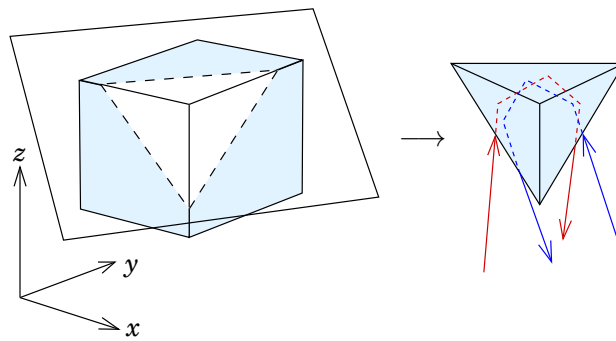


FIGURE 7.11. Corner-cube prism.

three reflective surfaces that are perpendicular to each other, of which the first inverts the x co-ordinate of the light ray to the $-x$ direction, the second, the y co-ordinate to $-y$, and the third, the z co-ordinate into the $-z$ direction. The end result is a complete inversion of the light ray:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \implies \begin{bmatrix} -x \\ -y \\ -z \end{bmatrix}.$$

The light ray hitting the prism will be reflected back in precisely the opposite direction, independently of which direction it came from — as long as it is within the prism's opening angle.

Over short distances, reflective stickers can also be used, or the light be reflected by the target itself, without aids. Then, the accuracy of the measurement is not necessarily the best possible!

Over longer distances one may use, instead of a single prism, an assembly of three prisms. Over very long distances (tens of kilometres) one can combine many prisms into a “pack”, figure 7.12. Nowadays such distances (vectors) are however measured using GNSS.

The prism assembly fits into a forced-centring device.

tähys **Warning** When using a signal equipped with a prism, one may not place the theodolite's crosshairs on the prism “crosshairs”! It likely is not aiming straight, and was never meant to be used for this. Use the triangular markings on the signal instead, figure 7.13.

hiusviiva-
ristikko



7.4.4 Instrumental errors of electronic range-finders

The systematic instrumental error of an electronic range-finder consists of two parts:



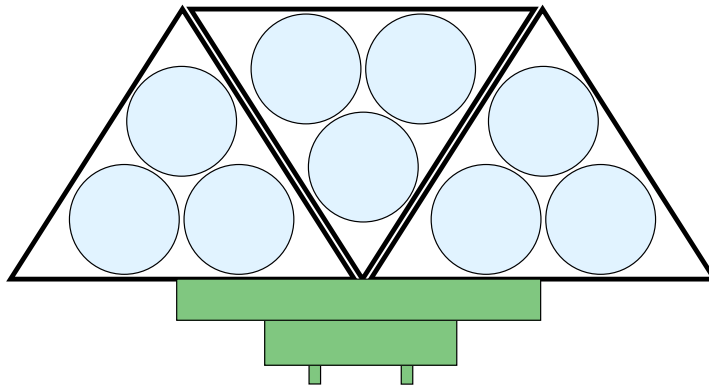


FIGURE 7.12. A prism pack for measurement over long distances.

- the zero-point or constant error
- the scale or frequency error.

The constant error is an instrumental constant to be determined by *calibration*. The frequency error is a scale error, which is also determined by calibration: frequency calibration.

The constant or zero-point error is caused by the circumstance that the electric centre of the instrument is at a different place than its nominal one. Inside the device, the signal path may contain unknown delays. The constant error may depend on temperature, may change slowly over time (“drift” or “creep”), and may change in connection with repairs. For this reason, *regular calibration* is recommended.

käynti,
ryömintä

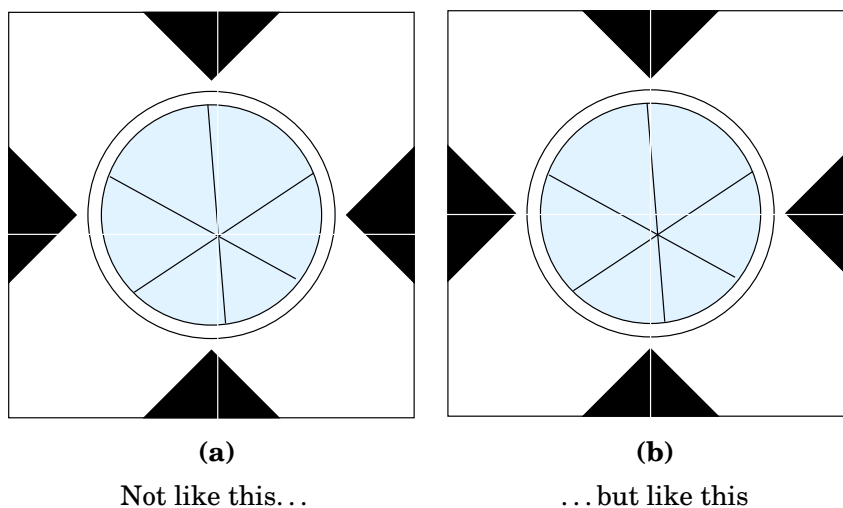



FIGURE 7.13. Incorrect and correct targeting at a signal equipped with a corner-cube prism.



 TABLEAU 7.1. Calculating the constant and frequency error of a range-finder by linear regression.

In order to calculate — *estimate* — the constant error a and the frequency error V from measurements on a calibration baseline, we use the following standard pair of equations for linear regression:

$$\hat{V} = \left(n \sum_{i=1}^n s_i \Delta s_i - \sum_{i=1}^n s_i \sum_{i=1}^n \Delta s_i \right) / \left(n \sum_{i=1}^n (s_i^2) - \left(\sum_{i=1}^n s_i \right)^2 \right),$$

$$\hat{a} = \frac{1}{n} \left(\sum_{i=1}^n \Delta s_i - \hat{V} \sum_{i=1}^n s_i \right).$$

In this, n is the number of calibration distances s_i used, at least 2. Δs_i is the difference: measured distance minus nominal distance, for each distance measured.

The reflector also has a zero-point error, and often, the sum of the zero-point errors of the instrument and reflector together are stated.

- The frequency error is often determined in the laboratory using a precise frequency standard:

$$V = \frac{f_{\text{measured}} - f_{\text{nominal}}}{f_{\text{nominal}}}. \quad (7.3)$$

In this, f_{measured} is the frequency value obtained in calibration measurements in the laboratory, and f_{nominal} is the frequency value stated by the manufacturer (and acting as the basis for the correction formulas in the instrument's firmware).

- The constant and frequency errors are determined by calibrating the instrument on a precise *baseline*:

$$s'_i = s_i - a - V s_i,$$

in which

s' “correct” distance given by the baseline

s distance as measured by the instrument to be calibrated

i number of the measurement point on the baseline, for example the number of the pillar.

We write

$$\Delta s_i \stackrel{\text{def}}{=} s_i - s'_i = a + V s_i$$



and solve for a and V by means of *linear regression*, tableau 7.1.

Of course obtaining a good calibration result requires the use of a sufficiently long calibration baseline. On a short baseline one can only determine a with sufficient accuracy. On the other hand, for many distance measuring devices, the electronic signal can be taken out and compared with a frequency standard directly, without using a baseline.

Thus, the instrumental correction for distance measurement is obtained:

$$s' = s - a - Vs,$$

in which

- s' corrected distance
- s measured distance
- $-a$ instrumental constant or zero-point correction
- $-V$ instrumental frequency correction.

The *random* total error of distance measurement, that is the *mean error* of the measurements, also generally depends on the length of the distance measured. An often-useful formula is

$$\sigma = \alpha + \beta s,$$

in which α is the random error for zero distance, β the distance-dependent random error, and σ the mean error of the observations computed from these. Here it is assumed that *systematic* errors — like constant error and frequency error — and the necessary observation reductions have already been taken care of as corrections.

7.5 Ray propagation in the atmosphere

Light — and other electromagnetic radiation like infrared or radio waves — travels slower in air, like in other media, than in a vacuum. The slowing-down effect of the medium is expressed by the *index of refraction* taitekerroin n . The definition of the index of refraction is

$$n = \frac{c}{c'},$$

in which c' is the speed of light in air, and c the speed of light in a vacuum, a constant of nature. Because air is a gas, a low-density medium, the values of n are always very close to unity. Therefore a definition is also



used that gives the deviation of the index of refraction from unity, in units of ppm, parts per million:

$$N = 10^6 (n - 1).$$

The index of refraction of air for the wavelengths of visible light is, according to the following approximate formula accepted by the International Association of Geodesy IAG at its general assembly of 1999 in Birmingham UK (Rüeger 1996, page 55, Anon., 1999):

$$N_L = N_0(\lambda) \frac{273.15 \text{ K}}{T} \frac{p}{1013.25 \text{ hPa}} - \frac{11.27 \text{ K/hPa}}{T} e, \quad (7.4)$$

¹² in which¹²

$N_0(\lambda)$ index of refraction of the light used (wavelength λ) in dry air under the following standard temperature and pressure conditions:

$$T = 273.15 \text{ K} = 0^\circ \text{ C}, \quad p = 1013.25 \text{ hPa}, \quad e = 0.0 \text{ hPa}$$

T air temperature, unit kelvin (K), i.e., absolute temperature

p air pressure, unit hectopascal (hPa), or millibar (mbar)

e partial pressure of atmospheric water vapour (“absolute humidity”), also in units of hectopascal.

N_0 depends only on the wavelength λ of the light used. An approximate formula for its calculation is

$$N_0 = 287.6155 + \frac{4.8866 \mu\text{m}^2}{\lambda^2} + \frac{0.0680 \mu\text{m}^4}{\lambda^4}. \quad (7.5)$$

ryhmä- ja vaihe-
taitekerroin
¹³
kantaalto

This is the *group index of refraction*, which differs from the phase index of refraction.¹³ In connection with electronic and electro-optic distance measurement equipment one must use the group index of refraction, because *information* travels in the *modulations* on the carrier wave, which propagate at group speed.

Example For a helium-neon laser (wavelength $\lambda = 632.8 \text{ nm}$) equation 7.5 yields $N_0 = 300.243$.

¹²Yes, the small anomalous water-vapour effect is treated as non-dispersive!

¹³In fact, the phase index of refraction is similarly (Anon., 1999):

$$N_0 = 287.6155 + \frac{1.62887 \mu\text{m}^2}{\lambda^2} + \frac{0.01360 \mu\text{m}^4}{\lambda^4}.$$



The index of refraction for *microwaves* again is

$$N_M = \frac{77.624 \text{ K/hPa}}{T} (p - e) + \frac{64.70 \text{ K/hPa}}{T} \left(1 + \frac{5748 \text{ K}}{T} \right) e. \quad (7.6)$$

Unlike the index of refraction for light and infrared, the microwave index of refraction *in the troposphere* is not dependent upon wavelength. However, the propagation of microwaves in the *ionosphere* is an entirely different matter relevant to satellite positioning; see subsection 12.7.3.

As an interesting detail, we may still note that, at a temperature of $T = 273.15 \text{ K}$, in the index-of-refraction equation for microwaves 7.6 the ratio of the effect of e to the effect of $p - e$ (dry-air partial pressure) is over twenty times larger than in the visible-light index-of-refraction equation 7.4!

Index	$p - e$	e	Ratio
Optical	0.29630	0.25504	0.86075
Microwave	0.28418	5.22133	18.373

Microwaves are thus very sensitive to water vapour,¹⁴ which is one of the drawbacks of this measurement technique. This problem is also manifest with GPS measurements.

The index of refraction of air affects the measured distance in the following way:

$$s - s' = (n - 1)s, \quad (7.7)$$

in which s is the measured distance, and s' the true distance, which would have been measured in a vacuum. The *refraction correction* is now applied as follows:

$$s' = s + \Delta s_n,$$

in which

$$\Delta s_n = -(n - 1)s = -10^{-6}Ns$$

is the correction. In the literature, the symbol K' is used for this. Because the correction is so small, it is permissible to use an approximate value for the distance s .

¹⁴This is caused by the non-symmetry and large dipole moment, or *chemical polarity*, of water molecules: in the H_2O molecule, the angle between the two O–H bonds is $104^\circ.5$. This is also one reason why water is a liquid at room temperature, and such a good solvent, and why a microwave oven is such a useful device for preparing food. Almost all other molecules in the atmosphere, like N_2 , O_2 , CO_2 , Ar, and CH_4 , are non-polar, and gases. [Wikipedia, Chemical polarity.](#)





7.6 “Curvature corrections”

kartoitus-
mittaus
runkomittaus

In distance measurement, the “curvature corrections” are second-order corrections: they are significant only when the distance of measurement exceeds many kilometres. Therefore they are inconsequential in detail or lower-order base-network surveys.

Because base-network surveys are nowadays done almost exclusively using GNSS, these corrections are mostly of historical interest, and we present them only briefly. More detailed expositions are found in the literature, for example Rüeger (1996, 2002).

refraktiokerroin

In the equations for curvature corrections, the *refraction coefficient* k , which we already encountered in connection with zenith-angle measurement (subsection 6.8.1), appears again. There are four different corrections:

The measurement-ray curvature correction The curving of the measurement ray causes a lengthening of the path. The measured distance is, for this geometrical reason, longer than the true distance. The equation for the correction is

$$\Delta s_{\rho} = -\frac{s^3}{24\rho^2} = -k^2 \frac{s^3}{24R^2}, \quad (7.8)$$

in which k is the refraction coefficient, R the radius of the Earth, and s the distance.

The Earth’s surface curvature correction After various reduction stages, the straight-line distance between the two projection points on a reference surface is usually obtained. It is, however, the longer distance over the curved Earth’s surface which is wanted. The equation for this correction is

$$\Delta s_R = \frac{s^3}{24R^2}.$$

The “second velocity correction” Usually, the effect of refraction on the propagation of the measurement ray is evaluated based on weather observations — measurements of air pressure, temperature, and humidity — made at both ends of the path. For very long paths, these measurements are no longer *representative* of the whole ray path.

The second velocity correction is a *systematic* effect caused by the ray curvature differing from the curvature of the Earth’s surface, together with the strong vertical gradient of the air pressure —



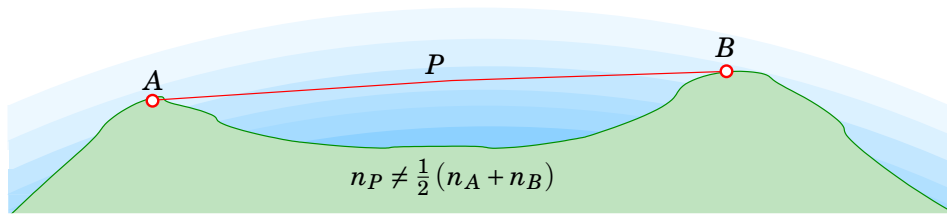


FIGURE 7.14. The second velocity correction: on a curved Earth, the indices of refraction at the end points of the measurement path are not representative.



and thus the index of refraction. When the curvature of the ray path is smaller than that of the Earth’s surface, at longer distances the path will “dive” deeper into the Earth’s atmosphere than what the end points of the path are telling us. See [Rüeger \(1996\)](#) page 81. The computational formula — the derivation of which is laborious — is

$$\Delta s_{n'} = -k(1 - k) \frac{s^3}{12R^2}.$$

The literature calls this correction K'' .

All three corrections can be combined into one equation:

$$\Delta s_{\rho R n'} = \Delta s_{\rho} + \Delta s_R + \Delta s_{n'} = (1 - k)^2 \frac{s^3}{24R^2}.$$

Let us calculate some values, assuming $k = 0.2$:

s (km)	1	3	10	30	100
$\Delta s_{\rho R n'}$	$0.66 \mu\text{m}$	$18 \mu\text{m}$	0.66mm	18mm	0.66m

The effect is thus normally very small.

The “terrain correction” The terrain correction of distance measurement ([Juhani Kakkuri](#), personal communication; figure

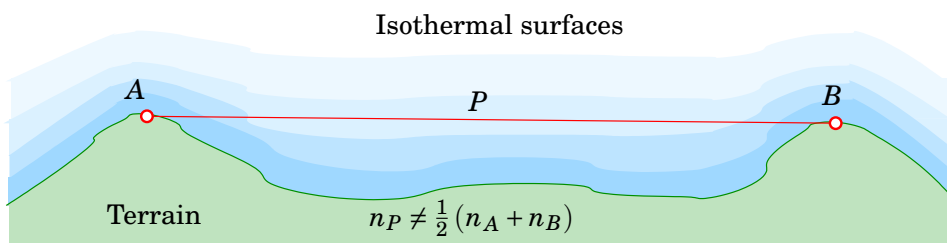


FIGURE 7.15. The terrain correction of distance measurement. Due to the shape of the terrain, the indices of refraction at the end points of the measurement path are not representative.



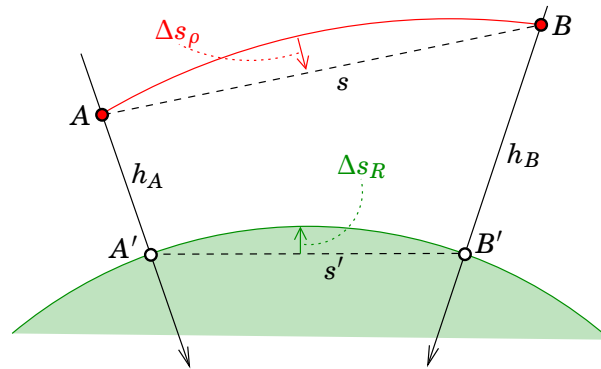


FIGURE 7.16. Reduction of distance measurement to a reference level, equation 7.9.

7.15; Kakkuri et al., 1981) is caused by the surfaces of equal temperature within the atmosphere, the *isothermal surfaces*, generally following the forms of the terrain. Therefore, in the same way as in the case of the second velocity correction, the weather observations at points *A* and *B* are non-representative of the mean value along the measurement path. There is no simple formula for this phenomenon.

7.7 Geometric reductions

7.7.1 Reduction to a reference surface

Thanks to the above described “curvature corrections”, the original measurement was reduced for the curvature of both the measurement ray and the Earth’s surface. Thanks to corrections Δs_n and $\Delta s_{n'}$ the delays by the atmosphere were also taken into account. The measurement path is however still sloping and above the Earth’s surface.

The correction, or *reduction*, to a chosen *reference level* corrects both the *slope* of the measured distance in space and its *height* above the chosen reference surface. The following may be chosen for reference surfaces:

- sea level
- the surface of a reference ellipsoid
- the height of a locally defined reference surface (“zero level”).

The reduction is carried out as follows:

$$(s')^2 = \frac{s^2 - \Delta h^2}{(1 + h_A/R)(1 + h_B/R)}, \quad (7.9)$$





TABLE 7.2. Examples of distance reductions.

		$s =$	100 m	1 km	10 km	100 km
$h_A = 0$	$h_B = 10 \text{ m}$		-0.50 m	-51 mm	-13 mm	-79 mm
$h_A = 0$	$h_B = 100 \text{ m}$		-	-5.0 m	-0.58 m	-0.83 m
$h_A = 0$	$h_B = 1000 \text{ m}$		-	-	-51 m	-13 m
$h_A = 100 \text{ m}$	$h_B = 100 \text{ m}$		-1.6 mm	-16 mm	-0.16 m	-1.6 m
$h_A = 1000 \text{ m}$	$h_B = 1000 \text{ m}$		-16 mm	-0.16 m	-1.6 m	-16 m

in which s is the measured slant range in space, s' the reduced distance (meaning the distance between the points A', B' projected onto the reference surface, see figure 7.16), $h_A = AA'$ and $h_B = BB'$ the heights of the points from the reference surface, $\Delta h = h_A - h_B$, and R the (approximate) radius of curvature of the Earth. The heights of the theodolite and signal are included in the heights h_A and h_B .

An approximate equation for when $\Delta h \ll s$ (and of course $h_A, h_B \ll R$) is

$$s' \approx s - \frac{\Delta h^2}{2s} - s \frac{h_A + h_B}{2R}.$$

An advantage of this equation is, that it gives separate terms for the influence of path slope and path mean height.

Table 7.2 gives some examples of corrections $s' - s$. It is seen that this correction can be already substantial also for short distances.

Modern theodolites and total stations are able to compute, in addition to device-specific corrections, at least the slope reduction of the measured range. takymetri



7.7.2 Map-projection reduction

A map-projection scale reduction is needed if we wish to have, instead of the length of an arc reduced to the reference ellipsoid s_{ell} , the length in the *map plane*, the projected length s_{proj} . For example, in the case of the Gauss–Krüger projection, as used in Finland, the approximate reduction is done as follows.¹⁵

$$s_{\text{GK}} = s_{\text{ell}} \left(1.0 + \frac{y_A^2 + y_A y_B + y_B^2}{6R^2} \right), \quad (7.10)$$

in which the distance s is between the points A and B , map co-ordinates (x_A, y_A) and (x_B, y_B) . Here, y is the raw distance from the central meridian, without the false easting 500 000 m. 15



This reduction *depends on the map projection chosen*, and is thus different for different map projections. For example, the equation for the Universal Transverse Mercator (UTM) projection is otherwise the same as 7.10, but the constant — the scale on the central meridian — is 0.9996 instead of 1.0.



Self-test questions

1. In what kinds of situations would tape measurement be a sensible option?
2. Derive the approximate slope correction $\Delta \ell_{\perp}$, equation 7.2, expressed in the slope angle α in degrees. Assume a small angle.
3. What is the relationship between wavelength and frequency for electromagnetic radiation?
4. What is the relationship between frequency and the energy of a photon for electromagnetic radiation?
5. What is the polarisation of electromagnetic waves?
6. Describe the correlation process between two identical but random signals. Why must the signals be random?
7. Describe ambiguity resolution in electronic distance measurement.
8. What are the factors affecting ray propagation in the atmosphere,

¹⁵The derivation starts from the scale of the Mercator projection on the sphere being, close to the equator,

$$m = \frac{1}{\cos \varphi} \approx \frac{1}{1 - \frac{1}{2}\varphi^2} \approx 1 + \frac{1}{2}\varphi^2 \approx 1 + \frac{1}{2}\frac{x^2}{R^2},$$

in which x is the metric distance from the equator. For a transversal Mercator projection, this becomes

$$m = 1 + \frac{1}{2}\frac{y^2}{R^2},$$

with y the distance from the central meridian. As clearly the scale distortion is quadratic in y , it will be, on a straight line on the map, a quadratic function of place along the line. Then, calculating the mean scale along the line by Simpson integration, equation 10.4, will be exact. As follows (M is the midpoint between end points A and B):

$$\begin{aligned} \bar{m}_{AB} &= 1 + \frac{1}{2}\frac{1}{R^2} \cdot \frac{1}{6} (y_A^2 + 4y_M^2 + y_B^2) = 1 + \frac{1}{12R^2} (y_A^2 + (y_A + y_B)^2 + y_B^2) = \\ &= 1 + \frac{1}{12R^2} (2y_A^2 + 2y_A y_B + 2y_B^2) = 1 + \frac{1}{6R^2} (y_A^2 + y_A y_B + y_B^2), \end{aligned}$$

from which the result 7.10 follows.



for visible light and for microwaves?

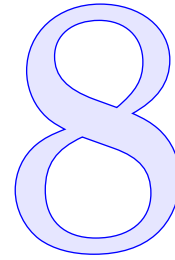
9. How are range-finders calibrated?
10. How are distance measurements reduced to a reference surface?
11. Prove equation 7.8, by defining $\psi \stackrel{\text{def}}{=} s/\rho$, the angle at the centre of the circular-arc beam path, so that

$$\Delta s_\rho = \rho \left(2 \sin \frac{1}{2} \psi - \psi \right).$$

Use the Taylor expansion for the sine function.



Base-network and detail-survey measurement



[...] Komitea [Karttakomitea] sitten ehdottaa toimenpiteenä tarkan karttalaitoksen aikaansaamiseksi, että perustettaisiin Geodeettinen komissioni, jonka tulisi toimittaa: 1) karttalaitoksen pohjaksi tarvittavat perustavat työt, 2) geodeettiset täytetyöt, 3) topografinen peruskartta mittakaavassa 1 : 20,000, Ensimmäinen ryhmä käsittäisi ensiluokan kolmiomittaukset ja tarkkavaakituksen ja toinen ryhmä pääasiallisesti alemman luokan kolmiomittaukset. [...]

Finnish Geodetic Institute director Ilmari [Bonsdorff \(1920\)](#)



8.1 Objective and planning of base-network measurement

The *task* of base-network measurement is to create, through a *network hierarchy*, the geometric foundation for mapping the country. For this purpose a permanent, sufficiently dense and precise benchmark set is created to which the local measurements of the various user groups will be tied. The co-ordinates of the benchmarks are known in the national co-ordinate reference frame, and by using them, the locally measured points and drafted maps will also be in the same frame.

runkomittaus

kiintopisteistö

Benchmarks are used both in detail surveys for mapping, and in setting out plans onto the terrain — “the inverse problem of mapping”.

kartoitus-
mittaus

The *planning* of base-network measurements starts from taking an inventory of the existing situation and an analysis of needs. The goal is to build a sufficiently precise and dense set of benchmarks, at a minimum cost. It pays off, however, to plan it for the future, especially in choosing the *benchmark substrate* and in *monumentation*. Possible future building activity is also taken into account in the choice of location, as this may

maastoon
merkintä

kiintopiste-
alusta

destroy a point or render it useless by destroying the visibility conditions for measurement: in the case of theodolite measurement, intervisibility between points, in the case of satellite measurement, sufficient visibility of the sky from the point.

rekognosointi *Reconnaissance* is part of the detailed planning of a network: an on-the-spot check that the measurements can be carried out as planned. Before reconnaissance, a “map reconnaissance” is done, in which the situation is judged from the office. If new points are created, a clear and useable **pistekortti** *point description* must be drafted for every point, with the aid of which also others can find it.

In addition to *precision*, attention must also be paid to *reliability*. **karkea virhe** Reliability means that possible gross measurement errors are noticed with the greatest possible ease, and that the effect on the end result — the co-ordinate solution — of the greatest possible gross error that remains undetected is as small as possible. To this end, there must be sufficient *redundancy* in the network: the measurement plan should always contain enough extra measurements above and beyond the required minimum.

kolmiomittaus-
verkko
monikulmiojono Nowadays, instead of the traditional solution — a triangulation network densified by traverses — **GNSS** networks are commonly measured. Those, too, need to be designed right, meaning hierarchically, and the measurements must be planned so that the accuracy and point-density objectives are achieved in an economical way.

The following alternative methods are on offer for base-network measurement:

- satellite positioning (**GNSS**)
- takymetri** ◦ traditional terrestrial measurement using total stations
- ilmakolmiointi** ◦ photogrammetric aerotriangulation.

The choice is dictated by the purpose of use — accuracy requirements and size of and ease of movement within the area — as well as by the visibility conditions at the points.

Measurement technology for base networks has, with satellite positioning, undergone a revolutionary change. Traditionally, base-network measurements were done for the horizontal using triangulation and traversing, and using precise levelling and lower-order levelling for measuring heights. Nowadays, satellite positioning is always used if at all possible. There are however situations where traditional techniques hold their own, like tunnel and mine surveying where the sky cannot be seen.



A good introduction to the subject is [Salmenperä \(1998\)](#), on which this presentation is partly based.



8.2 Guidance and standards

Many different standards and guidance documents exist to ensure the quality and effectiveness of geodetic measurements. We present here only those with a clear official status. ohjeistus

Important guidance was provided by the Zoning Survey Guide (“*Kaavoitusmittausohjeet*”) of the Finnish National Land Survey ([Anon., 2003](#)). The guidance concerns base-network measurement, detail survey, aerial mapping, and the drafting of a zoning base map, as well as documenting the work. Kaavoitusmittausohjeet
kaavan pohjakartta

Base-network measurement, co-ordinate reference systems, map projections, and zoning surveying has of late been the subject of guidance by [JUHTA](#), the Advisory Committee on Information Management in Public Administration, who publish the series [JHS](#), Recommendations for Information Management in Public Administration. In 2013 it was decided that from then on, the guidance concerning zoning surveying would be published in the [JHS](#) series.

Of the guidance documents relating to these subjects, we may mention the following on-line publications, unfortunately only in Finnish:

- [JHS 196](#): [EUREF-FIN co-ordinates in Finland](#) ([JUHTA, 2016a](#)).
- [JHS 197](#): [EUREF-FIN co-ordinate systems, related transformations and map-sheet division for ETRS89](#) ([JUHTA, 2016b](#)).
- [JHS 163](#): [Finland’s height system N2000](#) ([JUHTA, 2010](#)).
- [JHS 178](#): [Interface for geographic information services for local authorities](#). This document defines an interface called [kuntaGML](#)¹ (a variant of Geographic Mark-up Language) ([JUHTA, 2012a](#)).
- [JHS 184](#): [Point measurement in the EUREF-FIN co-ordinate reference system](#) ([JUHTA, 2012b](#)).
- [JHS 185](#): [Composing base maps for the city](#). This replaces in part the earlier *Zoning Survey Guide* and the *Zoning Base Map Guide* (“*Kaavan pohjakartta*”) by the Finnish National Land Survey ([JUHTA, 2014](#)).

Terminology work also represents important standardisation work in the sanastotyö

¹The name translates as “municipalityGML”.





TABLE 8.1. Methods for base-network measurement.

Areal extent	Order		Traditional methods	Modern methods
	Old	New		
Global	-	-	-	GNSS, VLBI, satellite laser, DORIS
1000 km	-	E1	(Stellar triangulation)	GNSS, continuously operating (FinnRef TM)
100 km	I	E1, E1b, E2	First-order triangulation	GNSS, EUREF-FIN densification
10 km	II, III	E3	Lower-order triangulation	GNSS, static
1 km	IV, V	E4	Traverses, aerotriangulation	GNSS, aerotriangulation, RTK with caution

Finnish-language area. Let us mention the Vocabulary of Geoinformatics (Sanastokeskus TSK, 2018) drafted in collaboration by the National Land Survey and the Finnish Terminology Centre TSK.

Guidance and standardisation is always a work in progress.



8.3 Network hierarchy and classification

During the past quarter of a century, almost all base-network measurements have already been done using the satellite positioning technique — with the exception of precise levelling, for technical reasons (the geoid problem). Table 8.1 catalogues the technologies used back then and now in connection with base-network measurement.

The table also illustrates well the concept of *network hierarchy*: more local networks are always tied to more extensive ones, which serve as the former's “formal truth”. The order of operations is always *from the large to the small*: first the most extensive networks are measured, which are then *densified* with measurements in a smaller area. In this way, a

kiintopisteistö benchmark set is obtained that

- Covers the whole country.
- Is dense enough: construction requires starting points sufficiently nearby, at most a few hundred metres from the project area and from each other.
- Is of homogeneous quality.



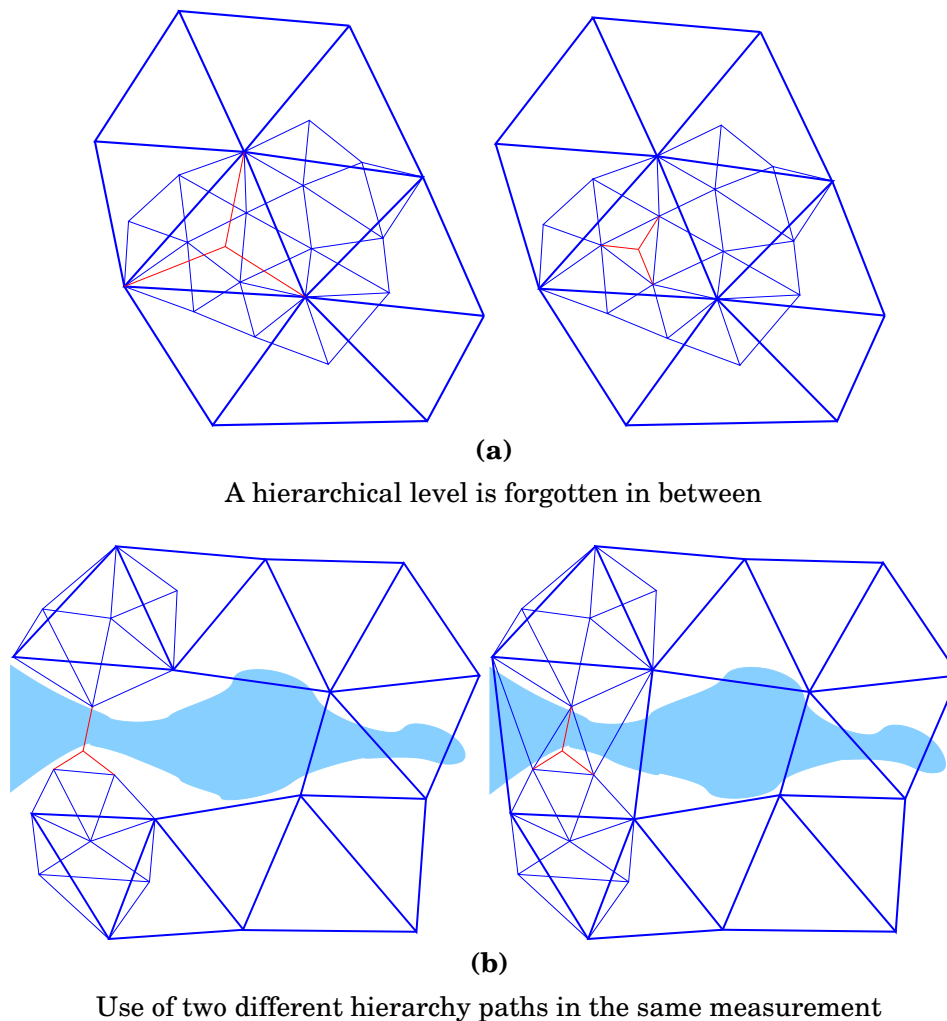


FIGURE 8.1. The significance of network hierarchy, and mistakes often made.
Not like this... but like this.

The hierarchical method is meant to prevent the unpleasant situation from occurring in which neighbouring points have co-ordinates determined for them along different paths, so that the *relative* location precision between them is weak.

The new measurement technologies mentioned in the table will be discussed in later chapters.

The planning of fundamental geodetic works in Finland started already before independence, and the work started immediately after (Bonsdorff, 1920). The network of the first-order triangulation mentioned in the table comprises 364 points and covers the whole territory of Finland. The network was measured by the Finnish Geodetic Institute during 1919–1987. Lower-order triangulations and traverses were measured by the then National Board of Survey, today the National Land Survey. Local

kolmiomittaus

jonomittaus



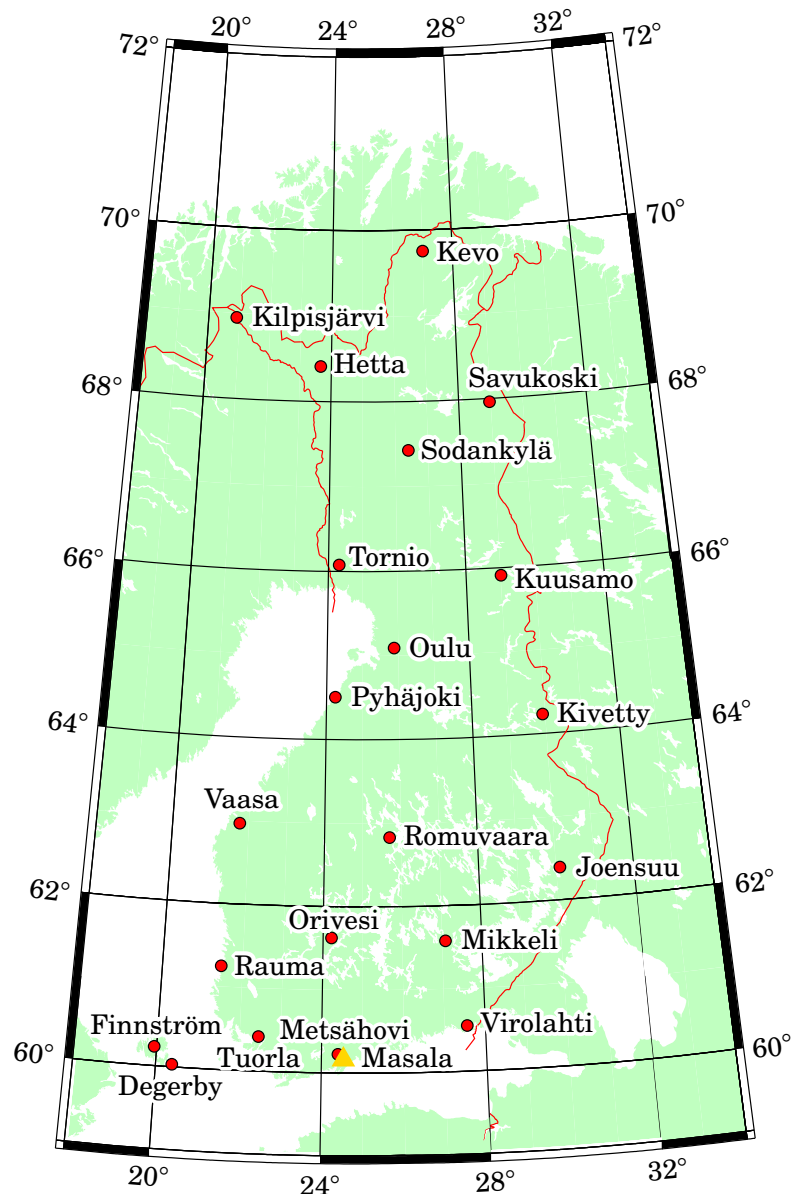


FIGURE 8.2. The Finnish continuously operating GNSS network FinnRef™, status 2018. The stations collect GNSS measurements continuously at a rate of one measurement event per second. The data centre is at the Finnish National Land Survey's Geospatial Research Institute FGI in Masala, Kirkkonummi, 30 km west of Helsinki.

measurements were carried out by many players, such as municipalities. In a similar way (JUHTA, 2012b) the points of orders E1 and E1b were measured by the Finnish Geodetic Institute, whereas the orders E2 and E3 were measured by the National Land Survey and the Finnish Maritime Administration. E4 and the use-point orders E5 and E6 are measured by municipalities.



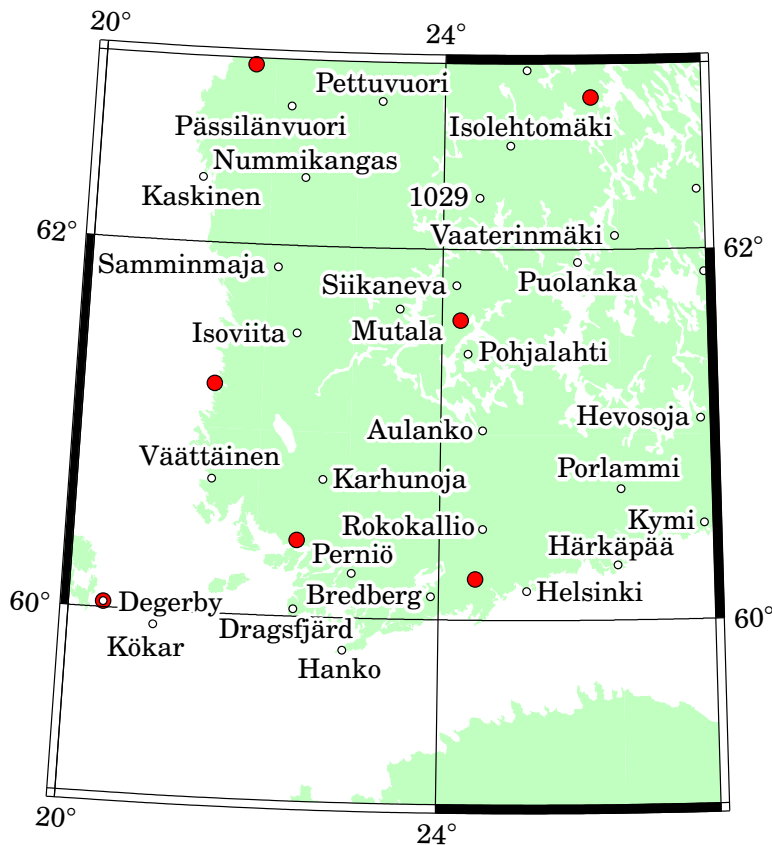


FIGURE 8.3. The Finnish **EUREF-FIN** first-stage densification network, detail. These points, together with the FinnRef™ points, form the E1 order.

Today, the highest level in the Finnish national **GNSS** network hierarchy is formed by the continuously operating **GNSS** network FinnRef™. Earlier on it consisted of 13 stations, in 2012–2013 a renovation was carried out, bringing it up to 20 stations. The observations are collected by the Finnish National Land Survey's Geospatial Research Institute **FGI**, the former Finnish Geodetic Institute. In 1996–1999 the **FGI** carried out a two-stage **EUREF-FIN** densification using the static **GPS** measurement technique, comprising in total some 450 points. The network of the first stage consists of 100 points, see figure 8.3 and **JUHTA (2016a)**. It was measured in 1996–1997. Together with the continuously operating **GNSS** network FinnRef, it forms the modern order I, or E1. Together they determine the **EUREF-FIN** co-ordinate reference frame. The other **EUREF-FIN** densification phase, which was measured in 1998–1999, comprises 350 points, and was designed to offer easier-to-reach points for practical measurements. Its order is E1b.

The National Land Survey has already for many years carried out base-



network measurements using the static GNSS measurement technique: there are some 2000 points of order E2, the measurements of which are tied directly to points of orders E1 and E1b. In recent years, kinematic GNSS measurement (RTK, *real-time kinematic*) has become common in lower-order base-network measurements, although its suitability for this has been credibly questioned. The technique should not be used without adequate reliability safeguards (JUHTA, 2012b).

A good description of the co-ordinate solutions used in Finland and the relationships between them is given in Häkli et al. (2009).



8.4 The terrain, the ellipsoid and the map plane

Geodetic measurement networks, like a triangulation network or a traverse, are in fact three-dimensional networks, figure 8.4. A logical idea is then to also carry out the computation of the network, the *adjustment*, three-dimensionally: the point co-ordinates are written three-dimensionally in the form of rectangular co-ordinates, and every observable is described as a function of those co-ordinates of the points between which the measurement takes place. This is how one obtains the *observation equations* upon which the adjustment of the network is based.

Three-dimensional network adjustment is a tempting thought, mostly because of the simplicity of the underlying idea. The formation of observation equations is nevertheless complicated, as the measurements are done, at every point, in *instrument co-ordinates*; co-ordinates in which the z axis points upwards along the local *plumb line*. The direction of the plumb line, which can be measured by astronomical means, is different at each point, as seen in figure 8.4.

This means that at least in the observation equations for the horizontal angles (azimuths) and zenith angles, the direction of the plumb line must be along. This makes these equations seriously complicated. For slant ranges, on the other hand, the observation equation is simple.

Terrestrial geodetic measurements are always made close to the physical surface of the Earth, generally between points located on the surface. Thus we may call the network geometry “quasi-two-dimensional”. It would also seem to make sense to try to carry out the computations in two dimensions, on a suitably chosen, mathematically simple computation or *reference surface* close to the Earth’s surface. The Earth’s physical



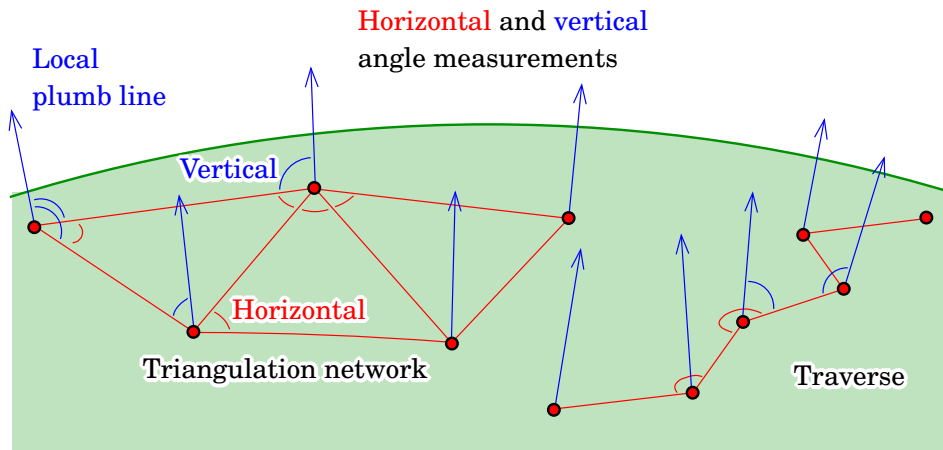


FIGURE 8.4. Geodetic networks are three-dimensional. A triangulation network and a traverse in space.

surface with its mountains and depths is however too craggy to serve as a computation surface.

More suitable computation surfaces are the *reference ellipsoid*, or — in a small area — the *map projection plane*, figure 8.5. In preparation for computation, the observations are *reduced* to this computation or reference surface.

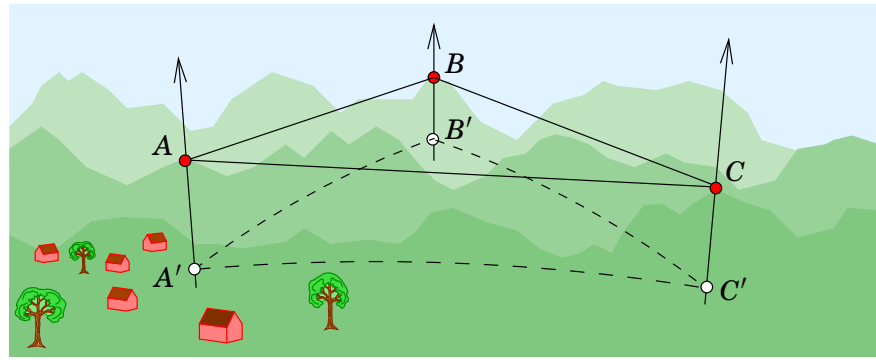
8.4.1 Adjustment on the reference ellipsoid

The reference ellipsoid coincides rather well with the Earth's surface, and as a simple mathematical surface it may serve as a computation surface. Table 8.2 shows the magnitude of the differences between the physical surface of the Earth and the reference ellipsoid.² And let us remark still, that humankind lives close to the surface of the solid Earth on land, but close to the sea *surface* on the seas: the impression given by the table exaggerates the thickness of the human living space.

For comparison, the difference between the equatorial and polar radii of the GRS80 reference ellipsoid is already 21.4 km or 0.336%. A “reference sphere” would be a clearly poorer approximation.

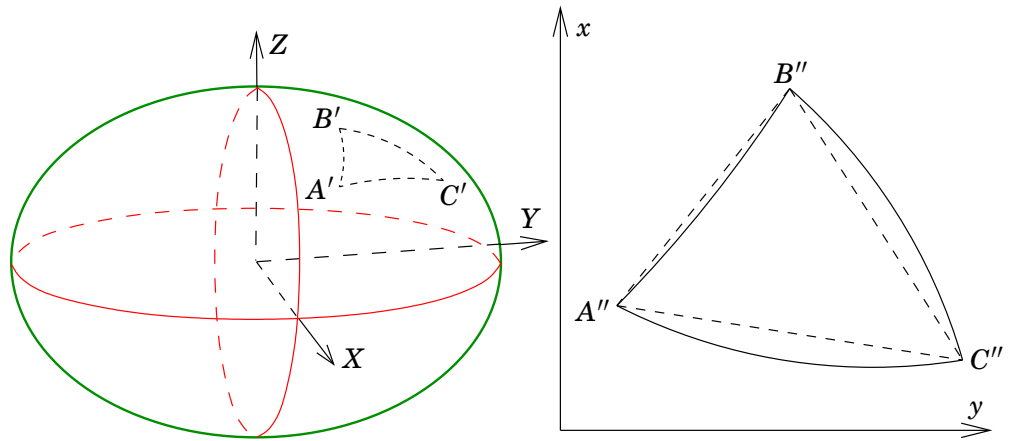
The reference ellipsoid was widely used as a computation surface as early as in the 19th century, before the existence of computing machinery and satellite positioning. The mathematics needed is complicated, but the method is more intuitive: terrestrial geodetic networks are on the

²For comparison: the mountain Olympus Mons on the planet Mars is 22 km above its surroundings, 0.65% of the radius of Mars. Gravity on Mars is only one-third of that on Earth.



(a)

Measurement in the terrain



(b)

Projected on the reference ellipsoid

(c)

In the map projection plane, "straight lines" are curved

FIGURE 8.5. Use of a reference ellipsoid and a map projection plane as reference surfaces when mapping the Earth.

Earth's surface, close to the reference ellipsoid, and the local plumb lines along which the vertical axes of measurement instruments are aligned

TABLE 8.2. Goodness of approximation by the reference ellipsoid: separation between the physical surface of the Earth and the reference ellipsoid, both in kilometres and in proportion to the Earth's radius.

Unit	km	%
Highest (Mount Everest)	+8.8	+0.138
Deepest (Mariana Trench)	-11	-0.17
Land mean height	+0.84	+0.013
Sea mean depth	-3.8	-0.06
Sea surface (geoid)	±0.1	±0.0016



are close to the *normals* to the reference ellipsoid surface.

Nowadays base networks are measured using satellite technology, and the traditional method has passed into history. GNSS networks are always adjusted truly three-dimensionally.



8.4.2 Adjustment in the map plane

The adjustment of small, local networks, such as traverses, can be done without significant error in the map projection plane.

A *map projection* is always applied in such a way that, first, for a terrain point, *geodetic co-ordinates* φ and λ are calculated on the surface of the reference ellipsoid. This is how the projection onto the surface of the ellipsoid is done. Then, the points on the ellipsoid are projected onto the map plane.

Of course the curved surface of the ellipsoid cannot be mapped onto the plane without error. The objects projected are *distorted*: directions, distances, surface areas, may all be wrong in the map plane. The map projection is chosen so, that some aspects that are considered important, are *not* distorted. Some other aspects then *are* distorted, sometimes badly. For example, a *conformal* projection maps angles and distance ratios correctly, but, as the classical Mercator projection demonstrates, it can depict surface areas spectacularly wrongly.

In conformal projections, small objects are nevertheless mapped with their correct shapes: their scales and absolute orientations may be wrong, but their *shapes are correct*. For larger objects, the “straight” lines projected from the ellipsoid are curved in the map plane. In the map plane, *directions* may be different from those on the surface of the ellipsoid³ — although in conformal projections the *angles* are identical.

Expressing and solving the network adjustment problem in the map projection plane is relatively simple, however it presupposes that

- the distance measurements are reduced, first to the reference surface, then to the map projection plane (so the *scale reduction* of the map projection has been done).
- The map projection is *conformal*, so the measured horizontal angles are directly useable without reduction. Projections used for general maps are conformal, like the Gauss–Krüger and UTM projections

³In the classical Mercator projection they are however identical, a valued property in navigation at sea.

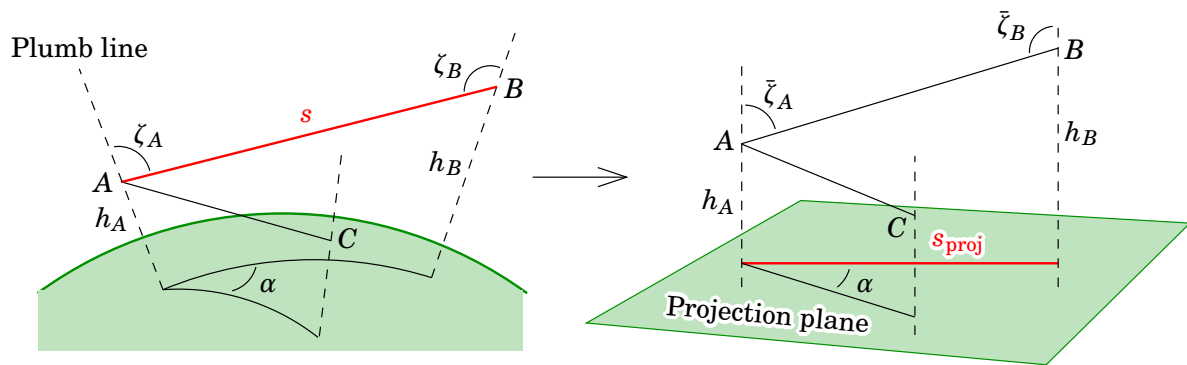


FIGURE 8.6. Transferring the geometry for adjustment of a small network to the map projection plane.



used in Finland. In Finland, map co-ordinates can thus be used directly in network adjustment.

- The zenith angles between two measurement points A and B have been measured in both directions, and the angles to be used are averages of sorts of the measurements: $\bar{\zeta}_A = \frac{1}{2}(\zeta_A + 200^g - \zeta_B)$, $\bar{\zeta}_B = \frac{1}{2}(\zeta_B + 200^g - \zeta_A)$. Then one may calculate heights from the reference level according to rectangular geometry (see subsection 6.8.3).
- Known points, like the starting and closing points as well as the auxiliary points of a traverse, have known co-ordinates in the map projection plane.

liitospiste

A visual explanation of this approach is presented in figure 8.6.



8.5 Detail survey

Detail survey (Kahmen and Faig, 1988, pages 285–303) is the stage of the measurement process, based on base-network measurements, that serves to map details in the terrain. It is the most laborious stage of the whole mapping project. A detail survey consists of collecting the data and processing it into the desired end product: a map or a digital geospatial data set. In the processing stage of detail surveying, the results of the base-network measurement come along, assuring the geometric correctness of the result.

paikkatieto

In the following we describe in more detail four classical terrestrial methods: *right-angle survey*, *tie-in survey*, *radial survey*, and *free-stationing survey*.

sidoslinja-
mittaus

A popular detail survey method is also *real-time kinematic* (RTK) satel-



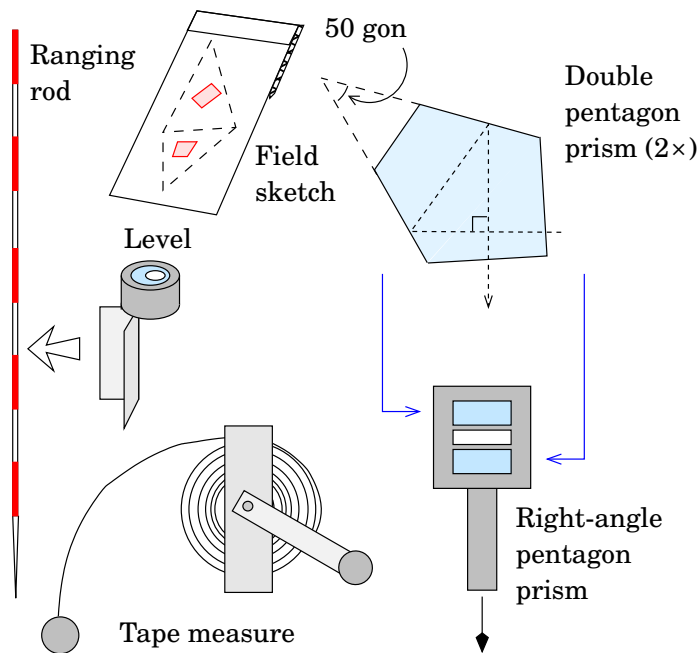


FIGURE 8.7. Tools of the right-angle survey method.

lite positioning, which is often competitive. Its use is however difficult, for example, in a high-rise urban landscapes, “urban canyons”, and impossible underground.

A new method for the detail survey of extended areas is the *mobile mapping system*. In it, a terrestrial laser scanner integrated with a GNSS receiver, an inertial measurement unit, and a cluster of cameras, is mounted on a car and driven around the area to be mapped. The amount of observational data produced is huge, and the processing work is demanding. The strength of the method is that manual field work is largely eliminated: only the signalisation of known points remains. A weakness is that occluded patches easily remain in the data. The results may be presented not only as a traditional map, but also as visual scenery like in Google Street View™.

katvealue

8.5.1 Right-angle survey (prism surveying)

A measuring tape, a double pentagon prism, ranging rods⁴ to mark out mapping lines, and drawing paper are needed for this: see figure 8.7.

linjaseiväs
4

The measurement is carried out according to figure 8.8. A and B are known points, often traverse points from a lowest-order survey. The right angles are created using a *double pentagon prism*: when one stands on

jonopiste

⁴German *fluchtstab*, French, Dutch *jalón*.



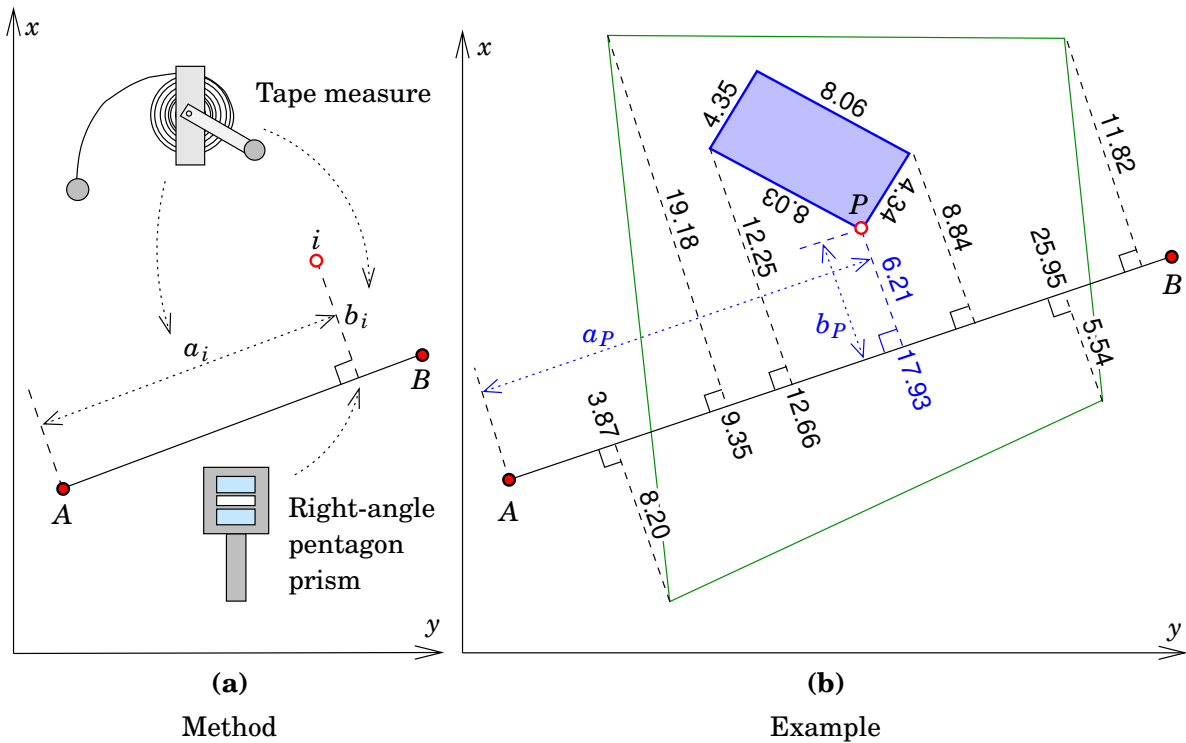


FIGURE 8.8. Right-angle survey.

the line AB , both end points (or rather, the ranging rods set up on them) show on top of each other in a device containing two pentagonal prisms. One of them looks by a right angle (100^g) to the right, the other by the same angular amount to the left, in the opposite direction.

The distances b may not be longer than one tape length (50 m).

On the right-hand side of the figure is shown how a building is measured using the right-angle method.

One should always take pains to make sure that there is sufficient *redundancy* to check for mistakes. In this example, the wall measures of the house could be measured.

eksteriööri The measurements with their number values are written on a field sketch, preferably neatly and systematically, in a way that will also be intelligible to others besides the drafter at the moment of drawing.

8.5.2 Tie-in survey

sidolinja-
mittaus Often, a method is used in which the measurement is carried out only with distance measurements in several densification steps (“Tie-in survey”, **Kahmen and Faig, 1988**) and figure 8.9. In this example a cross-measure (dashed line) of the parcel and the wall measures of the house were



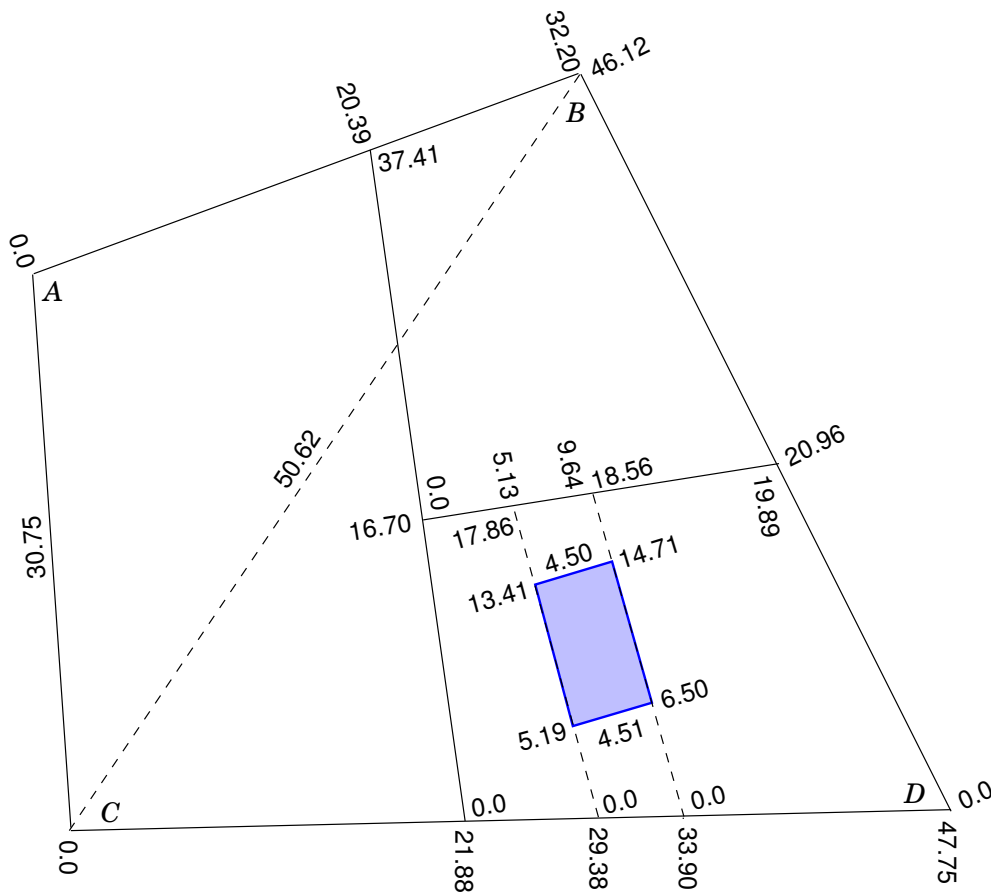


FIGURE 8.9. Tie-in survey. A cross-measure and the wall measures of the house serve as checks.

measured as checks. In this way one finds out in the field if there is an error in the measurements. A crude check can be done graphically.

More often, a mixed form is used in which the tie-in method complements the right-angle survey method.

8.5.3 Radial survey

Radial survey is explained in figure 8.10. Determination of the location in the plane of the points $i = 1, 2, \dots, n$ is done by means of measuring the angles θ_i and the (slant) ranges s_i . In the figure, the example point is $i = 2$. From a point with known co-ordinates — typically a traverse point — A , the horizontal angle θ_2 between another known *auxiliary point* B , and the point to be determined, point 2, is measured. The orientation direction α_{AB} is calculated by solving the inverse geodetic problem from the given location co-ordinates of points A and B .

geodeettinen
käänteistehtävä



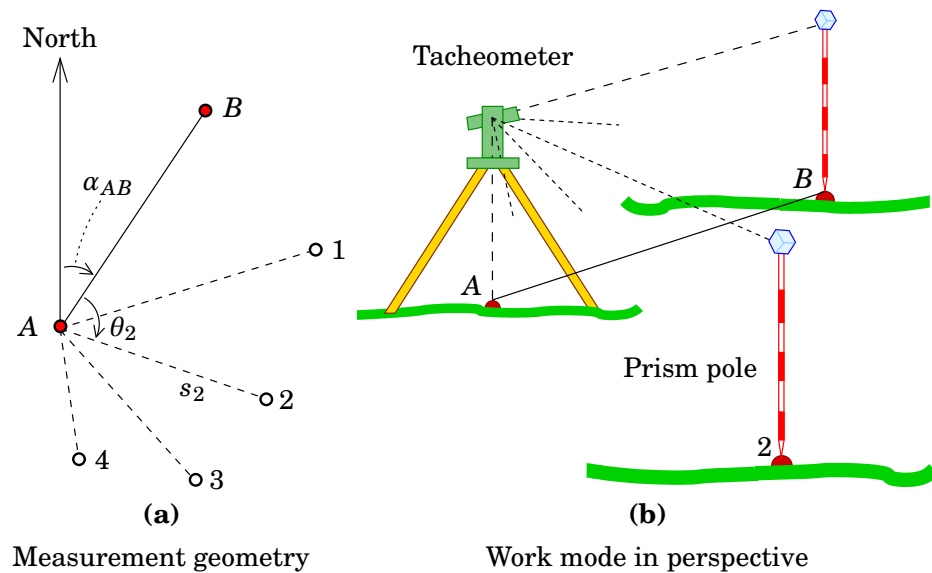


FIGURE 8.10. The radial survey method.

After this

$$x_i = x_A + \bar{s}_i \cos(\alpha_{AB} + \theta_i) = s_i \sin \zeta_i \cos(\alpha_{AB} + \theta_i),$$

$$y_i = y_A + \bar{s}_i \sin(\alpha_{AB} + \theta_i) = s_i \sin \zeta_i \sin(\alpha_{AB} + \theta_i).$$

Here, the measured slant range has been reduced to the *horizontal distance* $\bar{s}_i \stackrel{\text{def}}{=} s_i \sin \zeta_i$, in which ζ_i is the vertical or zenith angle, which is also measured.

Radial survey thus includes trigonometric height determination: as the instrument measures both the zenith angle ζ_i and the slant range s_i , one obtains, in addition to the horizontal distance \bar{s}_i , the third co-ordinate

$$z_i = z_A + s_i \cos \zeta_i.$$

The height of either point A or point B must be known as the starting value for the height computation. If the height of point B is given, it is not even necessary to measure the instrument height of the tacheometer over marker A, because the height of the prism pole will not change during measurement.

The equipment used is an electronic tacheometer or total station equipped with suitable software. The instrument is chosen based on the accuracy objective of the mapping to be undertaken.

Radial surveying is a *numerical mapping method* in which the map product is generated computationally based on the measurement values



θ_i , ζ_i , and s_i using their metadata. As the measurements are collected electronically, this is inexpensive and readily automated when the number of points to be measured is large and accuracy requirements high.

The radial surveying method is especially practicable in urban areas, because the setting out of mapping lines required in the right-angle method can be cumbersome due to traffic. Objects of which both plane and height information are required — special technical measurements, building and utility-line surveys — and busy work sites are also suitable for radial survey.



8.5.4 Free-stationing survey

In free-stationing survey — German: *freie Standpunktwahl* — the tacheometer is placed on a *freely chosen* point in the terrain, subject only to good visibility to the points to be measured as well as at least two, preferably three to four, points of which the co-ordinates are known, normally base-network points. The advantage of the method is, that the instrument does not have to be precisely centred over a known point marker or monument: the need for *centring*, and measuring instrument height, goes away. keskistys

The method has become widespread with the availability of electronic tacheometers and increased computing power. In principle, however, the method could be used with a theodolite and measuring tape. See figure 8.11. Note that there is no point marker (monument) under the tacheometer!

Let the co-ordinates of the points $A(x_A, y_A)$ and $B(x_B, y_B)$ be known. The instrument is set up on the unmarked point O . Measurements are made to the points A and B and the unknown points $i = 1, 2, 3, \dots, n$ (the example point in the figure is $i = 2$):

- horizontal directions θ
- distances s .

Ignoring heights for now, the measurement yields, in local or instrument co-ordinates (u, v) :

$$\begin{aligned} u_A &= s_A, & v_A &= 0, \\ u_B &= s_B \cos(\theta_B - \theta_A), & v_B &= s_B \sin(\theta_B - \theta_A), \\ u_i &= s_i \cos(\theta_i - \theta_A), & v_i &= s_i \sin(\theta_i - \theta_A). \end{aligned}$$

So we use the forward geodetic problem to compute (u, v) co-ordinates for geodeettinen
päätehtävä



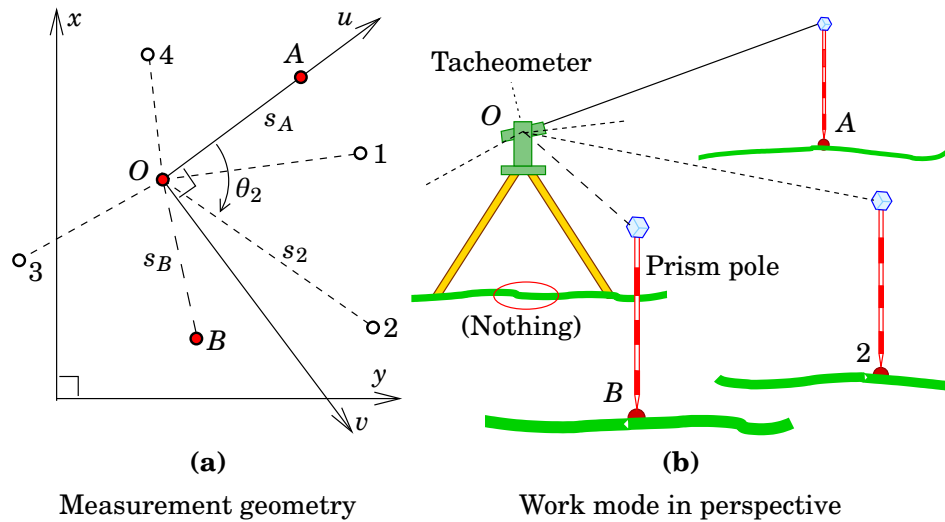


FIGURE 8.11. The free-stationing survey method.

all points $A, B, 1, 2, \dots, n$ in a co-ordinate frame of which the u axis is OA .

As we know, we can solve for the unknown parameters of the Helmert transformation if at minimum, the co-ordinates of two points, A and B , are known in both frames.

Using the Helmert transformation (section 3.6) we now transform the instrumental co-ordinates (u_i, v_i) into terrain co-ordinates (x_i, y_i) .



8.6 Carrying out a detail survey

maastomittaus Detail surveys can be carried out as topographic surveys, in which case they often cover limited areas. The instrument to use then is the *electronic tacheometer*. Alternatives are [GNSS-RTK](#) — the real-time kinematic method — or aerial mapping, which however may not always be suitable on their own due to lack of visibility in the terrain. In local measurements, traditional prism and tape measurement may be considered, but is used rarely nowadays due to its low productivity.



8.6.1 Data to be collected, mode of operation

From every measurement station, measure three-dimensionally, for every point to be measured, the horizontal direction (θ), zenith angle (ζ), and slant range (s). The instrument calculates the topocentric rectangular co-ordinates (x, y, z) and performs simple checks itself. When the whole object has been mapped, a draft printout is made at the site office, and an overall quality check is carried out.



The following data is collected:

General information

- site, date, time, weather, observer
- identifying codes as needed, see subsection 8.6.2.

From each station point

- station point (number, type), instrument height
- auxiliary points (number, type), horizontal direction, zenith angle, distance, prism height
- survey points (number, type), horizontal direction, zenith angle, distance, prism height.

Work phases

1. Choose terrain points based on the terrain and purpose of use of the measurement. For example, if forming a precise terrain model of an uneven area, points need to be collected at sufficient density.
2. Carry out the measurement: collection and pre-processing.
3. Process the material.
4. Present and archive the result. The result is a report on the measurement work, containing, among other things, a description of methods used, measurement conditions, point co-ordinates and their estimated accuracy, a draft map, and possibly calculations of areas or volumes or other relevant measurement results.

The working mode is fully digital.

Equipment and software

The processing capacity of total stations is sufficient for many uses. However,

- More and more, one sees a standard tablet or similar, loaded with versatile software, control the total station wirelessly.
- The tablet should be ruggedised for terrain use.
- Software guides the whole observation workchain in the field.
- Software enables *collection, testing, processing* and *reporting* in the field.

In the computations for topographic surveying, the separate phases of



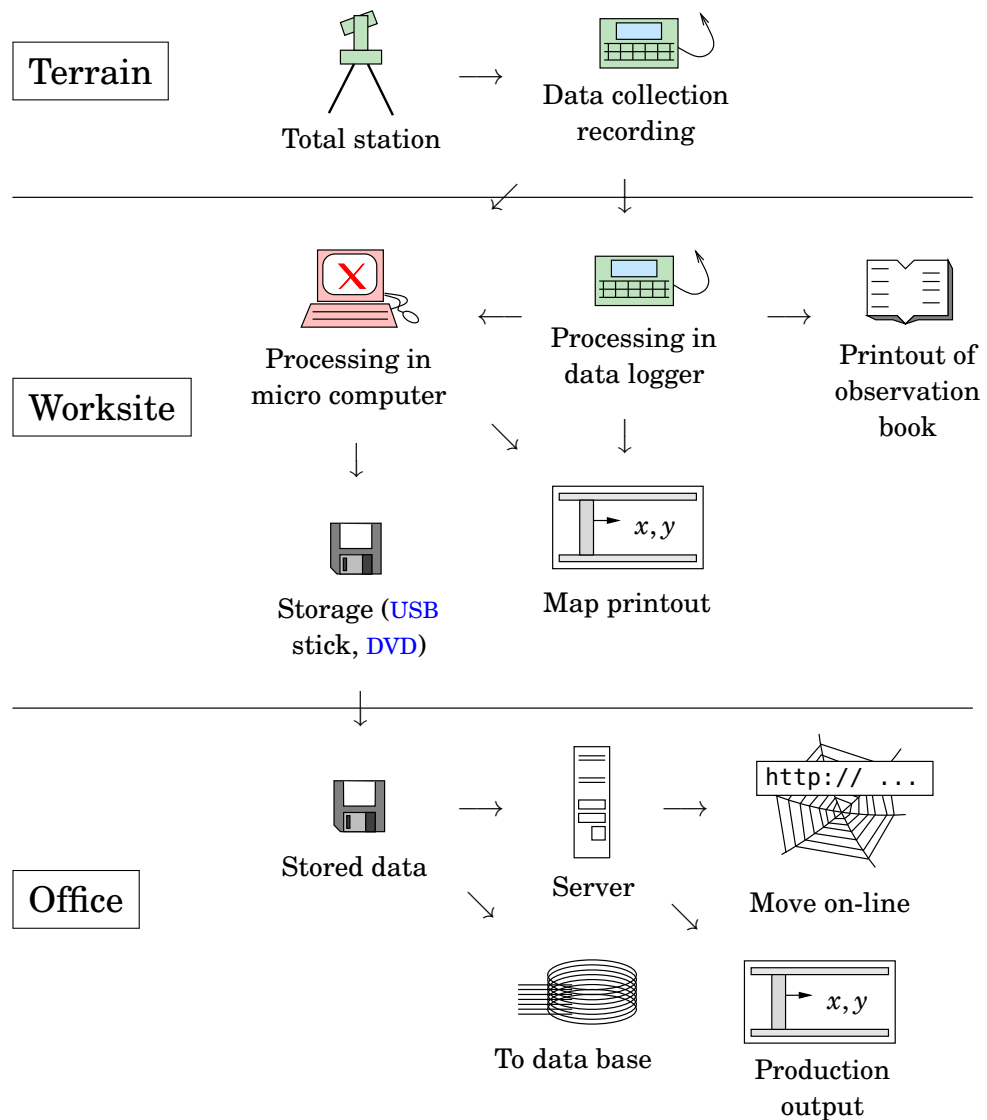


FIGURE 8.12. Workflow diagram of detail survey. The figure shows some archaic technologies such as pen plotters and diskettes that time has left behind although the work stages persist.

base-network measurement and *mapping* should also be distinguished. Both have their own routines.



8.6.2 Encoding field data

maastotieto-järjestelmä

A *topographic information system* is a special case of a geographic information system. It serves the efficient collection of topographic information by geodetic means for further processing, and thus differs from general geographic information systems.

As an extreme example of a topographic information system we can mention the proposal by the Finnish Ministry of Agriculture and Forestry



for a national topographic information system (Karlsson, 2015). It is defined as consisting of the following components (§ 3.1.2):

- The national co-ordinate and height system [probably intended are the horizontal and vertical co-ordinate reference frames].
- The *topographic database* and information management systems related to it. maastotietokanta
- All information corpora and information products collected or produced for the maintenance of the topographic database, and services maintained in order to obtain suggestions for its maintenance.
- All those information and service products that the administrator of the topographic information system produces from the topographic database in order to bring it on-line and improve its usability.

Of course most topographic information systems will not be on the national level.

Topographic survey data is collected in numerical form, and also has, besides point measurement data, *metadata*, data describing data. maastomittaus-tieto

The concept of metadata may be illustrated by the example of a topographic map: much more has been depicted on the map than just the points measured. Points form *objects*, linear (roads, streets, waterways, ...), area-shaped objects (parcel boundaries, buildings, forests and fields, ...) or three-dimensional (hills and valleys, terrain forms). Everything is depicted in different ways on the map, and the method of depiction is documented in the map's *legend*.⁵ The legend is thus in a way the metadata of an ordinary paper map. maastokartta

Documenting the measurements already in the measurement phase requires that, at the same time, *the metadata is also recorded*: does this point belong to a parcel boundary, is it the edge of a road, is it a tree (and which species), or is it just a height point in the terrain from which height contours or earthwork volumes will be calculated? To this end, *encoding methods* or *catalogues* have been developed, making the transfer of data easier and as automatic as possible. See for example OGC, Catalogue Service. korkeuskäyrä maamassa

Some aspects must be encoded *in the terrain* while doing the topographic survey:

- The point numbers (labels) of points measured in the terrain. These

⁵Legenda, Latin: *what can be read*.



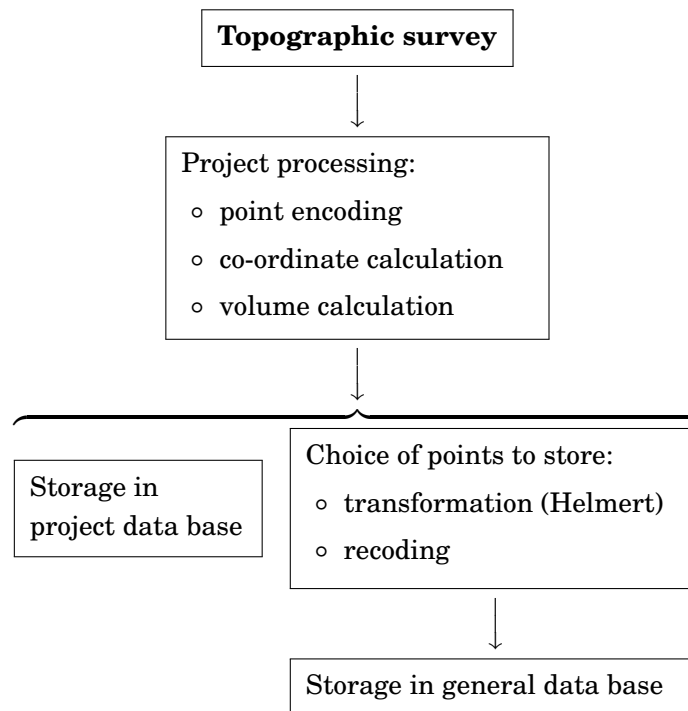


FIGURE 8.13. The encoding process for topographic data. The information may be stored into a general data base, on a case-by-case basis, for later use.



can also be generated (semi-)automatically.

- Point classification codes.
- Classification of other objects consisting of points, like lines and areas.
- Certain identifying attributes.

yksilöivä
ominaisuustieto

In connection with computation and data-base entry, the encoding may be *supplemented* in some respects, for example with topology data. The encoding carried out in the terrain is also not final because

- Not all terrain points will be entered into general data bases.
- A project-specific encoding is not suitable for general use.

Often, project-specific data is not entered into a general data base at all, but the area is mapped again when the work is done — “as-built” mapping. The idea is that one maps the finished situation, so no confusion can arise between what was implemented, and what was only planned but, after a differing implementation, never re-measured.

See chart 8.13.



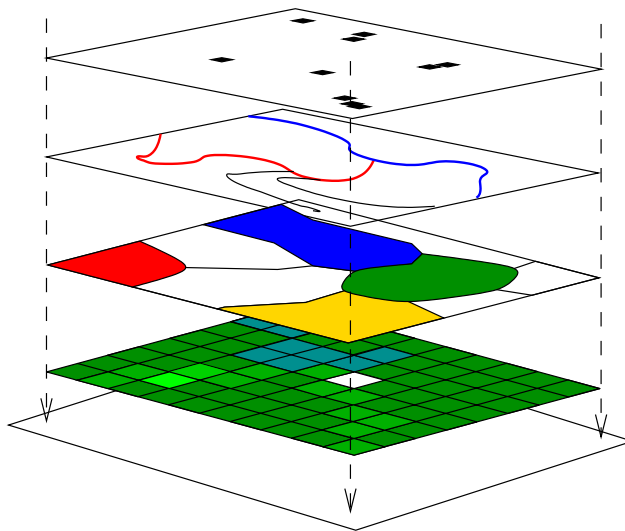


FIGURE 8.14. Attribute data of objects in multiple layers. With a geographic information system, spatial data can be efficiently combined, analysed, and refined.

Old encodings in widespread use are based on the classification of the Zoning Survey Guide, aimed at producing a map. The main starting point is *shared use* of geospatial information by the various players in the field.

Geographic information technology creates possibilities for this and many other efficient uses, such as

- The locations of objects are given in the same common co-ordinate frame. Earlier this was **KKJ** and its map projections, see subsections **3.2.1** and **3.3.1**. Today, it is always **EUREF-FIN** and its various map projection co-ordinate frames, see subsections **3.2.3** and **3.3.3**. The transformations between the co-ordinate frames are known. This facilitates the combination of different objects, producing added value.

TABLE 8.3. Classification of topographic data.

Data on nature	Cultural data	Location data	Attribute data
Topsoil type	Properties	Co-ordinate data	Identifying <i>a</i> .
Earth's surface forms	Buildings, structures	(<i>where</i>)	Locating <i>a</i> .
Soil, bedrock	Street and utility networks	Geometry data	Timestamp <i>a</i> .
Vegetation	Zoning	(<i>what shape</i>)	Descriptive <i>a</i> .
Waters	Street names, local names	Topology data	
		(<i>relations with neighbours</i>)	



- o Many methods and tools are on offer for combining, analysing and refining data from many sources.
 - ominaisuustieto – Often, various attribute data are presented on different *data layers* of a digital map, which can be processed together using various operators: *map algebra*, figure 8.14.
 - The information carried by objects can be sorted and classified according to different attributes, for example
 - * all GNSS points on a map sheet
 - * the drilling points in square $(x_a, y_a) - (x_b, y_b)$
 - kaivonkansi * the manhole covers of the municipal sewer network.
- o Spatial data may be *visualised* and, in this way, also made available to people who are not mapping professionals.

Topographic data can be classified by content into two main information types: nature data and cultural data (Salmenperä, 1998, sivut 83–84). Another way of classifying topographic data is as either location data, or attribute data. See table 8.3.

Many cultural attributes are invisible in the terrain, like place names, ownership, parcel boundaries, zoning, historical details, and so forth.



Self-test questions

1. What is the task of base-network measurement?
2. How does network hierarchy work? Why is it important? What could go wrong if it is not done properly?
3. What is reconnaissance? What are the requirements of a good point description?
4. What are commonly used computation or reference surfaces for geodetic network computation?
5. What terrestrial methods are available for detail surveys? Explain the strengths and weaknesses.
6. Explain why aerial photogrammetry cannot be the *only* method for executing detail surveys.
7. Explain why GPS — for example real-time kinematic GPS, RTK — cannot be the *only* method for executing detail surveys.
8. What is metadata, and why is it important? Give an example of metadata.





Construction surveying

9

Bergensbanen



Video: Norsk Rikskringkasting

Song: *Everloving* – Moby



9.1 Zoning plans and setting out

The types of zoning plan in use are regulated by law, in Finland the *Maankäyttö- ja rakennuslaki* (“Land Use and Building Act”) 1999/132 (Ministry of the Environment, Legislation on land use and building). They are the *local detailed plan* and the *local master plan*.¹ Both are approved by the municipality. In addition, there are still *regional plans*, which are plans at a higher level.

kaava

asemakaava
yleiskaava
1

A local master plan gives the outlines of a land-use plan for a municipal area. The plans comprise zoning base maps, the scale of which varies between 1 : 20 000 – 1 : 10 000 and 1 : 5 000 – 1 : 4 000.

kaavan
pohjakartta

According to the new Zoning Survey Guide (JUHTA, 2014), three measurement classes are defined. Every measurement class has a corresponding recommended scale for the zoning base map. Digital map products have no actual scale, but the precision at which map material is collected needs to be in accordance with the recommended scale. The co-ordinate and height reference frames are EUREF-FIN and N2000, and Gauss–Krüger is used as the map projection: ETRS-GK n , where n is the longitude of the municipality as an integer number.

Kaavoitus-
mittausohjeet

1. The *first measurement class* includes local detailed plan areas that

¹Internationally the nomenclature is highly variable: *general* or *comprehensive plan* is also used.

are built-up areas where land is extremely valuable, and where there is a local detailed plan with a binding parcel division, or a building prohibition aimed at drafting such a plan. The base-map scale of the local detailed plan is 1 : 500 or 1 : 1000. In maps that are intended to be used as part of a municipal geographic information system and used in technical planning requiring great precision, the still more precise *measurement class 1e* may be used.

2. The *second measurement class* includes local detailed plan areas that are built-up areas in which the local detailed plan to be drafted does not require a binding parcel division. The scale of the base maps is 1 : 1000 or 1 : 2000.

**ranta-asema-
kaava**
haja-asutusalue

3. The *third measurement class* includes shore detailed plan areas and lake-shore and sea-shore areas, as well as other such areas where land is clearly more valuable than land in agricultural and forestry use, such as scattered settlement areas. The base maps are drawn at a scale of 1 : 2000, in special cases 1 : 4000 or 1 : 5000.

**maastoon
merkintä**

The setting out onto the terrain of the boundary markers and location of buildings must be done according to need before actual construction starts. The accuracy of the setting out and the method of carrying out the works is strictly regulated, in Finland in the Zoning Survey Guide. The

runkoverkko base network of the zoned area may have to be brought up to standard.

**takymetri
tosiaikainen**

In setting out, an automatic total station, real-time kinematic (RTK) GNSS positioning, or any other sufficiently accurate measurement technology is used. The co-ordinates of both the known points and the points to be set out in the terrain are entered into the instrument's memory. The radial-survey or free stationing methods may be used. The instrument pre-calculates *setting-out measures*. Where to place the instrument may be flexibly decided in the terrain: point intervisibility is not always clear from the map.

**merkitsemis-
mitat**



9.2 Setting out and infrastructure

kaavoitus In connection with zoning, the infrastructure of the zoned area is planned:

- In the zoning plan, a certain intended use is assigned to an area.
- The formation of property organises the land ownership situation and boundaries as well as easements, like rights of way.
- Planning and construction implement the intended use as stated in the zoning plan, and the area is taken into use.

rasite



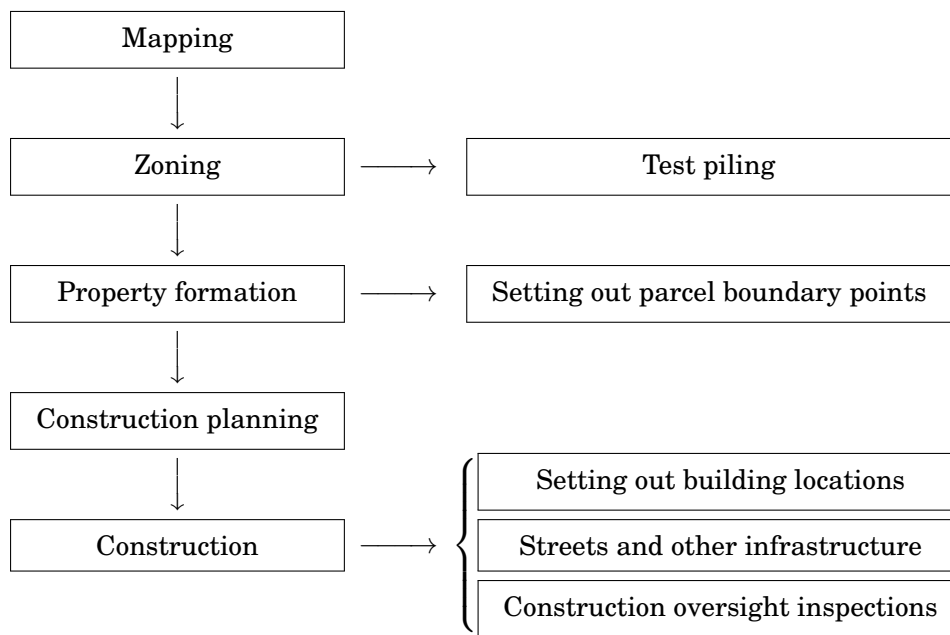


FIGURE 9.1. Setting out onto the terrain, process description.

The measurements related to the building of municipal technology or infrastructure — streets, roads, street furniture, utility lines, cables, among many other things — form their own sub-field of measurement.

Zoning-plan calculation:

- The sketch of the plan, a graphical presentation, is presented in numerical form.
- The plan is interpreted as circular arcs and straight line segments starting from known elements, figure 9.2. In the example, first a *polyline* A_1, \dots, A_7 is computed, and to this, the circular arcs K_1, \dots, K_7 are fitted.

The *zoning-plan boundaries*: the boundaries of blocks of houses, parcels, spaces, general traffic areas, recreational areas as well as areas to be built on, are to be set out onto the terrain.

- Co-ordinates are calculated for the objects to be set out onto the terrain.

See figure 9.3. Using the inverse geodetic problem, directions and distances are computed to the points to be set out, reckoned from the location of the instrument. When using the radial-survey method, the place of the instrument is chosen to be a known point.

The radial-survey method (subsection 8.5.3) in setting out works as follows:



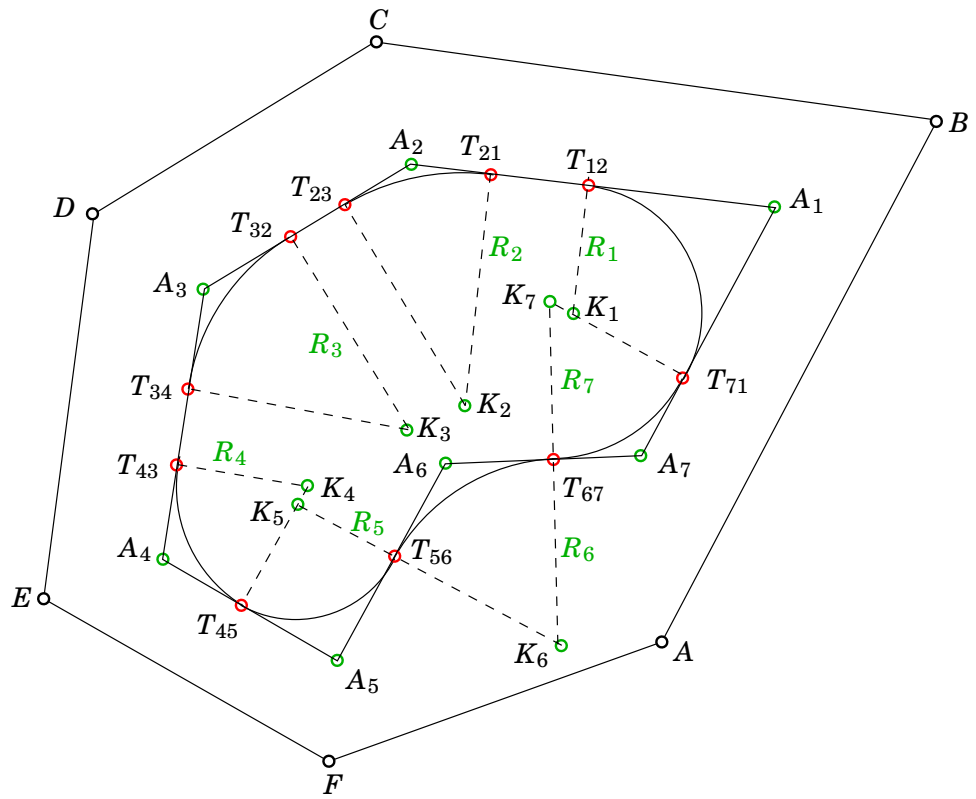


FIGURE 9.2. Zoning-plan calculation — an example.

1. The setting-out measures, angle θ_i and distance s_i , are calculated using the co-ordinates of points P_i , A , and B .

tähys

2. The instrument is set up on point A and aimed at signal B , both

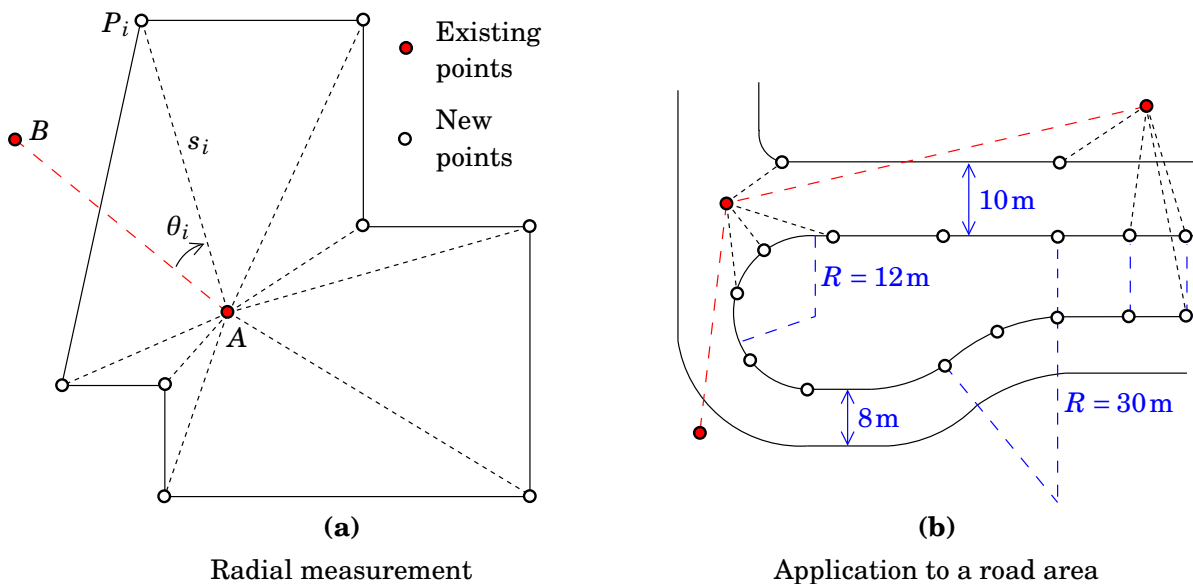


FIGURE 9.3. Setting out onto the terrain using the radial survey method.



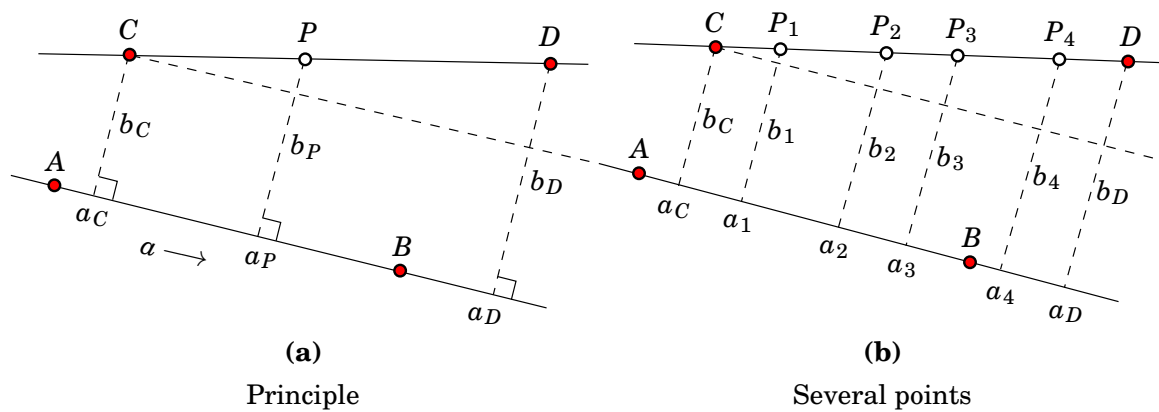


FIGURE 9.4. Straight setting-out method.

known. The reading on the horizontal circle is set to zero.

vaakakehä

3. The telescope is turned until the reading is θ_i .
4. The reflective prism is, within the field of view of the telescope, moved so that its distance reading becomes s_i .
5. A *pile* or *stake* is driven, a precise marker, on the location of which, for example, a boundary marker is built.

paalu

The free-stationing and right-angle survey methods are also being used.

9.3 Straight lines, circular arcs, rounding of corners

9.3.1 Setting out a straight line into the terrain

The *setting-out measures* (a_P, b_P) of unknown point P from the given straight line AB must be determined. If we know that point P is on the straight line CD , we may derive the setting-out measures of point P directly from the setting-out measures of points C and D : (a_C, b_C) and (a_D, b_D) , and the position of point P on the straight line CD , the distance CP :

$$a_P = a_C + \frac{CP}{CD}(a_D - a_C), \quad b_P = b_C + \frac{CP}{CD}(b_D - b_C).$$

This is the *straight method of setting out*, for example with a measuring tape and a right-angle or double pentagon prism.

Alternatively, *co-ordinates* are used. For example, if the co-ordinates of points A, B, C and D are already known, then the co-ordinates of point P are readily computed from the distance CP , and from those, the setting-out measures (a_P, b_P) of point P with respect to line AB are calculated.



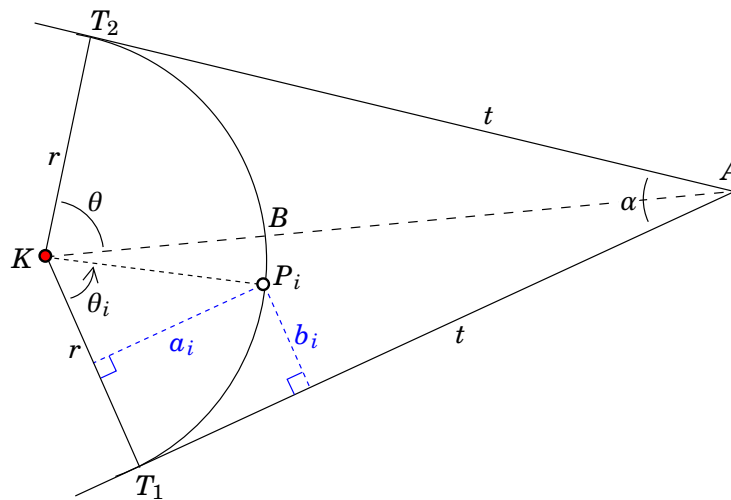


FIGURE 9.5. Rounding of corners with a circular arc.

Note that these setting-out measures are nothing but rectangular coordinates, and calculating them amounts to a co-ordinate transformation in the plane.

The result is easily generalised to the case in which there are several points P_i on the straight line CD , for all of which setting-out measures (a_i, b_i) are calculated.



9.3.2 Circular arc

Circular arcs are used widely in planning, due to their simplicity. A circular arc is defined by four parameters, figure 9.5:

- angle α between the tangents
- half the arc centre angle, θ
- arc radius r
- tangent length t .

There are two dependencies between these parameters:

$$\begin{aligned} \theta = 100^g - \frac{\alpha}{2} &\iff \alpha = 200^g - 2\theta, \\ t = r \tan \theta &\iff r = t \cot \theta. \end{aligned}$$

Setting out the arc onto the terrain is done as follows:

1. Normally the calculation of the straight lines has already determined the intersection point A of the two tangents, and the angle α .
2. Choose one further parameter: the radius r , and calculate the others using the equations given above.



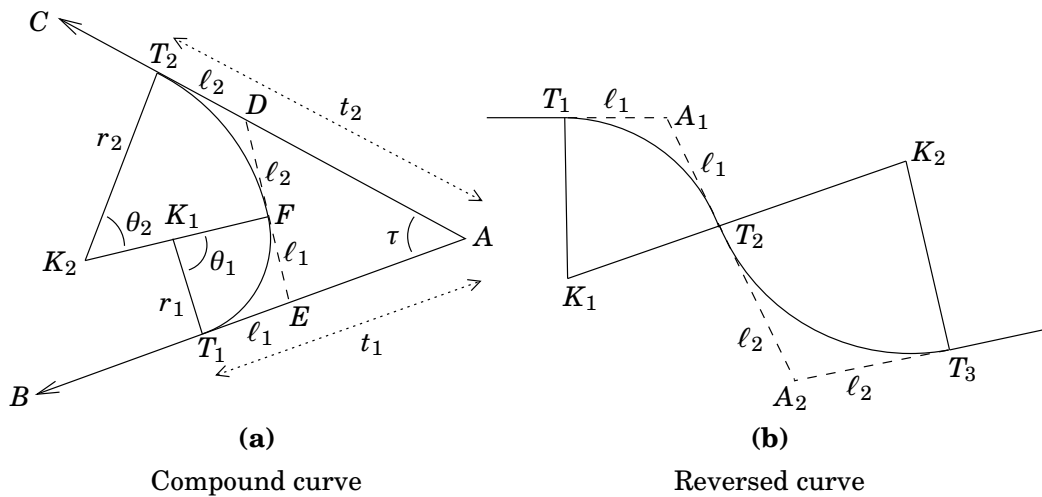


FIGURE 9.6. Rounding of corners with a compound curve. The reversed curve is a special case of the compound curve.

3. Measure the distance t from A along the tangent, yielding the tangent points T_1 and T_2 .
4. From both of these determine, with double pentagon prism and measuring tape, the centre point of the arc K (redundancy!)
5. On line KA now mark the arc point B .
6. From centre point K mark as many arc points as needed, using the distance r . One example point P_i is marked in the figure.
7. The rectangular setting-out measures of point P_i : (a_i, b_i) , are also readily obtained.

The rounding of corners is commonly done using circular arcs. Combining straight lines and circular arcs is very common in local detailed zoning maps, as in building plans in general. The lines are connected according to the situation using various conditions, which ensures continuity and the appearance of smooth curvature. Here we present a few examples.

9.3.3 Compound curve

An example case is the connection of two straight lines with two (or more) circular arcs bending in the same direction, a *compound curve*, figure 9.6a, Jamal (2017). So, at its simplest we have two straight lines and two circular arcs, that at their junctions — there are three of them — are parallel.

The situation of figure 9.6a has the following parameters:

- the lengths of the tangents $t_1 = AT_1$, $t_2 = AT_2$



- the radii of the arcs r_1, r_2
- the intersection angle of the tangents τ
- the central angles of the arcs θ_1, θ_2 .

There are multiple dependencies between parameters:

- In triangle $\triangle ADE$ one sees immediately that $\angle ADE = \theta_2$, because $K_2F \perp DE$ and $K_2T_2 \perp AC$, and similarly $\angle DEA = \theta_1$. Therefore

$$\tau + \theta_1 + \theta_2 = 200^g.$$

Furthermore $T_1E = EF = \ell_1$ and $T_2D = DF = \ell_2$.

- $\ell_1 = r_1 \tan \frac{1}{2}\theta_1, \ell_2 = r_2 \tan \frac{1}{2}\theta_2$.

- By the sine rule

$$\frac{AD}{\sin \theta_1} = \frac{AE}{\sin \theta_2} = \frac{DE}{\sin \tau} \implies \frac{t_2 - \ell_2}{\sin \theta_1} = \frac{t_1 - \ell_1}{\sin \theta_2} = \frac{\ell_1 + \ell_2}{\sin \tau}$$

and by substitution

$$\frac{t_2 - r_2 \tan \frac{1}{2}\theta_2}{\sin \theta_1} = \frac{t_1 - r_1 \tan \frac{1}{2}\theta_1}{\sin \theta_2} = \frac{r_1 \tan \frac{1}{2}\theta_1 + r_2 \tan \frac{1}{2}\theta_2}{\sin \tau}.$$

Thus we can calculate *all* seven parameters catalogued above, if given

- two of the three angles τ, θ_1, θ_2 , and
- two of the four lengths r_1, r_2, t_1, t_2 .

In the case depicted in the figure, the setting out onto the terrain is done as follows:

1. Measure the distances $t_1 - \ell_1$ and $t_2 - \ell_2$ from A along the tangents, yielding the points E and D .
2. The tangent intersection angles of the *individual circular arcs* at points D and E are $T_1EF = 200^g - \theta_1$ and $T_2DF = 200^g - \theta_2$.
3. After this, the setting out is done *separately* for circular arcs 1 and 2 in the way already explained above.



9.3.4 Reversed curve

The alternative case where the circular arcs bend in opposite directions (but otherwise the situation is similar to that of the compound curve) is the *S-curve*, also reversed curve, figure 9.6b, Jamal (2017). What makes the situation a little cumbersome is the possibility that the intersection point A of the straight lines² T_1A_1 and T_3A_2 may not exist, namely if the lines are parallel.

²The notation used in the S-curve figure does not directly match that in the compound curve figure.



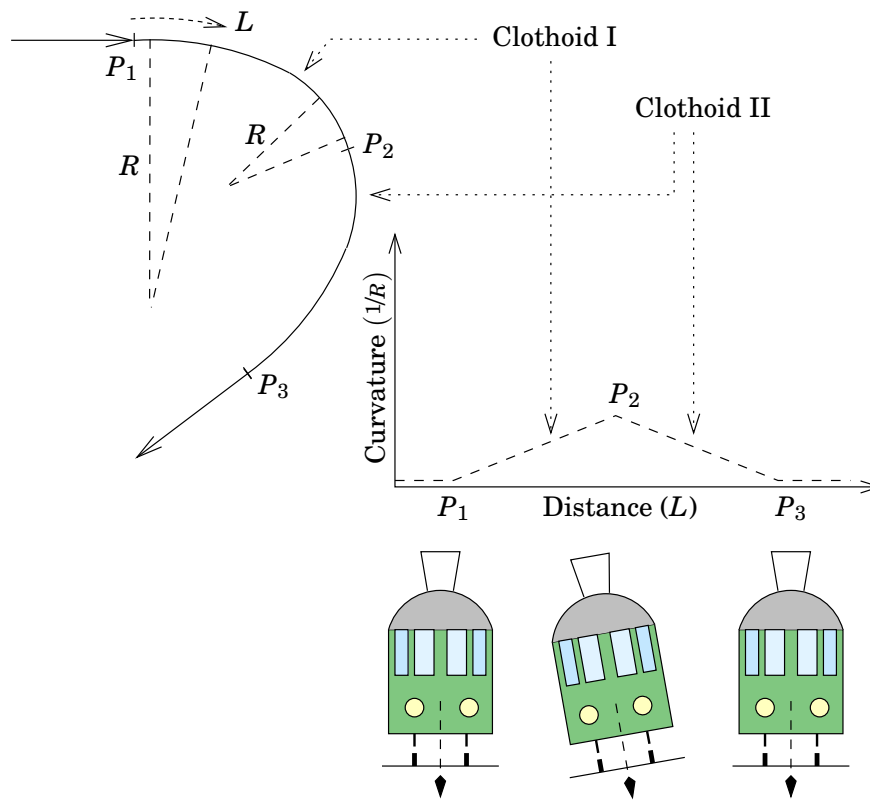


FIGURE 9.7. Principle of the clothoid.



9.4 Transition curve

A transition curve, Euler spiral ([Wikipedia, Euler spiral](#)) or *clothoid* is used in the planning of railways and fast motorways. In fast traffic, not only the centre line of the road, but also its *curvature*³ must be continuous, for the following reasons:

- For example, the control of the movement of an articulated lorry through the steering wheel is slow.
- The surface of the road or railway is tilted in the sideways direction against the centrifugal force. This *transversal tilt*, which is proportional to the curvature of the road, may change only slowly and continuously along the length of the road.

poikittais-
kallistuma

For these reasons, a combination of straight lines and circular arcs is unsuitable.

The equation of the clothoid is

$$RL = A^2,$$

³Curvature is the inverse of the *radius* of curvature!



in which A is the *parameter* of the clothoid, L is the distance along the clothoid, that is, along the road, and R is the local radius of curvature. As can be seen, the radius of curvature changes as a continuous function of the distance

$$R = \frac{A^2}{L}.$$

- ⁴ At constant speed, the centrifugal force⁴ F is inversely proportional to the radius of curvature:

$$F = \frac{v^2}{R} = \frac{v^2}{A^2}L, \quad (9.1)$$

so, assuming that the *design speed* of the road, v , is a constant, the centrifugal force, and thus also the transversal tilt of the road surface needed, is a *linear* function of distance travelled. This explains the suitability of the clothoid curve as the shape of motorway and railway curves.

Of course, care must be taken that the clothoid satisfies the continuity condition for the radius of curvature at the start and end points, with another clothoid, with a circular arc, or with a straight line ($A = R = \infty$). See figure 9.7, in which a straight line links to clothoid I (P_1P_2), which links to clothoid II (P_2P_3), which again continues as a straight line from point P_3 onward. The train arrives at point P_1 upright. In the interval P_1P_2 it tilts sideways at a fixed rate, and arrives at its maximum tilt angle at point P_2 . In the interval P_2P_3 the tilt diminishes linearly, and at point P_3 the train is again upright, and continues its journey straight ahead.

If the speed of the train is equal to the design speed v according to equation 9.1, the resultant of gravity and centrifugal force is always *perpendicular* to the floor of the train, and the passengers do not notice anything.

The clothoid is also used in planning motorways, although there, the true speeds of vehicles will vary.



9.5 Road and street surveying

Roadbuilding comes with the following measurement and calculation tasks:

- maastomittaus ○ Conducting a topographic survey in order to draft the road-plan

⁴... more precisely, the *pseudo-force* per unit of mass experienced by the travellers in the vehicle, which acts however from their viewpoint in precisely the same way as gravity.



base map, for example using aerial mapping (photogrammetry), as a basis for road planning.

- Calculating the trajectories in co-ordinates and setting-out measures.
- Calculating the co-ordinates, setting-out measures, if necessary earthwork masses and other volumes, for the road structures and special structures (bridges, tunnels, underpasses among others).
- Carrying out on-site measurements for building the road, setting out.
- After project completion, usually an “as-built” survey is carried out.

maamassa

The same work phases also occur in other technical measurements related to large-scale construction projects.



9.6 Construction surveying

Construction surveying measurements include the installation measurements of buildings, parts of buildings, bridges, tunnels, reservoir dams, industrial machines and similar structures.

vesiallas

- The measurement starts from the base network. First, a project *measurement base* is created, a sufficient set of horizontal and vertical benchmarks in the project area. Base-network points are used as benchmarks. Densification points or “use points” are created as needed, in a hierarchical fashion.
- The actual measurements are carried out from the use points. The measurements are done separately as horizontal and height measurements, and the point sets are also partly separate.
- Both national and international standards have been created to guide the measurements.

kiintopisteistö



9.6.1 Setting out a building location onto the terrain

When the building permit has been obtained, the builder can apply to the authorities for a decision to *set out the building location onto the terrain*.

paalutus,
maastoon
merkintä

The procedure has *three objectives*:

- Setting out the building location onto the terrain, verifying that no part of the building is too close to the parcel boundary.
- Checking the correctness of the wall measures of the building, important information for the builder.



- Verifying the correctness of the building's height location.

The land survey authorities only measure with a view to assigning where to build, unless agreed otherwise. The precision and number of points is not enough to actually start building. The builder can continue the survey work in connection with the actual building work.

jonopiste Either a traverse point or a boundary marker is chosen as a site reference point. If there are not enough of those nearby, or not of the required precision, the first task is measuring new *base points*. Map co-ordinates (x, y) for these are calculated and they are marked on the site plan.

Wall measures, cross-measures, distances from boundaries, and distances from pre-existing buildings may be used as *control measures*.

linjavaaitus korkeuskiintopiste The height location of a building is obtained by traverse levelling, which runs from one general height benchmark to another. Close to the building site, *at least two* height base points are created for the later work, unless nearby there are already enough general points. The correct height of a building is critical for functioning sewers and, in low-lying locations, flood safety.



9.6.2 Location review of a building

sijainti-katselmus

In *location review* it is verified that the building is in the correct place and at the correct height. The review is carried out when the foundation of the building is completed. After acceptance, construction may continue.



9.7 Other measurements



9.7.1 Technical measurements, deformation measurement

Technical measurements, or *engineering geodesy*, form their own specialised discipline. This also includes precise deformation measurements:

- Deformation monitoring, during construction and afterwards over the object's life cycle. Objects: reservoir dams, tunnels, bridges, other large structures, skyscrapers, and so on.
- Nowadays often *monitoring measurement* using automated equipment.
- Industrial measurements indoors or outdoors, installation measurements of large machines, paper machines, shipyards.
- Engineering surveying.



- Tunnel and mine surveying. Tight constraints on the measurement geometry, as well as the non-functioning of GNSS, are characteristic of these measurements.

In these measurements, *traceability* of the quantities measured is central. Careful (system) calibration and good metrological practice are of great importance.



9.7.2 Machine guidance of working machines

GNSS technology is used widely, in addition to more traditional positioning technologies, for guiding *working machines* in real time. The reliability requirements are obviously tough if an expensive working machine is being guided to lay, for example, asphalt on a motorway. Machine downtime is costly, errors even more so.

Terrestrial real-time guidance is usually used in the construction of reservoir dams, bridges, tunnels, and other infrastructure. This kind of technology was used in the construction phases of the Dutch Easter Scheldt storm-surge barrier (figure 9.8) and the Danish Great Belt and Sound bridge-tunnel solutions, as in many similar projects.

Shipping containers and cranes in harbours are also positioned in real time using GNSS technology in order to improve efficiency (Pitkä, 2009).

Agricultural and forestry machines may be guided by real-time GNSS (“precision farming”), so seedstock, fertiliser, and pesticides can be administered with precision according to very localised needs.



9.7.3 Underground utility-line mapping

Only part of the underground utility lines, like telephone, data and electric power cables as well as sewage, water and city heating pipes, have been mapped satisfactorily, mostly on maps prepared by various municipal institutions, of which there may be many per municipality. When utility-line surveys are this decentralised, the geodetic quality of the maps will vary.

Cities generally use the presentational and preparation practice of the 1974 utility-lines mapping standard SFS 3161. The standard was renewed in 1996.

Utility-line maps are used for many needs: in zoning for planning technical maintenance and the network, in connection with construction work, for maintaining the lines by the owner institution, and in connection with



jäljitettävyys

myrskyvuoksi-
suoja
Iso-Belt
Juutinrauma

täsmäviljely

johtokartoitus

johtokartta-
standardi



FIGURE 9.8. The AGA/Minilir infrared tracking device — French military technology! — in use on the building site of the Dutch Easter Scheldt storm-surge barrier, 1980. © 2016 Nicolàs de Hilster, (De Hilster), figure 4, with permission.

managing crisis situations and damages. The scale is most commonly 1 : 500.

Mapping methods:

kiintopiste

- The measurement should be based on general benchmarks available in the area, so that the result is obtained in the same system. If necessary, a *network densification* is carried out, meaning new benchmarks are created following the Zoning Survey Guide.

kartoitus-
mittaus

- The same measurement methods are used as more generally in detail survey, section 8.5.
- The mapping of new lines is done during construction, when the lines are still visible.

kaivonkansi

- The visible parts of old lines (manhole covers, distribution cabinets) are mapped. The underground parts can sometimes be located with a metal detector. Ground-penetrating radar (GPR) and electrical resistivity tomography (ERT) have also

maatutka



been used.

Underground utility-lines marking service: The lines are marked on the terrain for the builder, in order to avoid costly damage caused by excavation work. The service is usually offered by the _____ owner of the line, for example a power company.

johtojen
näyttöpalvelu



Self-test questions

1. What is zoning and why is it necessary?
2. How many measurement classes are used in Finland? How are they defined? What scales are the zoning maps on for each class?
3. Describe the various methods of setting out objects onto the terrain.
4. What is a compound curve? A reversed curve?
5. How many independent parameters uniquely determine a compound curve?
6. Explain why clothoids are used for fast roads and railways.
7. Describe the Finnish practice of “location review”. At what stage of construction does it take place?
8. What types of objects may deform in ways that can be monitored by precise measurements?
9. How are underground utility lines mapped?



Digital terrain models and volume calculation

10



Measuring the volume of a wine barrel with a measuring rod, inspiration for Johannes Kepler's "wine-barrel formula", or Simpson's rule (image [Deutsche Fotothek, 1523](#))

CONSTRUCTION IN THE built environment and its planning, as well as the technical measurements made in those contexts, make extensive use of *digital height and terrain models*.

The terms *digital height model*, **DHM**, *digital elevation model*, **DEM**, or *digital terrain model*, **DTM**, refer to a file consisting of points on the Earth's surface, which describes, more or less well, the forms of this surface. High-resolution terrain models are expensive to produce over large areas, but are nevertheless available for many countries. Terrain models for Finland are produced by the National Land Survey.

Information about the Earth's surface and its forms can be obtained using topographic surveying measurements, photogrammetrically from aerial imagery, and by *scanning* from aircraft or satellites. The scanner may be laser based or can be a microwave radar, a [SAR](#), or *synthetic-aperture radar*, which interferometrically achieves a very high resolution.

Global terrain models

1. The older model GTOPO30 (US Geological Survey, original 1996, [USGS](#), [GTOPO30](#)). The resolution is 30", about one kilometre. The model was updated with data from the [SRTM](#), see item 4 below. The model contains no sea-depth data.
2. The GLOBE model (Global One-km Base Elevation project ([NOAA](#) and many others, [GLOBE Task Team and others, 1999](#)). The resolution is also 30" \approx 1 km. The current model version contains no sea-depth data.
3. The ETOPO1 model, which supersedes the older ETOPO5 and ETOPO2 models ([NOAA et al. 2008](#)). The resolution is 1', about 1.8 km. The model contains depth data in addition to elevation data.
4. The [Shuttle Radar Topography Mission](#) imaged the Earth's topography between latitudes 60°N and 56°S. The flight took place in February 2000. The resolution is one second of arc on the Earth's surface, about 30 m. In 2014, *all* [SRTM](#) data was declared public. The data does not include sea depths.

The national terrain model New technologies, like airborne laser scanning, have already been operational for many years and are in widespread use. In Finland, the National Land Survey has also for a long time been scanning various areas in order to build a new, precise national terrain model. There are two models: model [KM10](#) is currently complete at a spatial resolution of 10 m, [Finnish National Land Survey, Elevation model 10 m](#), and model [KM2](#) is partially complete at a resolution of 2 m, [Finnish National Land Survey, Elevation model 2 m](#). The [KM2](#) model is planned to be completed by 2020. Based on the [INSPIRE](#) directive ([INSPIRE Knowledge Base](#)), the data is free of charge.

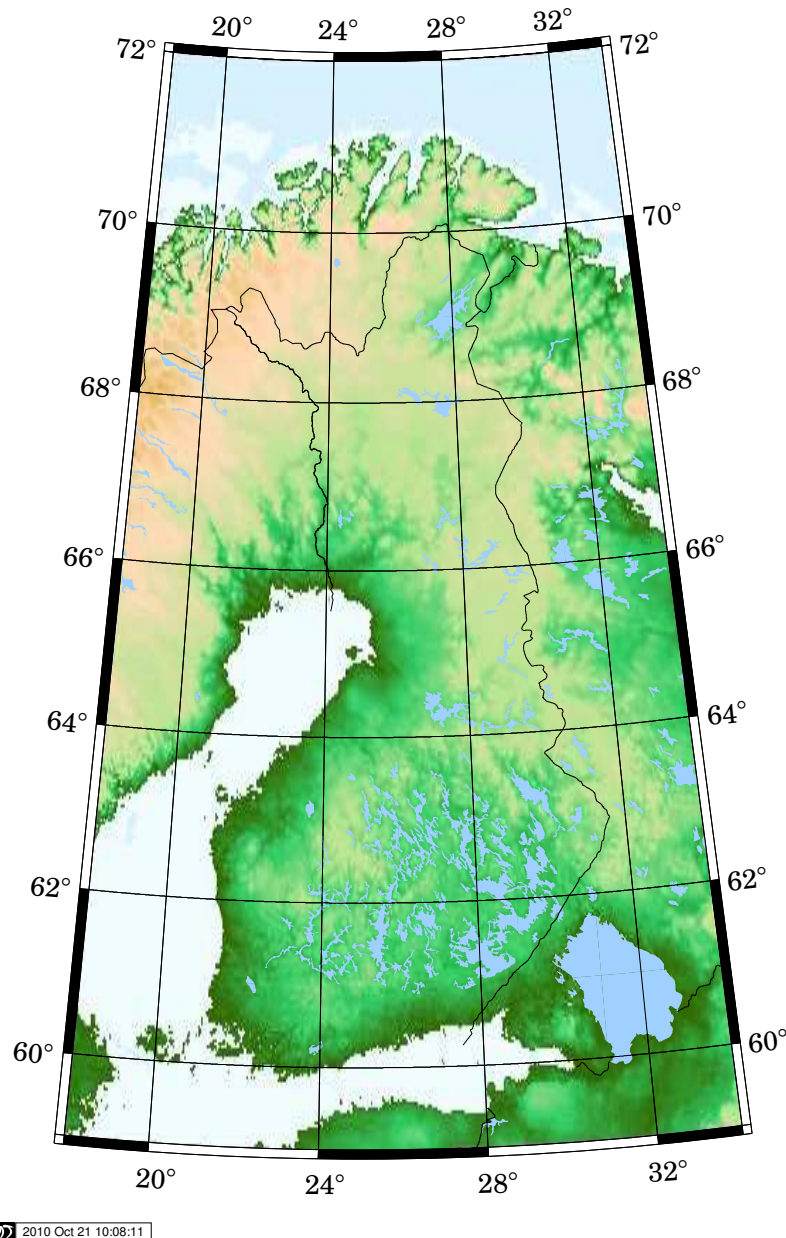


FIGURE 10.1. The global terrain model ETOPO2 version 2 on the Finnish territory.



10.1 Terrain-model measurement, construction, presentation



10.1.1 Measurement geometry

- The points to be measured do not generally form a regular pattern, and may be freely chosen within the constraints of the measurement technique used.
- The random-point method is a statistical sampling method. The sampling density may be higher where terrain forms are more otanta



10.1.3 Model presentation

With terrain models, different *forms of presentation* are used:¹

1

Point-grid presentation This agrees well with the way computers operate: the handling of large amounts of data is also straightforward and easy.

A regular grid may be square, rectangular, or more complicated, like a hexa tile pattern (“beehive”) or a triangular pattern.

Triangulated-network presentation Here, points representative of the terrain forms are chosen and connected by lines into a cover of triangles. A well-known mathematical triangulation technique is *Delaunay* triangulation, which gives beautiful triangles of which the sides are as equal as possible.

Delaunay'n
kolmiointi

A triangulated network type of terrain presentation is more difficult to manipulate, but it is also able to present difficult terrain forms, like sharp edges, better than a grid presentation, using a smaller number of points. Moreover, if the resolution of the terrain model varies by area, the triangulation presentation is better, because the sizes of the triangles vary with resolution.

In the literature, the method is referred to as **TIN**: *triangulated irregular network*.

CAD (computer-aided design) software, in use in planning offices, knows how to use digital terrain models and how to display them in many different ways; for example as a perspective image. Building plans are also in digital form and can be combined with this.

Let us also mention in this connection the multi-resolution “tiling” methods which are based on the discrete wavelet transform (**DWT**, [Wikipedia, Discrete wavelet transform](#)) and are meant for the interactive presentation of materials, for example [Fraser et al. \(2013\)](#). The image format **JPEG 2000**, as well as Google Earth™, are also based on this technique. This form of presentation enables rapid changes in viewing location and zooming. It is even more suited than Delaunay triangulation for presenting materials of highly variable resolution interactively.

aalloke-
muunnos

10.2 Use of terrain models

Terrain models are used for, among other things:

¹In image processing one speaks in analogous fashion of pixel and vector graphics.



- orto-oikaisu
 - The ortho-rectification of aerial photographs: the removal of projection errors of aerial imaging in the production of orthophoto maps.
- korkeuskäyrä
 - The calculation of the height contours to be shown on a map.
 - The planning of a traffic route (road, street, waterway, electric power line, ...), to minimise (under other constraints, like maximum slope or minimum radius of curvature):
 - earthwork volumes to be moved, section 10.4
 - the difference between soil to be removed and soil to be added
 - consumption of fuel, travel time by a typical vehicle using the route.
- maamassa
 - The creation and visualisation of three-dimensional landscape models, in support of planning and public debate over the plans.
 - The resolution of visibility issues, like in connection with the placement of cell-phone or radio masts.
 - The planning of ski pistes.
 - The support of a military application: automatic navigation in low flight of cruise missiles, but also jet fighters, hiding from radar “inside” the terrain.
 - The calculation of the gravity effect of terrain masses (terrain correction) in gravity-field and geoid determination.
 - The provision of the lower boundary in numerical weather prediction (NWP) and climatic general circulation models (GCM).
 - The creation of realistic landscapes for video games and flight simulators.
 - Many others.

Strictly speaking, a digital elevation model (DEM) describes the heights, not only of the terrain, but also of buildings, the forest canopy and the like, whereas a digital terrain model (DTM) describes only the heights of the terrain itself. Nevertheless the words are often used as synonyms.

latvusto



10.3 Calculating surface areas

Calculating surface areas is discussed in Kahmen and Faig (1988) in section 8.6.

merkitsemis-
mitat

A handy way of calculating is to use setting-out measures with respect



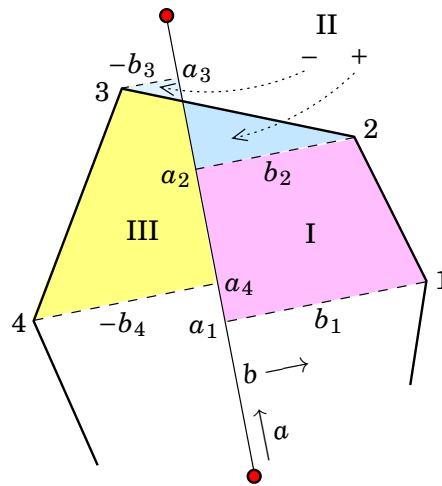


FIGURE 10.3. The use of setting-out measures in calculating surface areas.

to a baseline. In figure 10.3, the setting-out measures form trapezoids.

Area I is calculated as follows:

$$A_I = \frac{1}{2} (a_2 - a_1) (b_1 + b_2),$$

and area III as follows:

$$A_{III} = \frac{1}{2} (a_4 - a_3) (b_3 + b_4),$$

in which one should pay attention to the algebraic signs.

etumerkki

Area II is also obtained in a similar way, although it is the difference between two surface areas. It is nevertheless formally a trapezoid:

$$A_{II} = \frac{1}{2} (a_3 - a_2) (b_2 + b_3).$$

All equations can be made compatible by agreeing, for example, that the b values are positive on the right side seen when moving in the direction of a , and negative on the left side. Also, the a indices are followed in numerical order, in the example case, anti-clockwise. Then we obtain for the total area, by summation,

$$A = A_I + A_{II} + A_{III} + \dots$$

and all the algebraic signs, also those of the small cancelling triangles, are automatically correct.

If we have the use of co-ordinates, there are other ways of calculating surface areas. The total surface area is obtained as a sum of trapezoids (the i index is circular: $n + 1$ is the same as 1):

$$A = +\frac{1}{2} \sum_{i=1}^n (x_{i+1} - x_i) (y_{i+1} + y_i), \tag{10.1}$$



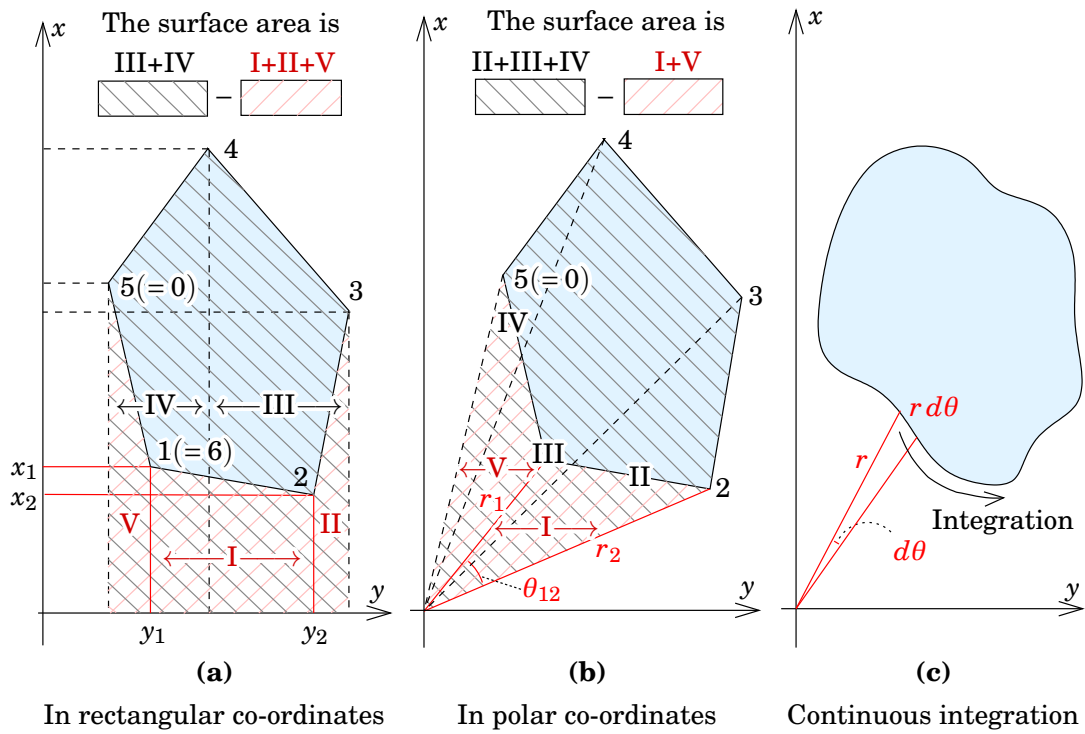


FIGURE 10.4. Calculating surface area.

and by interchanging x and y :

$$A = -\frac{1}{2} \sum_{i=1}^n (y_{i+1} - y_i) (x_{i+1} + x_i). \tag{10.2}$$

From equation 10.1 we obtain

$$A = -\frac{1}{2} \sum_{i=1}^n x_i (y_{i+1} + y_i) + \frac{1}{2} \sum_{i=1}^n x_{i+1} (y_{i+1} + y_i),$$

and by re-numbering the second term — as the i index is circular:

$$A = -\frac{1}{2} \sum_{i=1}^n x_i (y_{i+1} + y_i) + \frac{1}{2} \sum_{i=1}^n x_i (y_i + y_{i-1}) = \frac{1}{2} \sum_{i=1}^n x_i (y_{i-1} - y_{i+1}).$$

Similarly equation 10.2 yields

$$A = \frac{1}{2} \sum_{i=1}^n y_i (x_{i+1} - x_{i-1}).$$

These equations are known as the *shoelace formulas* (Wikipedia, Shoelace formula).

If equations 10.1 and 10.2 are added together and divided by two, we obtain

$$A = \frac{1}{2} \sum_{i=1}^n (x_{i+1}y_i - y_{i+1}x_i).$$



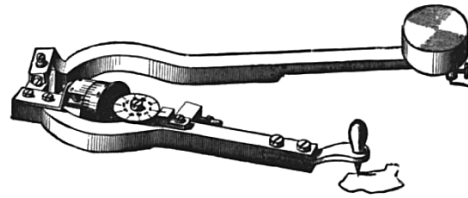


FIGURE 10.5. A polar planimeter from 1908 (Wikipedia, Planimeter). The surface area is measured by drawing along the outer edge of the figure.

This equation calculates the surface of a polygon as the sum of *triangles extending from the origin*. It can be shown — a graphical proof is given in figure 10.6 — that the surface area of such a triangle (example in the figure) is

$$A_{i,i+1} = \frac{1}{2}(x_{i+1}y_i - y_{i+1}x_i) = \frac{1}{2}r_i r_{i+1} \sin \theta_{i,i+1}. \quad (10.3)$$

This *planimeter equation 10.3* is the principle of operation of the *polar planimeter*.² Of course the equation can also be used directly numerically,² if the figure is given in polar co-ordinates:

$$A = \frac{1}{2} \sum_{i=1}^n r_i r_{i+1} \sin \theta_{i,i+1}.$$

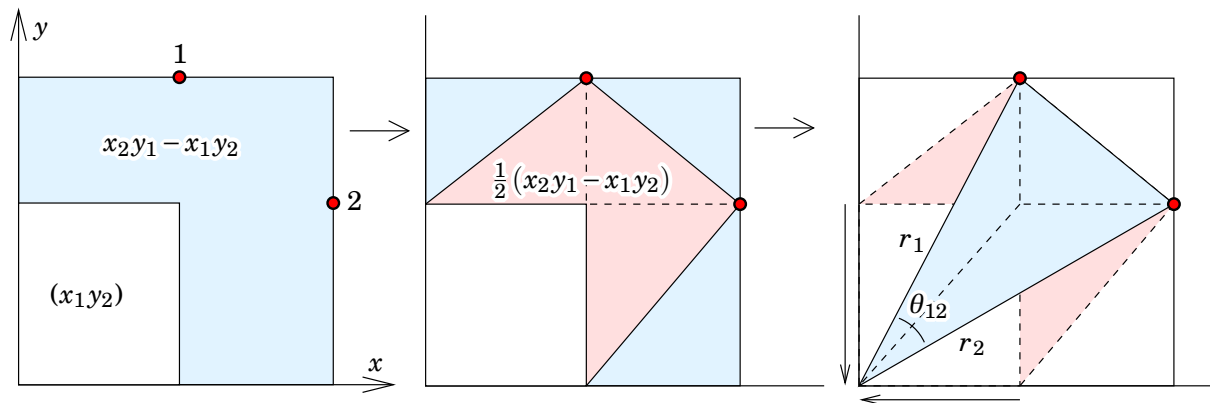


FIGURE 10.6. Graphical proof of the planimeter equation.

²The polar planimeter integrates mechanically the expression

$$\frac{1}{2} \oint r^2(\theta) d\theta,$$

which is the surface area of the closed figure.





10.4 Volume calculations

maamassa The calculation of earthwork masses is explained in **Kahmen and Faig (1988)** sections 14.2 and 14.3. Often, the case is about calculating amounts of gravel, sand and similar building materials to be moved in connection with, for example, road-building.

The means of measurement or determination are

- pintavaaitus**
 - Area levelling or use of a laser level. Only small projects, labour-intensive.
- ”pekkaniiskaTM”**
 - Photogrammetry from the air, or terrestrially from an aerial work platform.
- laserkeilaus**
 - Laser scanning, from the air or terrestrial — high productivity.
 - Height contours from a map. Computationally, the method is similar to surface-area calculation, section **10.3**.
 - A digital terrain model.



10.4.1 Simpson’s rule and quadrature

neliöinti A handy method for the numerical integration, or *quadrature*, of earthwork volume from profile data is Simpson’s rule,³ **Kahmen and Faig (1988)** subsection 14.2.1. The equation is

$$V = \frac{1}{6}(\omega_1 + 4\omega_m + \omega_2) \ell, \quad (10.4)$$

in which ω_1 , ω_2 are the surface areas of the end-point cross-sections, ω_m is the surface area of the midpoint cross-section, and ℓ is the length of the whole object.

Simpson’s rule can be proven as follows. Let the function to be integrated be $f(x)$, and let us have at our disposal function values at the points $(-\Delta x, 0, \Delta x)$:

$$f_{-1} = f(-\Delta x), \quad f_0 = f(0), \quad f_1 = f(\Delta x).$$

Approximate the function f by a fourth degree polynomial:

$$\tilde{f}(x) = a + bx + cx^2 + dx^3 + ex^4.$$

³Thomas Simpson (1710–1761) was the son of an English weaver and an autodidact mathematician, fellow of the Royal Society. He actually did not invent Simpson’s rule, although it appeared in his textbook: the rule was already known to Johannes Kepler.



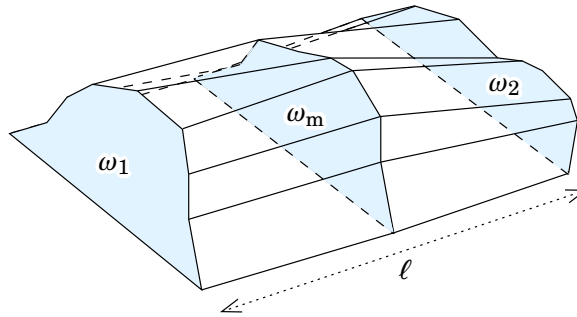


FIGURE 10.7. Simpson's integration rule in volume calculation.

The integral over the polynomial is

$$\begin{aligned} \int_{-\Delta x}^{+\Delta x} \tilde{f}(x) dx &= \left(ax + \frac{1}{2}bx^2 + \frac{1}{3}cx^3 + \frac{1}{4}dx^4 + \frac{1}{5}ex^5 \right) \Big|_{-\Delta x}^{+\Delta x} = \\ &= 2a \Delta x + \frac{2}{3}c \Delta x^3 + \frac{2}{5}e \Delta x^5. \end{aligned} \quad (10.5)$$

Write as well

$$\begin{aligned} \tilde{f}_{-1} &= a - b \Delta x + c \Delta x^2 - d \Delta x^3 + e \Delta x^4, \\ \tilde{f}_0 &= a, \\ \tilde{f}_1 &= a + b \Delta x + c \Delta x^2 + d \Delta x^3 + e \Delta x^4, \end{aligned}$$

so that the linear combination

$$\begin{aligned} I &= p_{-1}\tilde{f}_{-1} + p_0\tilde{f}_0 + p_1\tilde{f}_1 = a(p_{-1} + p_0 + p_1) + \\ &\quad + (b \Delta x + d \Delta x^3)(-p_{-1} + p_1) + (c \Delta x^2 + e \Delta x^4)(p_{-1} + p_1). \end{aligned} \quad (10.6)$$

Comparing equations 10.5 and 10.6 shows that, to get I as close as possible to the integral 10.5, we must choose

$$p_{-1} + p_0 + p_1 = 2 \Delta x, \quad -p_{-1} + p_1 = 0, \quad p_{-1} + p_1 = \frac{2}{3} \Delta x,$$

which yields

$$p_{-1} = p_1 = \frac{1}{3} \Delta x, \quad p_0 = \frac{4}{3} \Delta x.$$

Substituting this into equation 10.6 yields

$$I = 2a \Delta x + \frac{2}{3}c \Delta x^3 + \frac{2}{3}e \Delta x^5, \quad (10.7)$$

and the difference with the integral is

$$\int_{-\Delta x}^{+\Delta x} \tilde{f}(x) dx - I = -\frac{4}{15}e \Delta x^5,$$



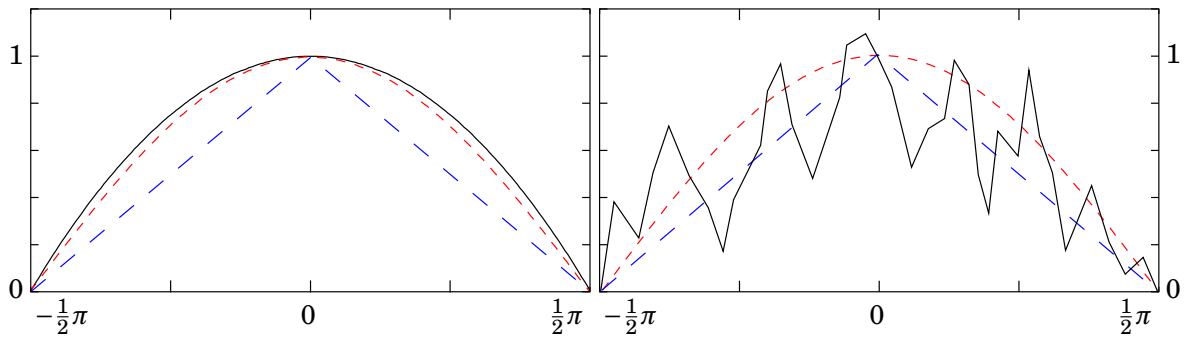


FIGURE 10.8. Alternatives for quadrature. Left, a mathematically well-behaved function ($\cos x$, black), Simpson (red) works best. Right, a “jagged” function. The trapezoid rule (blue) works just as well. Realistic terrain is between these extremes.



a fifth-degree function of the point spacing Δx . This means that by choosing Δx small enough, we can get the expression 10.7 very quickly close to the true value of the polynomial. We may write

$$\begin{aligned} I &= p_{-1}\tilde{f}_{-1} + p_0\tilde{f}_0 + p_1\tilde{f}_1 = \frac{1}{3}\tilde{f}_{-1}\Delta x + \frac{4}{3}\tilde{f}_0\Delta x + \frac{1}{3}\tilde{f}_1\Delta x = \\ &= \frac{1}{6} \left[\tilde{f}_{-1} + 4\tilde{f}_0 + \tilde{f}_1 \right] \cdot 2\Delta x. \end{aligned}$$

Into this we substitute the true values $f_{-1} \stackrel{\text{def}}{=} \omega_1$, $f_0 \stackrel{\text{def}}{=} \omega_m$, $f_1 \stackrel{\text{def}}{=} \omega_2$ of the function, as well as $\ell = 2\Delta x$, and obtain Simpson’s rule 10.4. If the function f to be integrated is not pathological — but terrain forms could well be pathological! — then Simpson’s rule will also converge quickly to it.

Note that Simpson’s rule is exact not only for quadratic, but even for cubic polynomials.



10.4.2 Alternative quadrature rules

neliöintisääntö Often, a simpler quadrature rule

$$V = \frac{1}{2}(\omega_1 + \omega_2)\ell$$

(the “trapezoid rule”) works, or even

$$V = \omega_m\ell$$

(the “rectangle rule”). They do not, however, converge as beautifully as Simpson’s rule: for both, the error is proportional to the point spacing, or object length, cubed, $\Delta x^3 = \ell^3$.



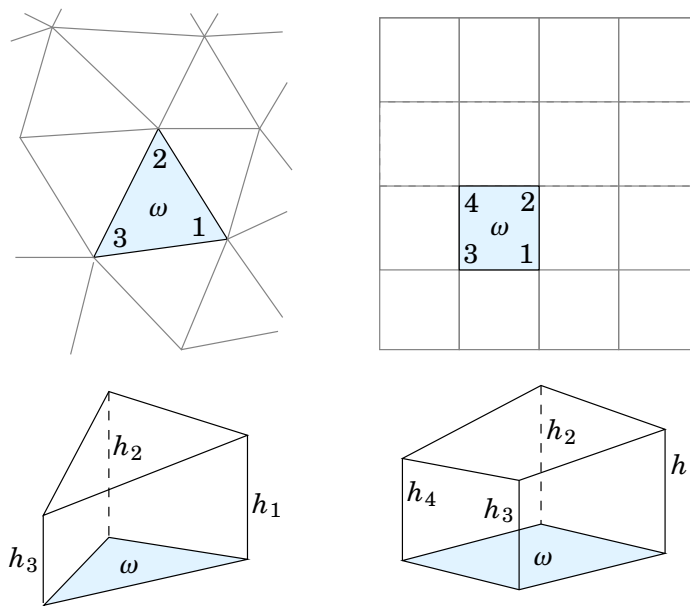


FIGURE 10.9. Volume calculation from digital terrain models.

If the accuracy of Simpson's rule is insufficient, because the object is too jagged or "pathological", one may achieve a better accuracy by dividing the object into slices, applying some simpler quadrature rule to each of those, and summing the contributions obtained.

Volume calculation for *triangulated-network* or *point-grid* type terrain models is depicted in figure 10.9. In the case of the triangulated-network model, the volume of a surface element is evaluated using the equation

$$V = \omega \frac{h_1 + h_2 + h_3}{3}.$$

In the case of a point-grid model, the equation to be used is

$$V = \omega \frac{h_1 + h_2 + h_3 + h_4}{4}.$$

Generalisation: the area of a surface element ω multiplied by the average height calculated from n corner points

$$\bar{h} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n h_i.$$

These equations are approximate but often sufficient.

Self-test questions

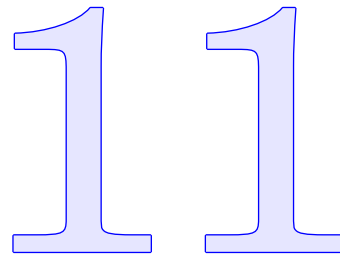
1. Which observation techniques are available for collecting terrain point information useable for the construction of terrain models?



2. Which are the two main techniques for the presentation of terrain models?
3. Discuss applications of terrain models.
4. Explain how the surface area of a parcel may be determined from setting-out measures of its boundary.
5. Explain how a polar planimeter works.
6. Explain how the quadrature of volumes from profile data using Simpson's rule works.
7. What was the mission concept of the Shuttle Radar Topography Mission (SRTM)?



The third dimension



Here as he walked by
on the 16th of October 1843
Sir William Rowan Hamilton
in a flash of genius discovered
the fundamental formula for
quaternion multiplication
 $i^2 = j^2 = k^2 = ijk = -1$
& cut it on a stone of this bridge.

Inscription on Broom Bridge, Dublin



11.1 Geocentric co-ordinate reference systems

In modern geodesy, the measurement methods of satellite and space geodesy are integral parts of the global geodetic observing system (**IUGG**, **GGOS**). Unlike traditional geodetic measurement methods, which carry out their measurements on, or close to, the Earth's surface, these measurements are *genuinely three-dimensional*, and they also require the use of three dimensions in computations involving them. In addition, the platforms of, at least, satellite measurements orbit the Earth, meaning that the centre of mass of the Earth becomes naturally the origin of the co-ordinate frame used. This is why, in satellite geodesy, we use *geocentric, three-dimensional* co-ordinate reference systems. See figure 11.1. Geocentric co-ordinates are often denoted by capital letters, like X , Y , Z .

Geocentric The origin of the system is in the centre of mass of the Earth, and the Z axis is directed along the rotation axis of the Earth.

There are two types of geocentric systems:

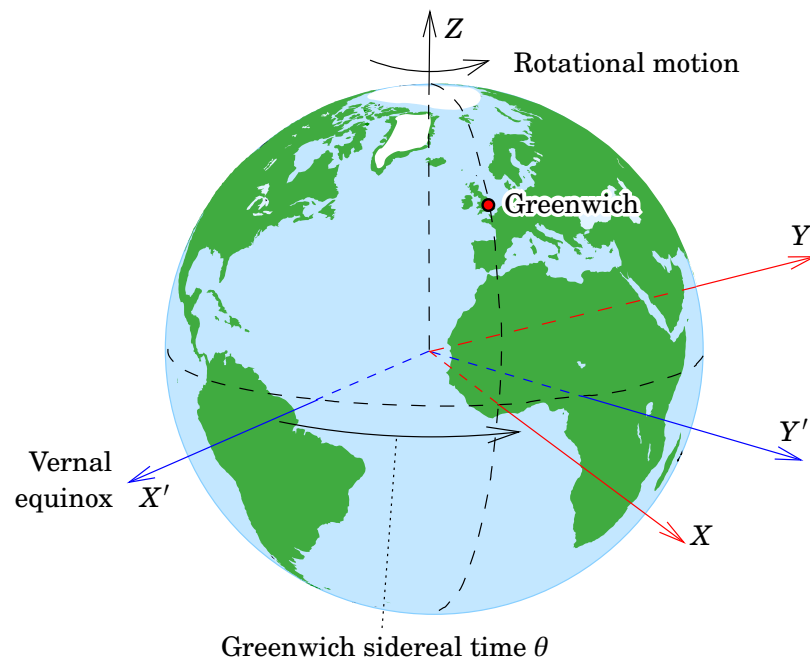


FIGURE 11.1. The inertial or celestial (X', Y', Z) and the terrestrial, co-rotating, or **ECEF** (X, Y, Z) co-ordinate reference systems.

kevät-
päiväntasaus

Inertial or celestial There is no rotational motion. The directions of the axes are fixed with respect to the stars.

The X axis points (usually) to the *vernal* or *spring equinox*, the “Greenwich of the sky”.

Terrestrial or co-rotating Also Earth-centred, Earth-fixed, or **ECEF**. The directions of the axes are fixed with respect to the solid, rotating Earth.

The X axis points in the direction of the *Greenwich* meridian.

kiertokulma
Greenwichin
tähtiäika

Between the inertial and the terrestrial systems there is a rotation angle called *Greenwich sidereal time*. It changes rapidly with time, at the same angular rate as the rotation of the Earth with respect to the stars.

Right-handed co-ordinate frame A corkscrew which progresses in the positive x direction, turns from the y -axis direction to the z -axis direction, figure 11.2.

A geocentric system is right-handed if the X axis points in the direction of the intersection of the planes of the Greenwich meridian and the equator, the Z axis to the North Pole, and the Y axis to 90° east.

Geocentric co-ordinates may be rectangular (X, Y, Z), spherical, geode-



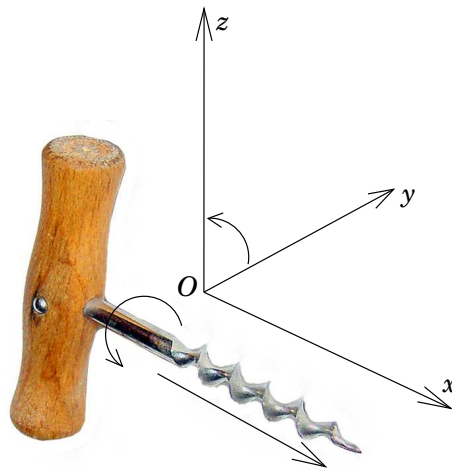


FIGURE 11.2. A right-handed co-ordinate frame. A $y \rightarrow z$ corkscrew (Wikimedia Commons, Corkscrew) progresses in the x direction — as does a $z \rightarrow x$ corkscrew in the y direction, and an $x \rightarrow y$ corkscrew in the z direction.



tic or geographical, or ellipsoidal co-ordinates. The following relations exist between them (figure 11.3):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = r \begin{bmatrix} \cos \phi \cos \lambda \\ \cos \phi \sin \lambda \\ \sin \phi \end{bmatrix},$$

in which (ϕ, λ, r) , the distance r from the geocentre and the *geocentric* latitude and longitude ϕ and λ , are spherical co-ordinates, and

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} (N(\varphi) + h) \cos \varphi \cos \lambda \\ (N(\varphi) + h) \cos \varphi \sin \lambda \\ \left((b^2/a^2) N(\varphi) + h \right) \sin \varphi \end{bmatrix},$$

in which (φ, λ, h) , the height h from the reference ellipsoid and the *geodetic* latitude and longitude φ and λ , are geodetic, also called geographical, co-ordinates. The quantities a and b are the semi-major and semi-minor axes of the Earth ellipsoid; in other words, the equatorial and polar radii. The transversal radius of curvature is

$$N(\varphi) = \frac{a^2}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}}.$$

The third geocentric co-ordinate type, ellipsoidal co-ordinates, are sometimes used in scientific work, but they have no practical significance in land-surveying; see Heiskanen and Moritz (1967) pages 39–45.

iso- ja
pikkuakselin
puolikkaat
poikittais-
kaarevuussäde



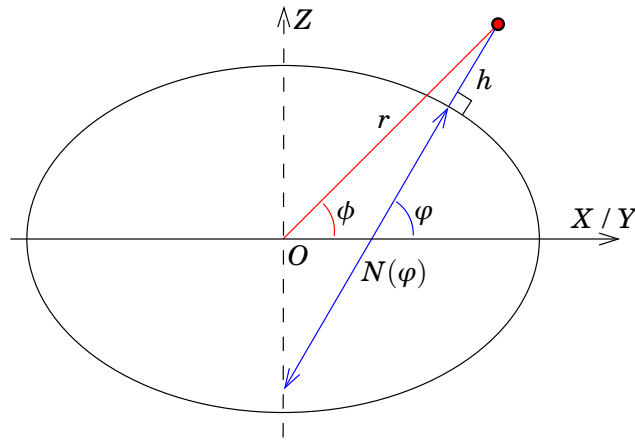


FIGURE 11.3. Geocentric (ϕ) and geodetic (φ) latitude and transversal radius of curvature N .



- 1 The advantage of rectangular¹ co-ordinates is that calculations are easy with them. For example, the distance s between two points (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) is simply²

$$s_{12} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2}.$$

If the points are given in the form $(\varphi_1, \lambda_1, h_1)$, $(\varphi_2, \lambda_2, h_2)$, the corresponding equation will not be quite as simple!



11.2 Topocentric co-ordinates

In practical measurement work, often local or *topocentric*, three-dimensional co-ordinates³ are used, in which the origin is the location of measurement itself, the instrument (*instrument co-ordinates*). It is natural to use spherical co-ordinates (α, ζ, s) , in which s is the slant range from the instrument, α is the azimuth or horizontal direction angle, and ζ is the zenith angle. From these, the rectangular co-ordinates of the signal

¹Rectangular co-ordinates are often called *Cartesian*, after René Descartes. Strictly speaking Cartesian co-ordinates have straight co-ordinate lines, whereas rectangular co-ordinates could also be curvilinear. In fact, both spherical and geodetic co-ordinates are rectangular in this more general sense.

²Of course this is the straight distance in space, often passing through the solid body of the Earth. Usually we are more interested in the distance over the Earth's surface.

³Greek *topos* = place, like *utopia* = non-existent place.



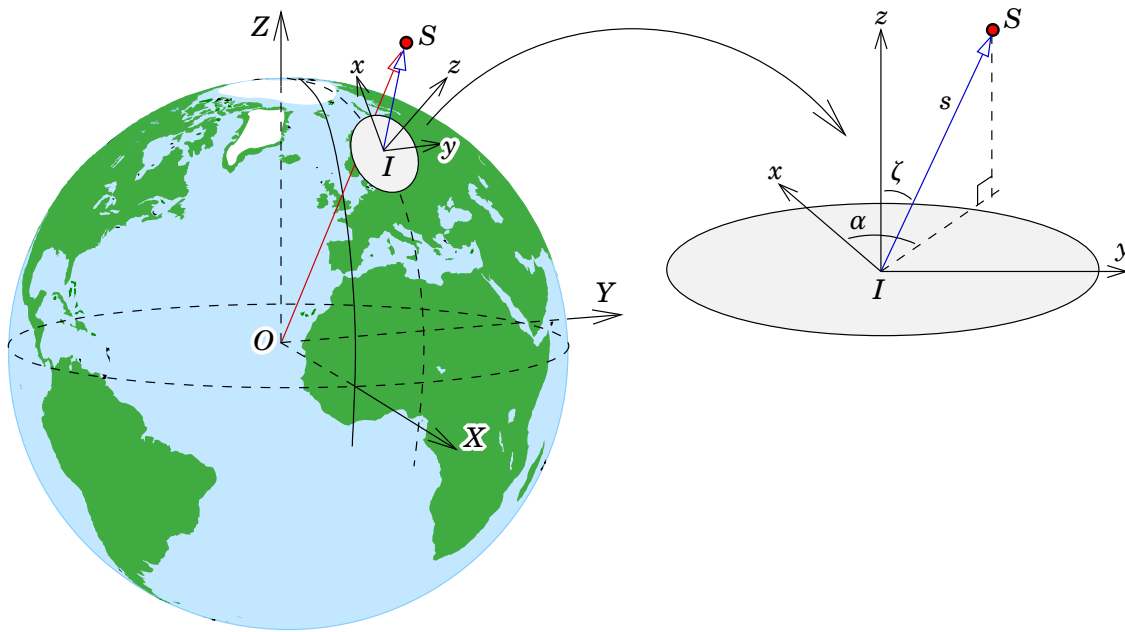


FIGURE 11.4. The topocentric or instrument co-ordinate frame, as well as the geocentric co-ordinate frame. The instrument is I , the geocentre O and the measured location or signal, S . Note that this co-ordinate frame is left-handed.



or target are easily calculated:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = s \begin{bmatrix} \sin \zeta \cos \alpha \\ \sin \zeta \sin \alpha \\ \cos \zeta \end{bmatrix}.$$

Total stations or electronic tacheometers can give precisely these instrument co-ordinates, either in spherical⁴ (α, ζ, s) or rectangular (x, y, z) ⁴ form. Conventionally we write topocentric co-ordinates in *lower case*.

Figure 11.4 depicts both the co-ordinate axes (x, y, z) of the topocentric system, and the axes (X, Y, Z) of the geocentric system. In this figure, the signal S may be a point to be measured on the Earth's surface, but also a satellite orbiting the Earth. In any case, the measurements are always obtained topocentrically first, with respect to the plane of the local horizon (grey circle) of the observation point I .

⁴Strictly speaking only gyrotheodolites can provide the absolute azimuth α . For an ordinary instrument, the unknown azimuth of the zero on the horizontal circle needs to be determined separately, typically in a network adjustment or by an astronomical azimuth determination (section 11.6).



The transformation between these two rectangular three-dimensional systems is a three-dimensional similarity of Helmert transformation, on which more in the following.



11.3 Three-dimensional transformations

A two-dimensional similarity or Helmert transformation is, if the turn from the X axis to the Y axis has the same direction as a positive angle α :

$$\begin{bmatrix} X' \\ Y' \end{bmatrix} = K \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} X - X_0 \\ Y - Y_0 \end{bmatrix},$$

in which α is the rotation angle, K the scale ratio, and $\begin{bmatrix} X_0 & Y_0 \end{bmatrix}^T$ are the co-ordinates of the origin of the new system, written in the old one.

The corresponding three-dimensional transformation equation is obtained by adding the Z axis and keeping its direction fixed:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = K \begin{bmatrix} \cos \alpha_3 & \sin \alpha_3 & 0 \\ -\sin \alpha_3 & \cos \alpha_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix}.$$

The size 3×3 rotation matrix visible in the equation may be called $R_3(\alpha_3)$.

In the same way as around the Z axis, rotations may also take place around the Y or X axis. In that case, we obtain analogously the rotation matrices

$$R_1(\alpha_1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha_1 & \sin \alpha_1 \\ 0 & -\sin \alpha_1 & \cos \alpha_1 \end{bmatrix}, \quad R_2(\alpha_2) = \begin{bmatrix} \cos \alpha_2 & 0 & -\sin \alpha_2 \\ 0 & 1 & 0 \\ \sin \alpha_2 & 0 & \cos \alpha_2 \end{bmatrix}.$$

The general similarity or Helmert transformation containing all three rotations (and three translations, and a scaling) can now be expressed in the following compact form:

$$\bar{\mathbf{R}}' = KR(\bar{\mathbf{R}} - \bar{\mathbf{R}}_0). \quad (11.1)$$

Here we write the vectors as column vectors of their components or co-ordinates of location, as follows:

$$\bar{\mathbf{R}}' = \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}, \quad \bar{\mathbf{R}} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad \bar{\mathbf{R}}_0 = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}.$$



The overbar signals that we are dealing with column vectors of component values, not the vectors in space themselves. It may be left off when the context is clear.⁵

5

Furthermore, the rotation

kierto

$$\mathbf{R} = \mathbf{R}_3(\alpha_3)\mathbf{R}_2(\alpha_2)\mathbf{R}_1(\alpha_1)$$

is the combination, or “chaining”, of three rotations. Equation 11.1 is called the (three-dimensional) Helmert or similarity transformation. The elements of the size 3×3 matrix \mathbf{R} representing the rotation of the co-ordinate frame are complicated trigonometric expressions in the angles α_1 , α_2 , and α_3 , and we do not derive them here.



11.4 Transformation in the case of small rotation angles

Often, the axes of two co-ordinate frames are very close to each other. In that case, the rotation angles are small and one may make the approximation that $\sin \alpha \approx \alpha$ and $\cos \alpha \approx 1$. Then, the equations become simpler. If, in addition, it may be assumed that the scale ratio K is close to unity, one may write

$$K = 1 + m,$$

in which m , the *scale distortion*, is small.

In addition

$$\mathbf{R}_1 \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha_1 \\ 0 & -\alpha_1 & 1 \end{bmatrix}, \quad \mathbf{R}_2 \approx \begin{bmatrix} 1 & 0 & -\alpha_2 \\ 0 & 1 & 0 \\ \alpha_2 & 0 & 1 \end{bmatrix}, \quad \mathbf{R}_3 \approx \begin{bmatrix} 1 & \alpha_3 & 0 \\ -\alpha_3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If all α_i are small, one may furthermore assume that all $\alpha_i \alpha_j \approx 0$.

⁵The vectors themselves are

$$\mathbf{R} = k(X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}) = \mathbf{R}_0 + k'(X'\mathbf{i}' + Y'\mathbf{j}' + Z'\mathbf{k}'), \quad \mathbf{R}_0 = k(X_0\mathbf{i} + Y_0\mathbf{j} + Z_0\mathbf{k}),$$

in which $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ and $\{\mathbf{i}', \mathbf{j}', \mathbf{k}'\}$ are *orthonormal bases* of the space, for the old and the new co-ordinate frame respectively. Also $K = k/k'$, the ratio of the scale deformations of the two frames. For a pure rotation this becomes

$$\mathbf{R} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k} = X'\mathbf{i}' + Y'\mathbf{j}' + Z'\mathbf{k}'.$$



We obtain

$$R = R_3(\alpha_3)R_2(\alpha_2)R_1(\alpha_1) = \begin{bmatrix} 1 & \alpha_3 & -\alpha_2 \\ -\alpha_3 & 1 & \alpha_1 \\ \alpha_2 & -\alpha_1 & 1 \end{bmatrix} = I + \Delta R,$$

in which I is the 3×3 unit matrix, and

$$\Delta R = \begin{bmatrix} 0 & \alpha_3 & -\alpha_2 \\ -\alpha_3 & 0 & \alpha_1 \\ \alpha_2 & -\alpha_1 & 0 \end{bmatrix}$$

is a skew-symmetric (antisymmetric) matrix: $\Delta R^T = -\Delta R$.

Now we obtain from equation 11.1:

$$\begin{aligned} \bar{\mathbf{R}}' &= (1+m)(I + \Delta R)(\bar{\mathbf{R}} - \bar{\mathbf{R}}_0) \approx (I + mI + \Delta R)(\bar{\mathbf{R}} - \bar{\mathbf{R}}_0) \\ &\implies \bar{\mathbf{R}}' - \bar{\mathbf{R}} = (mI + \Delta R)\bar{\mathbf{R}} - (I + mI + \Delta R)\bar{\mathbf{R}}_0. \end{aligned}$$

siirto If we now also assume the translations $\bar{\mathbf{R}}_0$ to be small, then from this follows the *co-ordinate correction equation*

$$\bar{\mathbf{R}}' - \bar{\mathbf{R}} \approx (mI + \Delta R)\bar{\mathbf{R}} - \bar{\mathbf{R}}_0 = \begin{bmatrix} m & \alpha_3 & -\alpha_2 \\ -\alpha_3 & m & \alpha_1 \\ \alpha_2 & -\alpha_1 & m \end{bmatrix} \bar{\mathbf{R}} - \bar{\mathbf{R}}_0, \quad (11.2)$$

in which m , α_1 , α_2 , α_3 , $\bar{\mathbf{R}}_0$, and $\bar{\mathbf{R}}' - \bar{\mathbf{R}}$ are all small (but $\bar{\mathbf{R}}$ and $\bar{\mathbf{R}}'$ are large).

The form 11.2 is the “small” form of the general Helmert transformation, between two realisations of co-ordinate reference systems that are close together, like, for example, the different realisations of **ITRS**, the *International Terrestrial Reference System*. In that case, the rotation angles α_i are of the order of a fraction of a second of arc, and the translation vector $\bar{\mathbf{R}}_0$ is under 10 cm.



11.5 The transformation between two reference ellipsoids

A classical case is the transformation between two geodetic datums which are defined on two differently non-geocentric reference ellipsoids, for example in Europe between the Hayford ellipsoid of the **ED50** datum and Eastern Europe’s Krasovsky ellipsoid. A brute-force method is then to first convert geodetic co-ordinates (φ, λ, h) to rectangular (X, Y, Z) , carry out a three-dimensional Helmert transformation between the two



datums, and convert the result back into geodetic co-ordinates in the other datum, (φ', λ', h') .

If the difference between the two datums is small and consists only of a shift of the reference ellipsoid's centre, there is a simple relationship between the centre shift and the change in geodetic co-ordinates. See figure 11.5.

Let the location vector of a point from the centre of one ellipsoid be \mathbf{R}_1 , and from the other, \mathbf{R}_2 , and the difference, in rectangular geocentric co-ordinates,⁶

$$\Delta \bar{\mathbf{R}} = \bar{\mathbf{R}}_2 - \bar{\mathbf{R}}_1 = \begin{bmatrix} X_2 - X_1 \\ Y_2 - Y_1 \\ Z_2 - Z_1 \end{bmatrix} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix}.$$

Around the point, we define local topocentric co-ordinates (x, y, z) on a unit-vector, or orthonormal, basis $\{\mathbf{N}, \mathbf{E}, \mathbf{U}\}$ ("north, east, up"). At the location of the point, the principal radii of curvature of the reference ellipsoid are $M(\varphi)$, the meridional radius of curvature, and $N(\varphi)$, the transversal radius of curvature. Now the topocentric co-ordinate differences are

$$\Delta x = M(\varphi) \Delta \varphi, \quad \Delta y = N(\varphi) \cos \varphi \Delta \lambda, \quad \Delta z = \Delta h,$$

and

$$\Delta \mathbf{R} = \mathbf{N} \Delta x + \mathbf{E} \Delta y + \mathbf{U} \Delta z = \mathbf{N} M(\varphi) \Delta \varphi + \mathbf{E} N(\varphi) \cos \varphi \Delta \lambda + \mathbf{U} \Delta h.$$

In matrix notation, this is

$$\begin{aligned} \Delta \bar{\mathbf{R}} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} &= \begin{bmatrix} \bar{\mathbf{N}} & \bar{\mathbf{E}} & \bar{\mathbf{U}} \end{bmatrix} \begin{bmatrix} M(\varphi) \Delta \varphi \\ N(\varphi) \cos \varphi \Delta \lambda \\ \Delta h \end{bmatrix} = \\ &= \begin{bmatrix} N_X & E_X & U_X \\ N_Y & E_Y & U_Y \\ N_Z & E_Z & U_Z \end{bmatrix} \begin{bmatrix} M(\varphi) \Delta \varphi \\ N(\varphi) \cos \varphi \Delta \lambda \\ \Delta h \end{bmatrix}, \end{aligned}$$

in which the matrix is orthogonal, in fact a rotation matrix:

$$R = \begin{bmatrix} N_X & E_X & U_X \\ N_Y & E_Y & U_Y \\ N_Z & E_Z & U_Z \end{bmatrix} = \begin{bmatrix} -\sin \varphi \cos \lambda & -\sin \lambda & \cos \varphi \cos \lambda \\ -\sin \varphi \sin \lambda & \cos \lambda & \cos \varphi \sin \lambda \\ \cos \varphi & 0 & \sin \varphi \end{bmatrix}.$$

ortonormaali
kanta



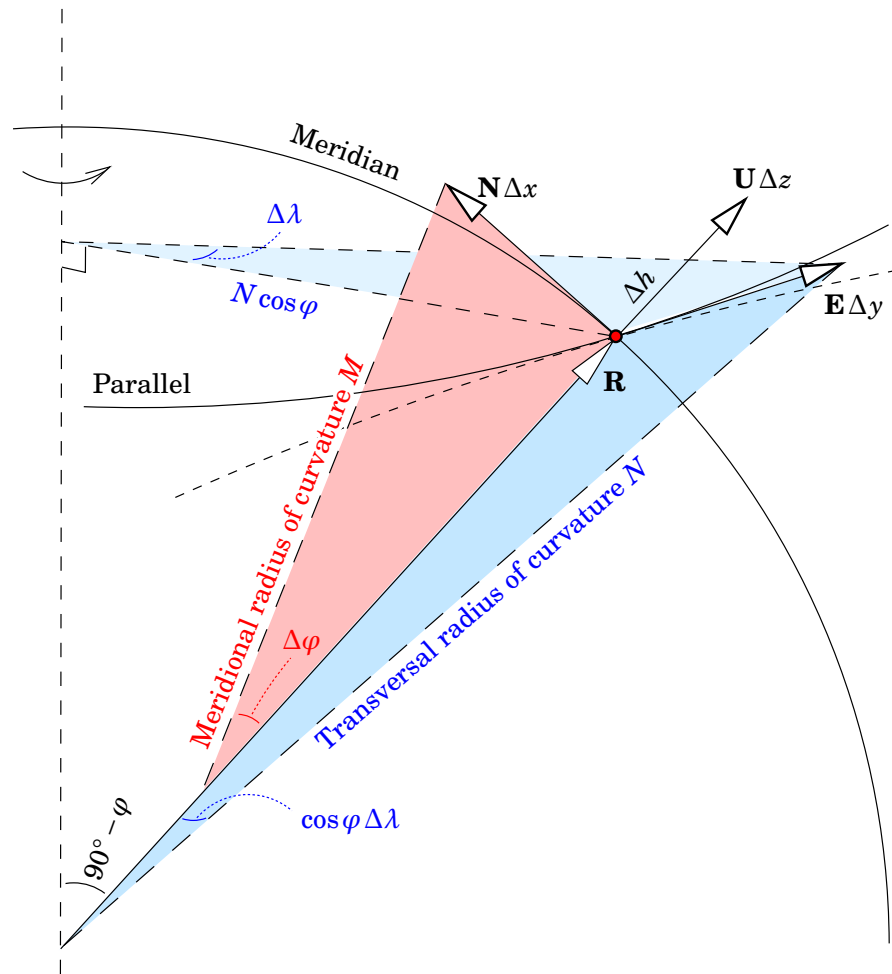


FIGURE 11.5. The differential connection between rectangular ($\mathbf{N}, \mathbf{E}, \mathbf{U}$) co-ordinates and geodetic co-ordinates on the reference ellipsoid.

The columns of the matrix are the geocentric components of the vectors \mathbf{N} , \mathbf{E} , and \mathbf{U} .

Inverting an orthogonal matrix is easy: $R^{-1} = R^T$, or

$$\begin{bmatrix} M(\varphi) \Delta\varphi \\ N(\varphi) \cos\varphi \Delta\lambda \\ \Delta h \end{bmatrix} = \begin{bmatrix} -\sin\varphi \cos\lambda & -\sin\varphi \sin\lambda & \cos\varphi \\ -\sin\lambda & \cos\lambda & 0 \\ \cos\varphi \cos\lambda & \cos\varphi \sin\lambda & \sin\varphi \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix}.$$

Thus we may easily calculate what are the effects of shifting the centre of the reference ellipsoid on geodetic co-ordinates (φ, λ, h) , evaluated on the ellipsoid:

$$\begin{bmatrix} \varphi_2 \\ \lambda_2 \\ h_2 \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \lambda_1 \\ h_1 \end{bmatrix} + \begin{bmatrix} (M(\varphi))^{-1} & 0 & 0 \\ 0 & (N(\varphi) \cos\varphi)^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix} R^T \begin{bmatrix} X_2 - X_1 \\ Y_2 - Y_1 \\ Z_2 - Z_1 \end{bmatrix}.$$

⁶This is also the vector connecting the centres of the two reference ellipsoids!



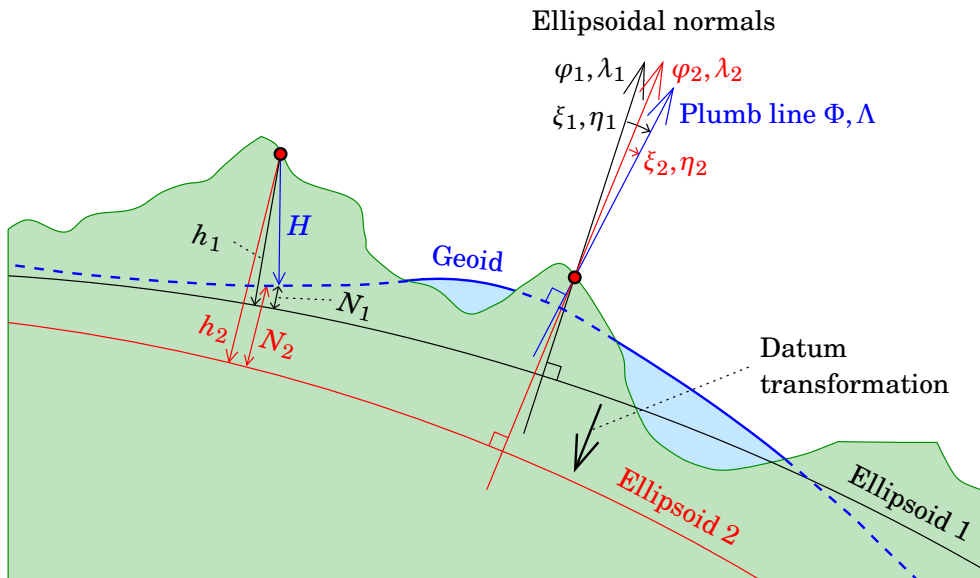


FIGURE 11.6. The effect of a datum transformation (shift of the centre of the reference ellipsoid) on geodetic latitude and longitude φ and λ , deviations of the plumb line ξ and η , geoid heights N , and the heights h of points from the reference ellipsoid.



And when the geodetic co-ordinates (φ, λ, h) change, then the deviations of the plumb line and the heights of the geoid also change, their definitions being

luotiviivan
poikkeama

$$\xi = \Phi - \varphi, \quad \eta = (\Lambda - \lambda) \cos \varphi, \quad N = h - H, \quad (11.3)$$

in which (Φ, Λ) are astronomically determined latitude and longitude, (ξ, η) are deviations of the plumb line in the north and east directions, h is the height above the ellipsoid, and H the height above sea level, that is, the geoid, while N is the geoid height reckoned from the reference ellipsoid. From this is obtained directly

$$\begin{bmatrix} -M(\varphi) \Delta \xi \\ -N(\varphi) \Delta \eta \\ \Delta N \end{bmatrix} = \begin{bmatrix} -\sin \varphi \cos \lambda & -\sin \varphi \sin \lambda & \cos \varphi \\ -\sin \lambda & \cos \lambda & 0 \\ \cos \varphi \cos \lambda & \cos \varphi \sin \lambda & \sin \varphi \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix},$$

because Φ , Λ , and H can be calculated directly from measurements, without any use of a reference ellipsoid. See figure 11.6.



11.6 Laplace azimuth measurements

A geodetic network computed on the reference ellipsoid is oriented by astronomical observations. Two of the three orientation degrees of free-



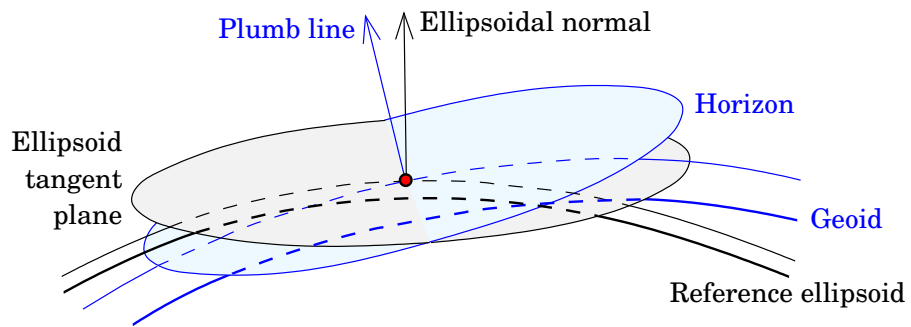


FIGURE 11.7. The deviation of the local plumb line from the normal on the reference ellipsoid surface. The general situation in a point neither on the reference ellipsoid nor on the geoid.



dom are fixed by making the direction of the ellipsoidal normal in the datum point or points equal to the astronomically determined direction of the plumb line. That leaves a third degree of freedom, the network's orientation with respect to the local north.

luotiviiva

The directions measured in the network are projected onto the plane of the local horizon. If one ignores the local deviations of the plumb line — meaning that one assumes that the local plane of the horizon is parallel to the local tangent plane to the reference ellipsoid — one may say that the sighting azimuths are projected onto the reference ellipsoid. This assumption is, however, *not correct*. The local horizon is perpendicular to local gravity, the direction of which differs a little from that of the normal to the reference ellipsoid. The phenomenon is precisely the *plumb-line deviation*, see figure 11.7.

The deviations of the plumb line are ξ in the south-north direction, and η in the west-east direction. Their equations 11.3 were already given:

$$\xi = \Phi - \varphi, \quad \eta = (\Lambda - \lambda) \cos \Phi,$$

with (Φ, Λ) astronomical latitude and longitude. φ and λ , geographical latitude and longitude, are *geodetically computed* co-ordinates such as can be found on maps, computed with respect to a certain reference ellipsoid.⁷

Let the *astronomical azimuth* (absolute direction) measured with respect to the plane of the local horizon be A , and the same sighting mea-

⁷So this means that the deviations of the plumb line will depend on the reference ellipsoid chosen. The choice of the local reference ellipsoid or *datum* was often made so that the sum of the squares of plumb-line deviations was minimised over the area of interest. In other words, so the ellipsoid would fit as well as possible to a level surface of the very local gravity field, the *geoid*.



sured with respect to the reference ellipsoid, that is the *geodetic azimuth*, α .

Then we may say that

$$\begin{aligned} A - \alpha &= \eta \tan \Phi + (\zeta \sin \alpha - \eta \cos \alpha) \cot \zeta = \\ &= (\Lambda - \lambda) \sin \Phi + ((\Phi - \phi) \sin \alpha - (\Lambda - \lambda) \cos \alpha \cos \Phi) \cot \zeta, \quad (11.4) \end{aligned}$$

in which ζ is the zenith angle. Equation 11.4 is called the *Laplace azimuth equation*. If the sighting direction is in the horizontal plane, then $\cot \zeta = 0$ and the correction above is a constant for the observation site, as the dependence on azimuth α vanishes.

Figure 11.8 explains where both terms come from:

- The first term $\eta \tan \Phi$ comes from the projection of the direction of the celestial pole onto the local horizon being different from the projection onto the tangent plane to the reference ellipsoid. It depends on the height of the celestial pole, or astronomical latitude Φ .
- The second term $(\zeta \sin \alpha - \eta \cos \alpha) \cot \zeta$ comes from the difference in projections of the signal direction onto the local horizon and onto the tangent plane of the ellipsoid. It depends on the zenith angle ζ of the signal, and vanishes if $\zeta = 100^\circ$.

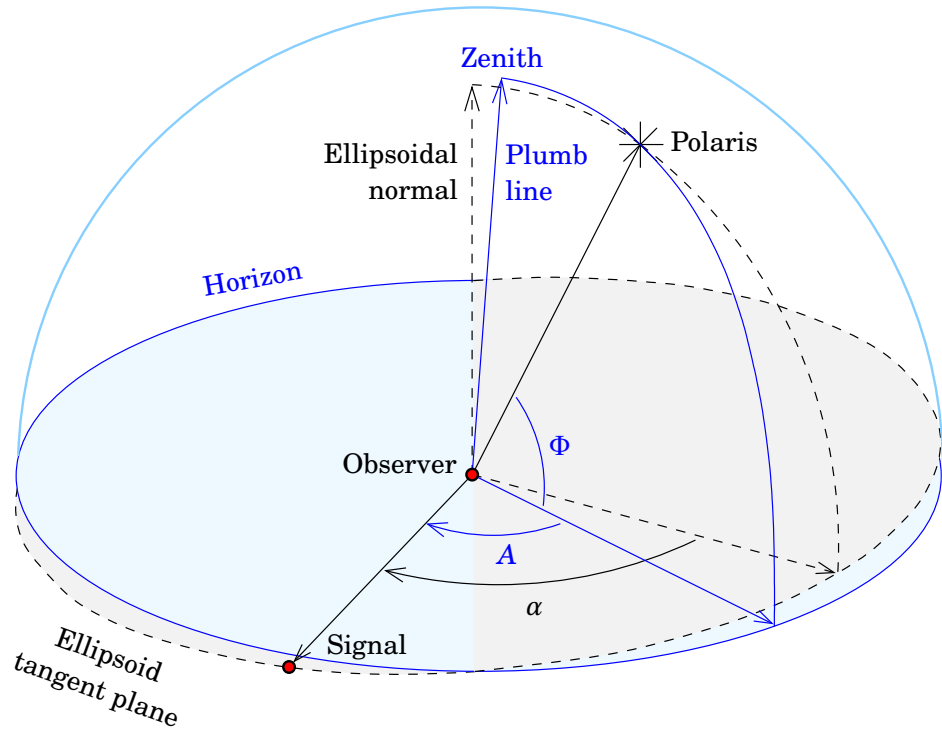
11.7 Traditional “2D+1D” co-ordinates

Co-ordinate frames in which horizontal location and height are given separately have long been in widespread use. An example of this is the **KKJ** system, the Map Grid Co-ordinate System, which was in use in Finland but is now obsolete, and the height system **N60**. **KKJ** gives horizontal co-ordinates in the Gauss–Krüger projection on the Hayford ellipsoid, also known as the International Ellipsoid of 1924. The co-ordinates are based on the **ED50** (European Datum 1950) system, which was created in 1950 by a joint adjustment of the triangulation networks of all Western European countries. This is a traditional, non-geocentric datum.

yhteistasoitus
kolmiomittaus-
verkko

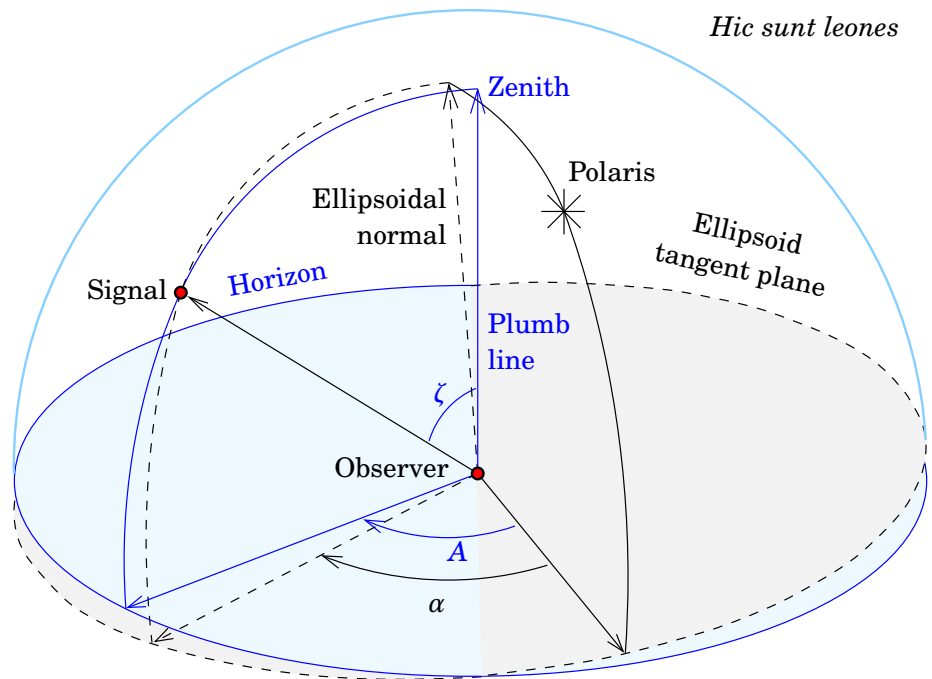
The **N60** system gives orthometric heights, meaning heights from the *geoid*, not the reference ellipsoid. The geoid (see section 16.4) is an undulating reference surface similar to mean sea level. With the International Ellipsoid, one must use the *Bomford*⁸ *geoid model*, which back in the day

8



(a)

Part related to pole elevation Φ



(b)

Part related to signal elevation $100^g - \zeta$

FIGURE 11.8. The Laplace phenomenon: the effect of the plumb-line deviation on the azimuth.



was determined in connection with the ED50 project.


Stating the three-dimensional location of a point in the form (x, y, H) , in which (x, y) is a **KKJ** co-ordinate pair, and H an orthometric **N60** height, is problematic: the connection with the systems used by satellite positioning is complicated. Transforming the co-ordinates (x, y, H) into geocentric (X, Y, Z) co-ordinates involves the following steps:

1. The **KKJ** co-ordinates (x, y) have already undergone a two-dimensional Helmert transformation, aimed at achieving an approximate compatibility with the still older **VVJ** or “Helsinki System”. This transformation is documented in **Ollikainen (1993)**. The inverse of this transformation needs to be applied: $(x, y) \implies (x', y')$.
2. The Gauss–Krüger map projection is applied in the inverse direction, $(x', y') \implies (\varphi, \lambda)$, on the International or Hayford ellipsoid.
3. In order to transform the orthometric height H into a height h from the reference ellipsoid, we need a geoid model that is compatible with the reference ellipsoid used, thus *Bomford’s geoid model*. The geoid height $N = h - H$ is needed in every point.
4. The geodetic co-ordinates (φ, λ, h) are to be transformed into rectangular (X', Y', Z') co-ordinates. These three-dimensional co-ordinates are still in the European Datum 1950 system, which is not (precisely) geocentric.
5. As the **ED50** datum is non-geocentric, we still need a three-dimensional Helmert transformation to arrive at precisely geocentric co-ordinates (X, Y, Z) , see section **11.8** for details.

KKJ has already been replaced by new map projection systems, which are based either on the Gauss–Krüger or the **UTM** (Universal Transverse Mercator) projection on the **GRS80** reference ellipsoid, and by the **EUREF-FIN** datum, the **ETRS89** (European Terrestrial Reference System 1989) system’s Finnish national realisation. The new **N2000** system is used as the height system, and its connection to the ellipsoidal heights is given by the *geoid model FIN2005N00*, see **Bilker-Koivula and Ollikainen (2009)**. The above points **2–4** continue to apply, albeit with other names.

⁸Brigadier Guy Bomford (1899–1996) was a gifted British geodesist and student of geoid determination.



 TABLE 11.1. Transformation parameters from [Ollikainen \(1993\)](#) for [EUREF89](#) → [ED50](#).

Parameter	Value	Precision	Unit
ΔX	93.477	± 3.345	m
ΔY	103.453	± 5.534	m
ΔZ	123.431	± 2.736	m
e_x	-0.246	± 0.168	"
e_y	0.109	± 0.106	"
e_z	0.068	± 0.112	"
m	-2.062	± 0.417	ppm

11.8 Case: the transformation between ED50 and EUREF89

This is a Helmert transformation of type [11.1](#), to an accurately geocentric system. Because the Hayford ellipsoid on which [ED50](#) is based is not (accurately) geocentric, the origin of the co-ordinate frame must be shifted from the centre of the reference ellipsoid to the centre of mass of the Earth. The shifts (translations) are of the order of a hundred metres, and the rotations and the scale change are also substantial, see [Ollikainen \(1993\)](#) page 15 and table 2 on page 13:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{ED50}} = (1 + m) \begin{bmatrix} 1 & e_z & -e_y \\ -e_z & 1 & e_x \\ e_y & -e_x & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{EUREF89}} + \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix},$$

for which table [11.1](#) gives the transformation parameters according to Matti Ollikainen's solution for the territory of Finland.


[ED50](#) (European Datum 1950) is the traditional European datum on which [KKJ](#) is based. It was created well before the satellite age. As can be seen from the table, it is *non-geocentric*. [EUREF89](#) is a modern, [GNSS](#)-based European co-ordinate reference frame. The precision figures given in the table are large because co-ordinates determined in the traditional way over a large area are just not very precise.

More recent information on matters of co-ordinates and transformations for the Finnish territory can be found in [Häkli et al. \(2009\)](#).

11.9 Case: the transformation between ITRF and ETRF

A three-dimensional, satellite-based, thus geocentric co-ordinate reference frame called [EUREF-FIN](#) is in use in Finland. It is the national



 TABLE 11.2. Transformation parameter values ITRF2005 \rightarrow ETRF2005 from Boucher and Altamimi (2007) tables 3 and 4.

Parameter	Value (cm)	Parameter	Value ($10^{-3}''/a$)
T_1	5.6	\dot{R}_1	0.054
T_2	4.8	\dot{R}_2	0.518
T_3	-3.7	\dot{R}_3	-0.781

realisation of ETRS89, the European Terrestrial Reference System, on the Finnish territory.

All geodetic satellite measurements, however, give a position solution in the same frame as that in which the orbital elements of the GPS satellites are given, like ITRF2005. Then, the following transformation is needed to the corresponding ETRS89 realisation, ETRF2005:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{ETRF2005}}(t) = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{ITRF2005}}(t) + \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}_{\text{ITRF2005}}^{\text{ETRF2005}} + \begin{bmatrix} 0 & -\dot{R}_3 & \dot{R}_2 \\ \dot{R}_3 & 0 & -\dot{R}_1 \\ -\dot{R}_2 & \dot{R}_1 & 0 \end{bmatrix}_{\text{ITRF2005}}^{\text{ETRF2005}} \times (t - 1989.0) \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{ITRF2005}}(t),$$

in which the dot on the R parameters (Newton's dot notation) indicates derivation with respect to time. The \dot{R} parameters in this equation include the tectonic motion of the Eurasian plate.

The parameter values for the equation are found in the instructions written by the EUREF subcommission, and are presented in table 11.2.

As can be seen, the transformation parameters in this case are many orders of magnitude smaller than in the earlier-described transformation between EUREF-FIN and ED50. Both co-ordinate reference frames, ETRF2005 and ITRF2005, are geocentric on the centimetre level.

Self-test questions

1. Which are the two main types of geocentric co-ordinate reference systems?
2. What is sidereal time and what does it describe?
3. How many transformation parameters does a three-dimensional Helmert transformation have?

tähtiaika



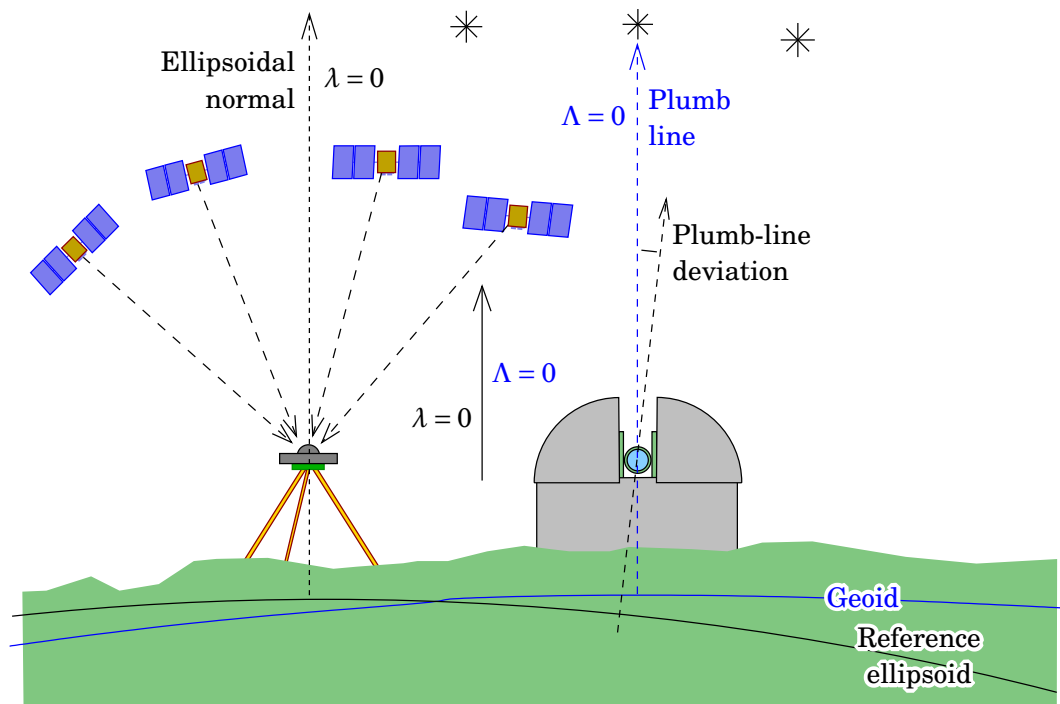


 FIGURE 11.9. Greenwich geometry: longitude is a *direction*, not a *place*.

4. Name the parameters of the three-dimensional Helmert transformation.
5. Describe the Laplace azimuth measurement and what it is used for.

Exercise 11–1: Greenwich: explain this

People have been taking their inexpensive hand-held GNSS receivers — mobile phones, even — to the Greenwich or zero meridian, figure 2.11, finding that *they do not show zero longitude*.

How is that possible?

There is a lot of good explanation for this on the Internet, including one *Journal of Geodesy* article from 2015 — and a lot of plain old tabloid nonsense. Do not buy the nonsense. In this exercise, provide an explanation in your own words, showing that you “get” it.

Imagine that you’re called up by a journalist who has heard about this, and wants your take on it as a geodesist. Your explanation — an elevator speech — should be so lucid that she “gets” it, and what’s more, that when she gets home and tells her husband what she has learned, he “gets” it too...





The Global Positioning System (GPS)

12

Wouldn't it be noteworthy that Transit 5B-5 (launched 1964) is the oldest active satellite in space? It never went into operation because the data up/download link system failed after deployment into LEO. But since that day it transmits its doppler carrier and marker signals on 136.650 MHz, of course the batteries failed many years ago and so the electronics only function when the satellite is in sunlight. But it still manages to boot up most days, after more than 50 years of operation, and can be received using simple handheld scanners. It's considered an admirable "space monument" by sat enthusiasts.

176.0.30.145, November 29, 2015. [Wikipedia, Transit 5B-5 — oldest active satellite](#)

IN LAND-SURVEYING, THE ROLE of the [GPS](#) has during the last two or three decades grown to be dominant, both in Finland and worldwide. The professional literature, especially in the English language, is extensive. [Poutanen \(2017\)](#) is a significant work in Finnish. A good basic work in English is [Hofmann-Wellenhof et al. \(2001\)](#).

Here we will concentrate on the [GPS](#), which has been fully operational for a long time. In parallel to this system, which is operated by the US military authorities, similar systems by other countries have in recent years appeared on the scene. The Russian [GLONASS](#) deserves a special mention: after a time of decay it has now grown back to operability. Satellites from both systems are today routinely used together in land-surveying.

The Europeans have developed their own Galileo system, and the Chinese their BeiDou or "Compass" system. Satellites of both systems

are functioning in orbit and both systems are close to full operationality.

These systems are, in the aggregate, called “GNSS” or *Global Navigation Satellite Systems*. However, the relative simplicity of the way the original GPS works makes it a suitable model for teaching the basics, and therefore we shall concentrate on it.

The GPS is originally a *navigation system*. It was not the first radio navigation system using satellites: an earlier satellite navigation system was the Transit system, or NNSS (Navy Navigation Satellite System), unofficially the “Doppler positioning system”. This system, which was operational 1964–1996, consisted of five satellites orbiting the Earth in low orbits. Geodetically useful positioning required several satellite passages overhead, amounting in practice to at least 24 hours of observations.

See figure 12.3 for a visual illustration of the NNSS.

The GPS satellites are in much higher orbits, and anywhere on Earth, at almost any time, at least six of them are “visible”. More commonly, ¹ the number of visible satellites ranges from eight to over ten.¹ Therefore GPS positioning can be done almost instantly, within a few seconds or minutes.



12.1 Radio navigation and hyperbolic systems

Of the older terrestrial radio navigation systems for seafarers may be mentioned the Decca system, closed down in 2000, which is an example of a *hyperbolic system*. Other systems worth mentioning are Loran-C and ² Omega, both no longer in use.²

kantaaalto Decca transmitted non-modulated carrier waves in the frequency band 70–129 kHz, corresponding to wavelengths 2.3–4.3 km. Its use required at least three transmitters, one “master” and at least two “slave” or auxiliary stations. The transmissions of the stations were precisely synchronized, though they were all transmitting on different wavelengths.

The term “hyperbolic system” is based on the circumstance that a vessel without a clock synchronised with the transmitters can observe only the

¹Positioning instruments making use of more than one satellite system, like GPS and GLONASS, “see” even more satellites, even in poor-visibility locations like the centres of big cities.

²Some countries, including the US, however, are considering bringing Loran-C back to life under the name “eLoran” (“Enhanced Loran”), because, using high-power long-wavelength radio transmitters, it is less susceptible to interference compared to GPS.



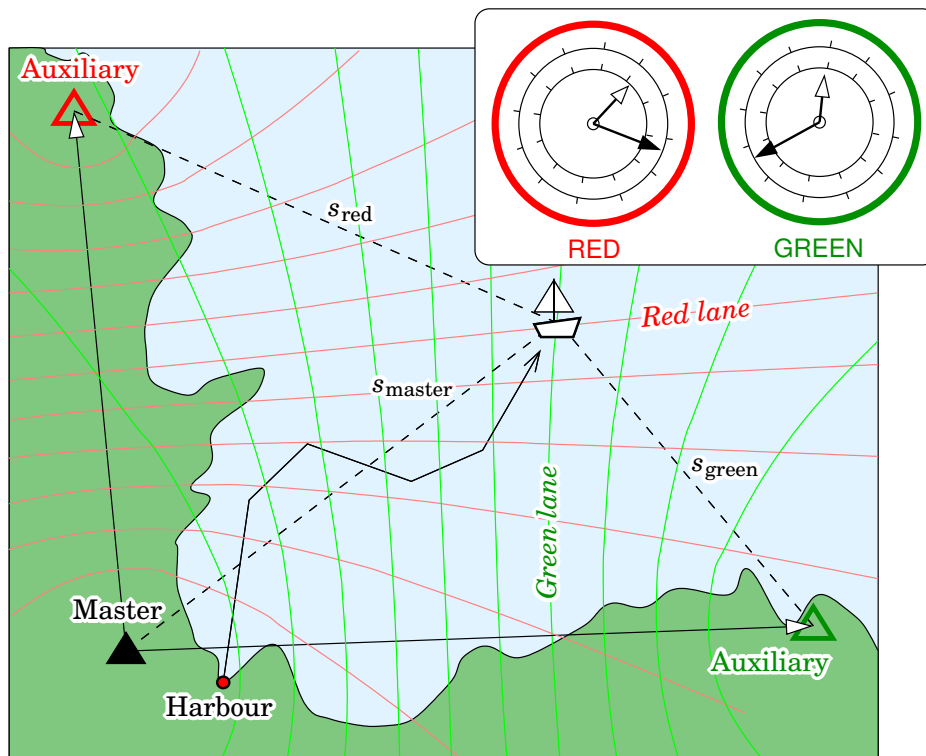


FIGURE 12.1. The Decca system.

difference in reception times between the waves from two transmitters. See figure 12.1.

On the map are drawn, in two different colours, *hyperbolas*, curves of which the points have a *difference* in distance from the master and a certain auxiliary that is constant. For example, for the red hyperbolas we



FIGURE 12.2. A Decca receiver. [Wikimedia Commons, Decca Navigator Mk 12.](#)



have

$$s_{\text{red}} - s_{\text{master}} = \text{constant}$$

and for the green hyperbolas

$$s_{\text{green}} - s_{\text{master}} = \text{constant.}$$

Every curve has its own constant. Unfortunately *this constant cannot be observed*, because all carrier waves look the same.³

kujalaskuri Therefore, an *incremental* measurement method is used. Each *lane counter* of the **Decca** instrument has two hands, one for counting the lanes, the other for the fraction within a lane. The hands move together, just like the hands of a clock.

vaihe-ero The ship must set the lane counters to the correct starting values at a known location, such as the port of departure. After that, during cruising, they follow the development over time of the lane values: every lane hand on the phase-difference dial tracks how many full turns the phase hand has made — see figures 12.1, 12.2. It is a precondition that the *radio connection with the base stations stays uninterrupted*.⁴

tukiasema At any point in time, one may use the two⁵ lane numbers and residual phase differences — fractional numbers $\in [0, 2\pi)$ — to read one's own location on a sea chart, on which the hyperbolas are pre-drawn.

Like the **GPS**, the Navy Navigation Satellite System **NNSS** was also a hyperbolic system: the hyperbolas were formed with two successive satellite locations in space, for times t_0 and $t_0 + 2 \text{ min}$, as the focal points. The difference between the distances to the two locations is obtained by integrating the Doppler shift of the received frequency over this time span, yielding the difference in cycles. The equation is

$$s(t_0) - s(t_0 + 2 \text{ min}) = \frac{c}{f} \int_{t_0}^{t_0 + 2 \text{ min}} (f' - f) dt,$$

in which f is the nominal frequency of the transmission, f' the received frequency, and c the speed of light. This requires the nominal transmission frequency to be accurate: the **NNSS** system was also the first satellite system that served the dissemination of precise time.

pyörähdyshyperboloidi Every hyperboloid of revolution in space intersects the Earth's surface

³In connection with the **GPS** this is called the *ambiguity problem*.

⁴This is referred to as a *kinematic method*. Absolute location is not measured: instead the user follows continuously how they are moving in relation to a known starting point.

⁵Actually **Decca** used *three* colours: *red, green* and *purple*.



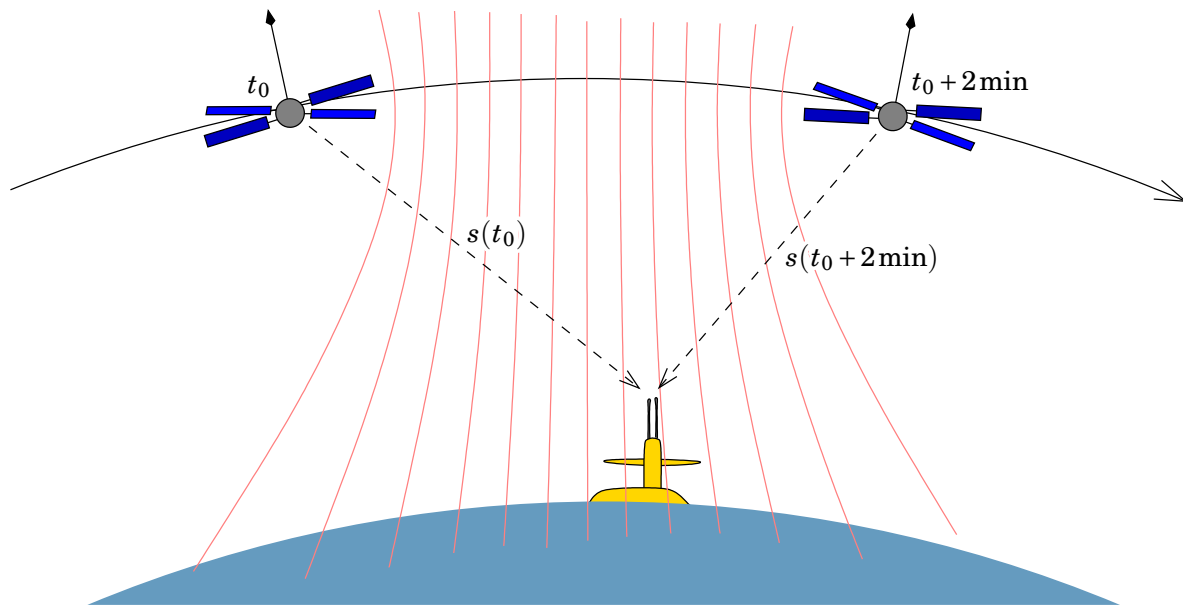


FIGURE 12.3. The NNSS Transit system.

in a curve. The orbits of the Transit satellites are polar, meaning that these curves are running in the general west-east direction. A fix could be obtained already from a single overpass, and using multiple, both north- and southgoing, overpasses in a least-squares adjustment⁶ improved the precision of the determination of both latitude and longitude.

pienimmän
neliösumman
tasoitus
6

12.2 The GPS satellite

Characteristic of the GPS, being originally a military positioning system, is that the *satellites* are *active* and the *users* *passive*. So, the positioning instruments used by the users, the GPS receivers, are quiet, whereas the satellites contain radio transmitters. A GPS satellite is in a way a flying Decca base station.

A GPS satellite is big, a contraption appearing as in figure 12.4a. It contains, among others, the following components:

- A precise *atomic clock*, either a caesium, rubidium, or hydrogen-maser clock. This clock synchronises all signals (carrier waves and modulations) which the satellite transmits. Because the clocks have

⁶At the time of deployment of the NNSS, the computing capacity required was still hard to come by, especially in a form factor suitable for submarine use. A special-purpose computer was designed for this, weighing a quarter of a ton, but small enough — and with rounded corners! — to go through the hatch of a nuclear submarine, [Wikipedia](#), [The AN/UYSK-1 \(TRW-130\)](#)!

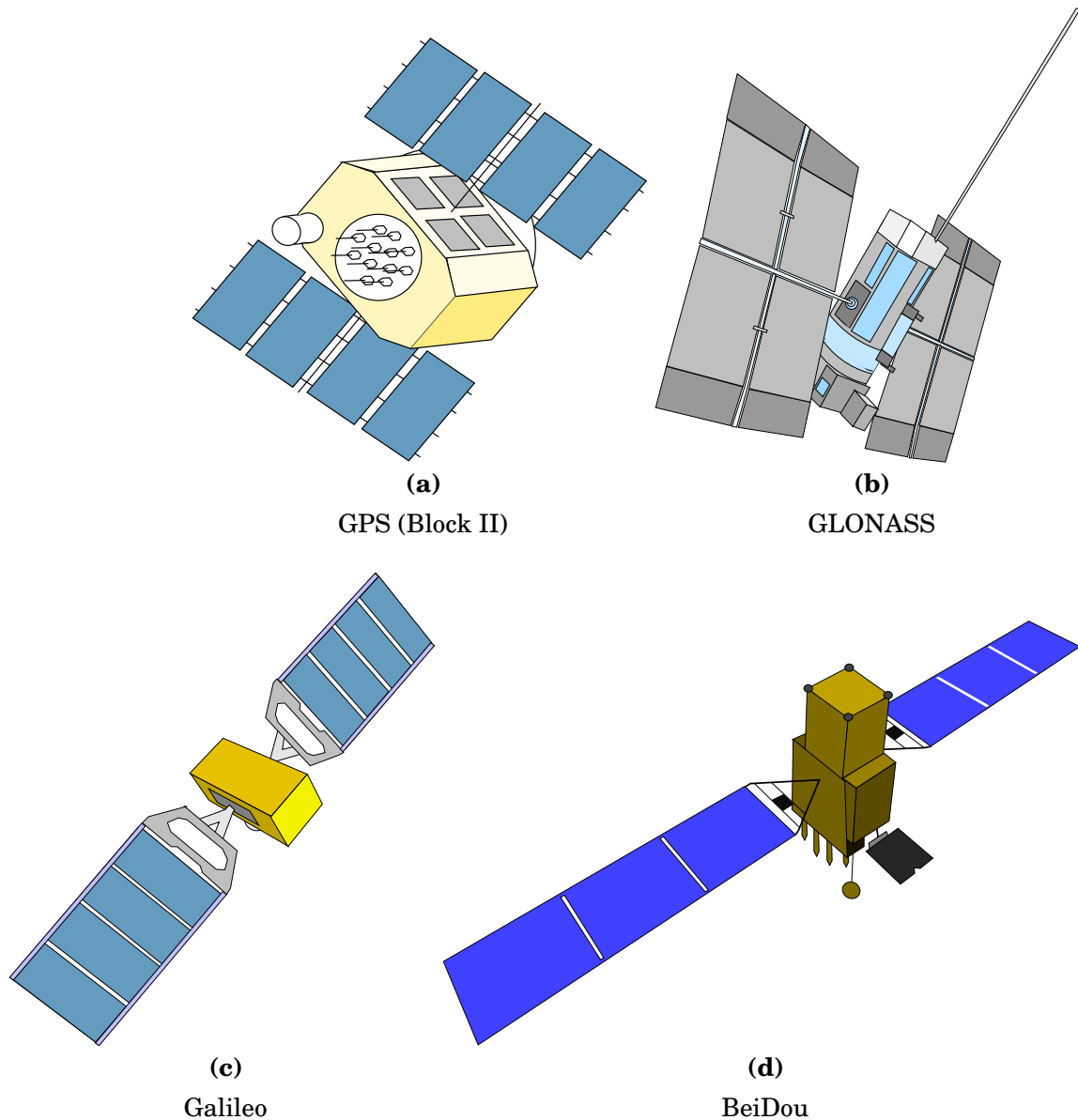


FIGURE 12.4. Positioning satellites.

a limited lifespan, each satellite includes spare clocks.

- *Radio transmitters.* The antennas of the satellite point all the time to the Earth. Transmission power is significant only within a directional cone that contains the Earth as a whole; the total power is about 50 W. Two carrier-wave frequencies are used, 1575.42 MHz (L_1) and 1227.60 MHz (L_2), which enables the elimination of the effect of the ionosphere. The various codes that are used in positioning, as well as a code containing orbit and other information for the users, are modulated onto the carrier waves.



- *Communication channels.* The satellite receives data from the GPS control centre. Besides control commands, this data contains information on the orbits, clock corrections, “health”, etc., of all GPS satellites. The information is stored in the satellite’s memory and transmitted forward to users as a message modulated onto the carrier wave.
- *Solar panels* produce the power needed by the equipment. The amount of power has grown from 400 W (Block I) to almost three kilowatts (Block IIF). At times, the satellite moves through the Earth’s shadow; for this, there are batteries.
- Small rocket engines or *thrusters* for controlling attitude and orbit, as well as a *stock of propellant* (monopropellant hydrazine). Because of orbit perturbations, orbit corrections are needed at regular intervals.
- The satellites are stabilised on three axes: the antennas are pointing to the Earth, the solar panels to the Sun. For stabilisation, *reaction wheels* (“flywheels”) are used.

As, during the lifetime of the GPS, the field of electronics has seen huge developments, there are several satellite generations: “Block I”, “Block II/IIA”, “Block IIR”, “Block IIR-M” and “Block IIF”, see [Misra and Enge \(2010\)](#). The first Block-IIIA satellite was launched in 2018.

Satellites in operation today are all Block II or higher. The masses of the satellites are 845 kg (Block I), 1660 kg (Block II) and 2269 kg (Block III). The design lifetime of the satellites — limited by the propellant stock for orbit maintenance, the diminishing power of the solar cells and batteries, and the development of defects in clocks and electronics in the Earth’s outer radiation belt — is 4.5 (Block I), 7.5 (Block II) or today even 12–15 years. The satellites have regularly exceeded their design lifetimes.



12.3 The GPS segments

The GPS consists of three *segments*, figure 12.5:

Space segment The satellites themselves.

[avaruuslohko](#)

Control segment Command centre, telemetry and control, tracking stations, orbit determination, time synchronisation.

[valvontalohko](#)

The main command centre, or Master Control Station, of the GPS is located in Colorado Springs. There are some two dozen orbit



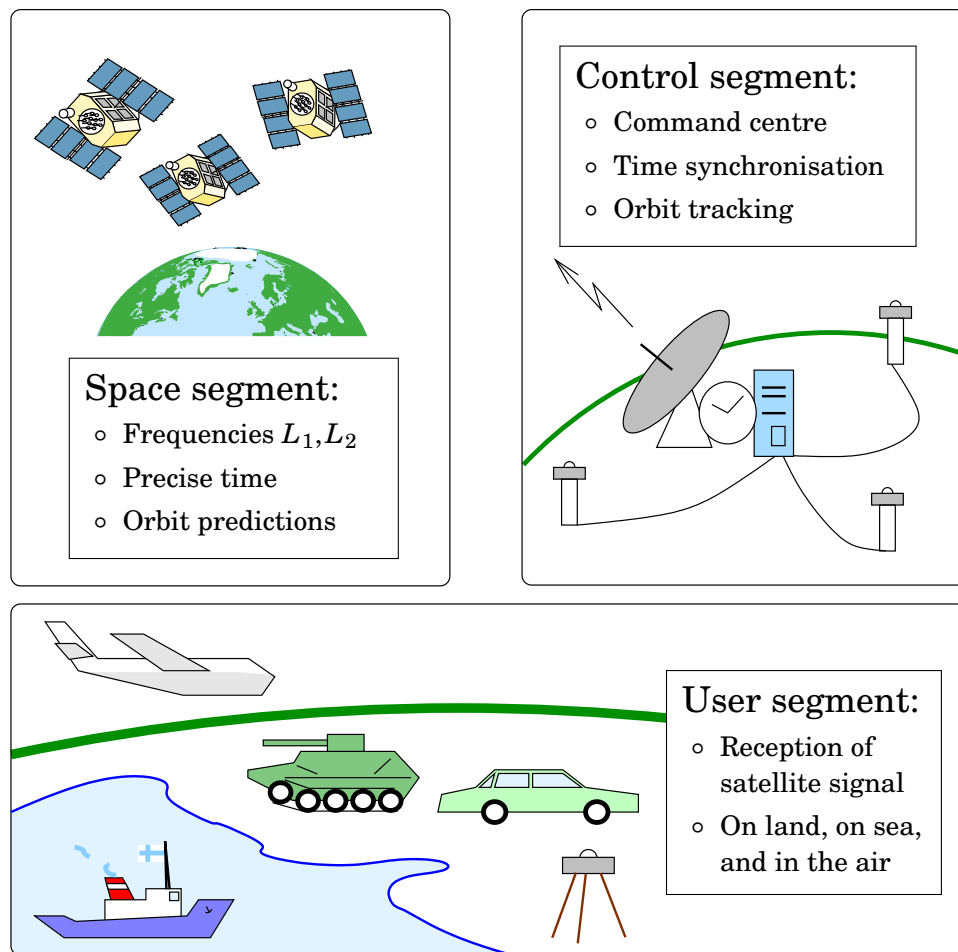


FIGURE 12.5. The three segments of the [GPS](#).

tracking and control stations around the globe. Like the whole [GPS](#), the control segment also resides under the US Department of Defense, more precisely the US Air Force.

Every control and tracking station is equipped appropriately with, among other things, a [GPS](#) receiver using a precise caesium clock. Once every 24 hours, new orbital data — “broadcast ephemeris” — and correction information for the satellite’s atomic clock is uploaded to the satellites. The satellites include these orbit and clock data into the radio signal they transmit, to be used by all users.

käyttäjälohko **User segment** All users, on land, at sea and in the air — and more and more also in space, in low Earth orbits — with their receivers.



12.4 The GPS constellation

The design constellation of the GPS consists of 24 satellites and three “active spares”, which can be moved into place immediately when an active satellite breaks down. The satellites are in six different orbital planes, with four satellites in each plane. In reality there are over 30 satellites operating today.

The height of the orbits from the Earth’s surface is 20 200 km. The orbital period around the Earth is $11^{\text{h}}58^{\text{m}}$, so, in the time it takes the Earth to rotate once around her axis, $23^{\text{h}}56^{\text{m}}$, the satellites are “seen” again in the same spots in the sky as the previous day. *The GPS measurement geometry repeats every day four minutes earlier*, because the length of the day in the mean solar time used by our clocks is four minutes longer than the rotation period of the Earth.

The tilt of the orbital plane, or *inclination*, is⁷ $i = 55^\circ$. Because of this,⁷ the geometry of the GPS constellation is not very strong at high northern latitudes, from where the satellites are seen mostly in the southern sky.

The system has currently such complete coverage that at least four satellites are “visible” (meaning that they are above an elevation angle of 10°) anywhere on Earth at any time. The number of visible satellites is almost always, and usually substantially, larger than this. korkeuskulma



12.5 Codes in the GPS signal

Two *pseudo-random codes* are modulated onto the carrier waves transmitted by the GPS satellites on two different frequencies: the *C/A code* and the *P code*. In addition, there is still the navigation message — containing “broadcast ephemeris” and almanac data — which is also modulated onto the carrier waves.⁸ See table 12.1. 8

The modulation technique used is *phase modulation*:⁹ the phase of the vaihe-
modulaatio

⁷The Block I satellites had a different inclination, $i = 63^\circ$. None of those satellites are working today.

⁸In connection with GPS modernisation, frequency L_5 , 1176.45 MHz, has been added to the signal. It is meant to be used in Safety of Life (SoL) critical operations, like aviation and rescue services. In addition, new civilian and military codes have also been added to the L_1 and L_2 frequencies.

⁹Other modulation types in existence are *amplitude modulation* — where the strength of the carrier, the amplitude, is varied in the rhythm of the signal — and *frequency modulation* — where the frequency of the carrier is made to vary. Common radio

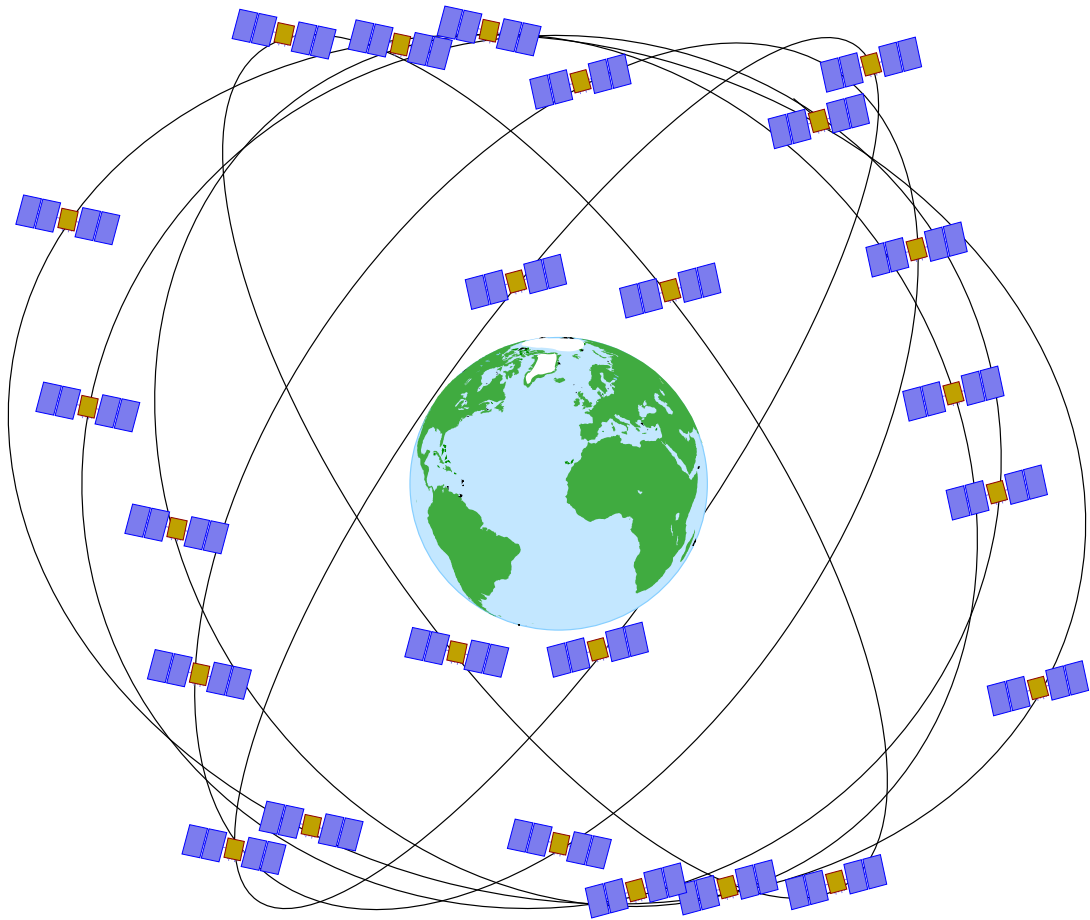


FIGURE 12.6. The original GPS constellation. The orbits and locations of satellites are realistic with respect to the Earth.

carrier is turned “upside down”, meaning it is shifted by an angle of 180 degrees, or π , when the code switches between zero (0) and one (1).

Because of the modulation, the signal transmitted by the GPS satellites

TABLE 12.1. Codes included in the GPS signal.

Name	Explanation	Modulation frequency	Repeat period	Carrier wave
C/A code	Coarse / Acquisition, Civilian Access	1.023 Mb/s	1 ms	L_1
P code	Precise / Protected	10.23 Mb/s	1 week	L_1, L_2
P(Y)	Combination of P code and secret W code	Same		Same
-	Navigation message	50 bits/s	12.5 minutes	L_1, L_2

stations use amplitude modulation, and FM stations frequency modulation.



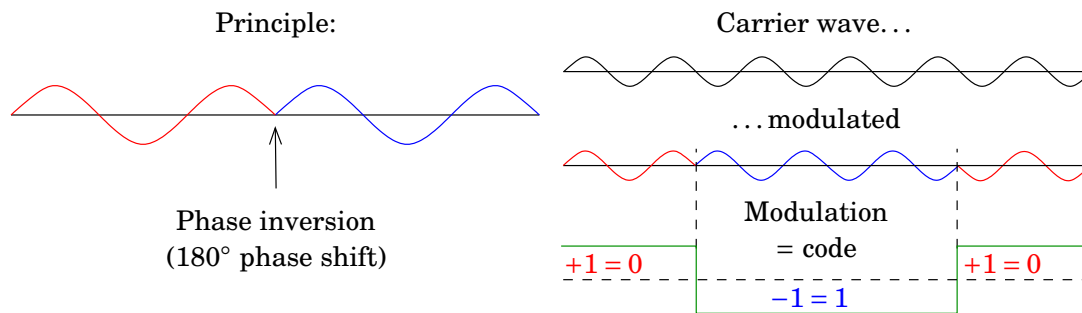


FIGURE 12.7. The principle of phase modulation.

is pretty broadband. The effective bandwidth is several times the bit frequency of the **P code**, that is several tens of MHz.

When already the transmission bandwidth requirement for a single satellite is this large, one might think that the bandwidth demand of the whole constellation would be huge. This is however not so: *all satellites use the same carrier-wave frequencies L_1 and L_2* . The receiver is able to separate the signals of different satellites from each other with the aid of their different *pseudo-random codes (C/A code and P code)*. Every satellite has its own code, or “fingerprint”, in the same way that in navigation at sea, every lighthouse has its own flashing sequence. The technical solution goes by the name **CDMA**, or “code division multiple access”.

kantoaalto-
taajuus

koodijako-
kanavointi

The *navigation message* is a bit stream containing, besides precise orbital information on the broadcasting satellite, crude orbital information — an *almanac* — on *all* satellites. Locking on to one satellite signal is sufficient to receive the approximate orbital information for all satellites.

The pseudo-random codes are generated according to a documented mathematical recipe ([Wikipedia, Gold code](#)). The codes thus are not genuinely random: they can be exactly reconstructed using the same recipe. They however behave statistically like genuinely random bit sequences.

12.5.1 The “tree rings” of GPS

The **GPS** receiver’s antenna on the Earth’s surface receives signals from *all* the satellites that are in the sky at the moment of observation. This whole “soup” travels along the cable to the receiver’s electronics.¹⁰ Here, the first two tasks are performed:

¹⁰In the pre-amplifier of the antenna, the analogue signal is processed to bring the carrier-wave frequency down to a much lower value, without affecting the modulation (“downbanding” or “heterodyning”). This prevents crosstalk of the amplified signal back

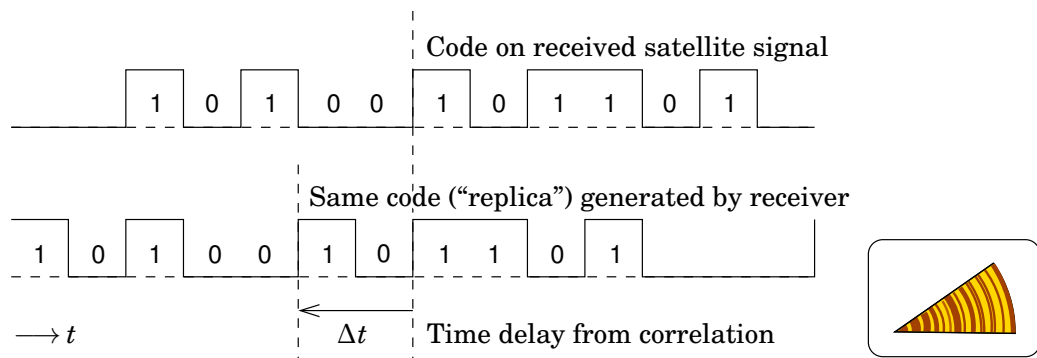


FIGURE 12.8. The correlation method for determining the travel time Δt of the GPS signal.

1. Separate the signals of different satellites from each other using their individual pseudo-random code or “fingerprint”.¹¹
2. Determine the *travel time* of the signal of each satellite from satel-

TABLEAU 12.2. How does dendrochronology work?

The method works as follows: the laboratory holds a *reference sequence* of tree rings, which has been built from partially overlapping tree-ring sequences. Rainy years show as thick, dry years as narrow rings.^a In the reference sequence, the true, absolute year number of every ring is known.

Building a reference sequence is challenging. After doing so, however, the age of any wooden object can be determined by comparing its tree rings with the reference sequence, until the place is found where they match (*correlate*). The method works because the succession of wet and dry years is largely random. Similar methods are in use in many fields of science: dating and correlation of ice drilling cores or geological deposits, correlation of magnetisation stripes on the sea floor, and others.

^aHowever, at the tree line in the mountains or the Arctic, tree-ring thickness is controlled mostly by *temperature*. This circumstance has been used for reconstructing paleotemperature time series. It can be said that the tree-ring width is a *proxy* for temperature.

into the antenna, and makes further processing easier, like the digitisation of the signal in an A-D (analogue-to-digital) converter.

¹¹The pseudo-random codes of the different satellites have been carefully designed to be *orthogonal* to each other, meaning the true signal of one satellite correlates as weakly as possible with the replica signal of another satellite generated in the receiver, even with the correct time shift.



lite to receiver.

Both tasks are carried out by a *correlation method* which seeks the time shift Δt between the signal received from the satellite and a signal of the same form, a “replica”, generated by the receiver, which would make them identical. The replica sequence is shifted in time alongside the received one until a strong *correlation* is found, a similarity or correspondence of the pattern. The time difference obtained is in essence the GPS observable.

A suitable metaphor for the correlation method employed is the use of year rings to date wooden objects, or *dendrochronology*, which was already briefly explained in figure 7.7 and on which more in the text tableau 12.2. vuosilusto

In a way, the comparison of the received GPS code with the replica code generated by the receiver is also “dating”: the “age” of the signal travelling from the satellite to Earth is determined. . .

When the correlation processing yields $\Delta t = t_{\text{rcv}} - t_{\text{xmit}}$, multiplying it by the propagation speed c of the signal will give the *pseudo-range* to the satellite, the basic observable of GPS measurement. It is called a *pseudo-range*, because it contains more than just the geometric distance, among other things the *clock offsets*. We will return to this presently. pseudoetäisyys kellopoikkeama



12.5.2 C/A code and P code

Because the one-millisecond C/A code consists only of 1023 bits, one must look at only 1023 alternative time shifts Δt . This goes very quickly. Each satellite has its own “personal” C/A code, therefore, initially $N \times 1023$ comparisons have to be made, where N is the number of satellites.

The C/A code can be used to determine the pseudo-range only modulo 300 kilometres, because the code repeats every millisecond, the time in which the signal travels 300 kilometres. See figure 12.9. This is good enough if the approximate location of the receiver is already known with this accuracy. The state of modulation of the C/A code switches, if it switches, at intervals of one microsecond (“chip rate”), the time in which the radio signal travels 293 metres. The *accuracy of measurement* when using the C/A code is better than this: if the receiver electronics can measure the phase of the modulation with an accuracy of 1% of a full cycle, then the measurement accuracy is ± 3 m. modulaation vaihe

Better accuracy is offered by the P code. It, too, is a pseudo-random code, but its length is no less than 267 days. Every satellite uses its own,



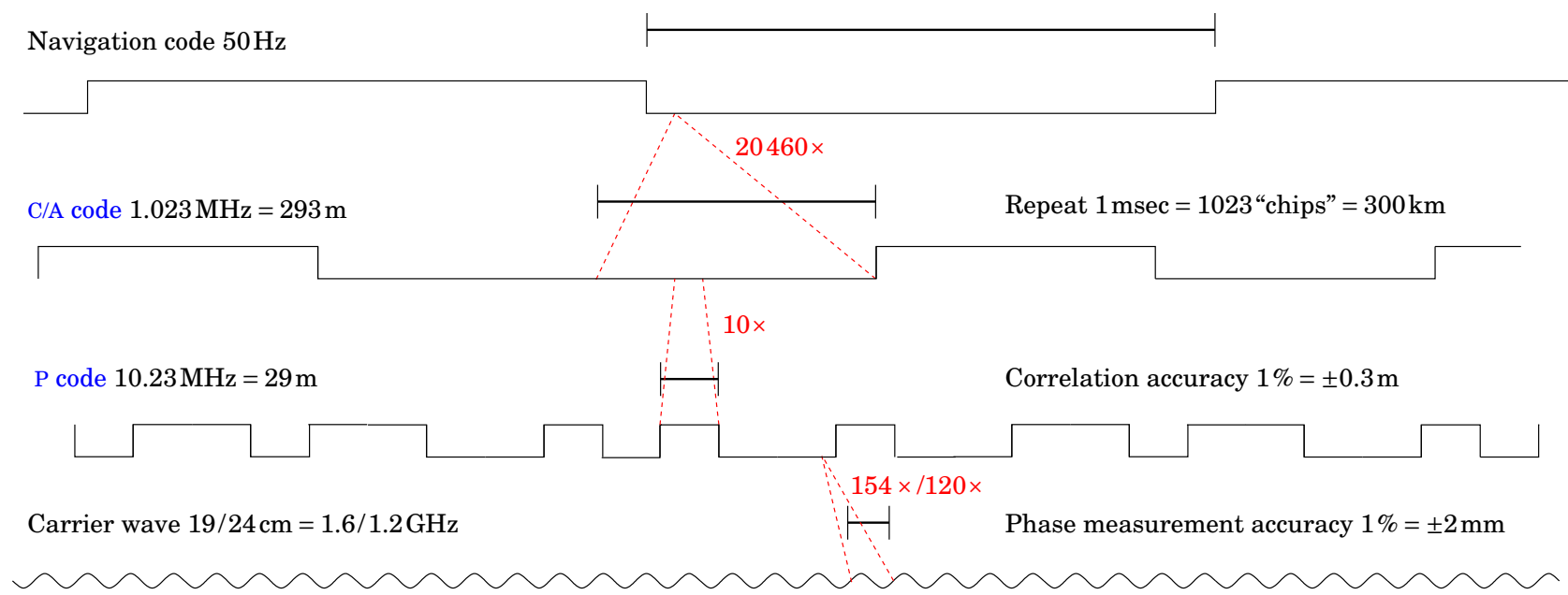


FIGURE 12.9. The various frequencies and effective wavelengths of the GPS signal.



satellite-specific, one-week long subinterval of this period. The receiver must also in this case be able to generate a “replica” of the code. Because, however, the quantity Δt has already been obtained, using the **C/A code**, to microsecond accuracy, only decimals more precise than this need to be looked at.

The **P code** “chip rate” or bit frequency is ten times faster than that of the **C/A code**, 10.23 Mb/s , which corresponds to a travelled distance of 29 m.¹² Again assuming a modulation phase measurement accuracy of 1%, this corresponds to an accuracy of the pseudo-range observation of $\pm 30 \text{ cm}$.

The **P code** is kept from civilian users by encryption. The encryption is done by modulating, on top of the **P code**, a **W code**, the generating algorithm for which has not been published.

Both the **P code** and the **C/A code** are thus modulated on the L_1 carrier frequency. Distinguishing between them has been made possible by using *phase quadrature*: while the **P code** is modulated with phase shifts 0 (bit value 0) and π (bit value 1), the **C/A code** is similarly modulated with phase shifts $+\frac{1}{2}\pi$ and $-\frac{1}{2}\pi$, or $+90^\circ$ and -90° . One speaks of “in-phase” and “quadrature” modulations.

The phase angles of the navigation message are the same as those of the **C/A code**. This is not a problem as their frequencies are so different: twenty repeats of the whole **C/A code** fit within one bit of the navigation message, duration 20 ms! See figure 12.9.

12.6 GPS receivers

The receivers intended for precise geodetic work are always *dual-frequency* instruments that can measure the *carrier phase* of the **GPS** signal. Unlike with inexpensive hand-held devices, the antenna is usually separate and connected to the receiver by a coax cable. The weak satellite signal is already amplified inside the antenna by a *pre-amplifier*.¹³

¹²This is the “effective wavelength” of the **P code**. It is calculated as follows:

$$\lambda_{\text{eff}} = \frac{c}{f},$$

in which f is the “chip rate”, a frequency-like quantity, unit s^{-1} or Hz, and c is the speed of light. So if $c = 300\,000\,000 \text{ m/s}$ and $f = 10\,230\,000 \text{ s}^{-1}$, it follows that $\lambda_{\text{eff}} = 29.3 \text{ m}$.



12

suora vaihe

kaksitaajuus-
koje
kantoaallon
vaiheesivahvistin
13



FIGURE 12.10. Control panel of the Ashtech Z-12.

pakkokeskistys-
laite
tosiaikainen

The antenna may be mounted, using a standard forced-centring device, onto a geodetic tripod. In real-time mapping surveys however, a *measuring pole* or staff is used, with the antenna screwed onto the top. Attached to the pole is the GPS receiver (if separate) with peripherals. The antenna has a standard $\frac{5}{8}$ inch hole at the bottom with a similar standard screw thread (11 threads per inch) as in most geodetic instruments.

With every receiver type comes its own antenna type; the antennas and receivers of different manufacturers are not generally electrically intercompatible. However, the *choke-ring* antenna model (figure 12.11) is available for precise geodetic work from many different manufacturers. Choke rings lessen the problem of reflection of radio waves from the Earth's surface and other surfaces, or *multipath*.

The *electric centre* of the antenna, the point where, in the geometric interpretation, the radio waves are apparently received, is not the same as the antenna's official reference point (ARP). It is not even unambiguously defined, but depends somewhat on the cut-off elevation angle used for the observations, see figure 12.12. As a metaphor, one may think of the apparent place of a fish under water, which also depends on the angle of view. One speaks of the *variation of the antenna's phase centre*, see Poutanen (2017) section 7.4 or Hofmann-Wellenhof et al. (2001) section

¹³The DC voltage feed required by the pre-amplifier comes from the receiver, also through the coax cable. This may complicate or prevent the mixing of receivers and antennas from different brands.



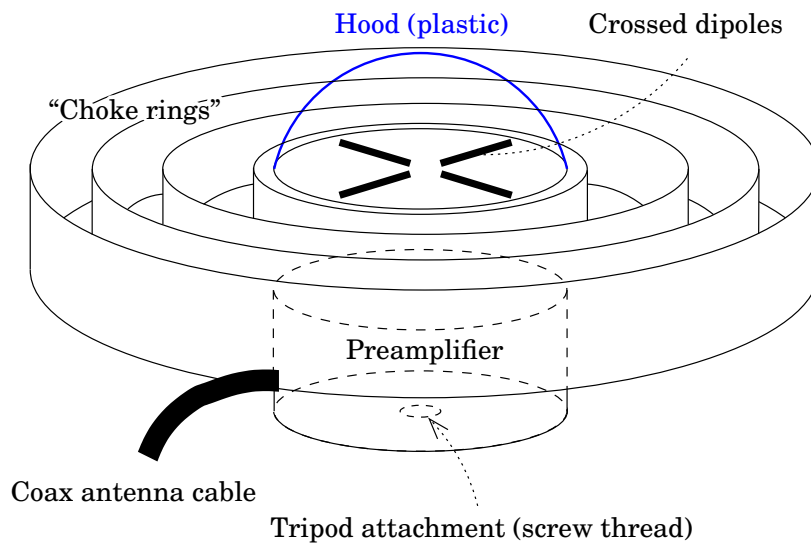


FIGURE 12.11. A *choke-ring* GPS antenna for precise geodetic work. The antennas provided by manufacturers are simpler and especially smaller and lighter.



6.5.

When measurements are carried out over a relatively small area, using only one type of antenna, this variation of the phase centre vanishes from the end result, and from the position *difference* vectors computed between different points in the network. If, however, antenna types are mixed, or extensive networks measured — hundreds or thousands of kilometres across — one ought to *calibrate* the phase delay patterns of the antennas, which are fairly complicated functions of both the elevation angle η and the azimuth direction α . The calibration in which this phase delay pattern $\Delta\phi(\eta, \alpha)$ is determined can be carried out in the laboratory using an artificial GPS signal source, or as a field calibration in which two antenna types are always compared to each other. Field calibration is thus always *relative*, referring to some agreed reference type of antenna.

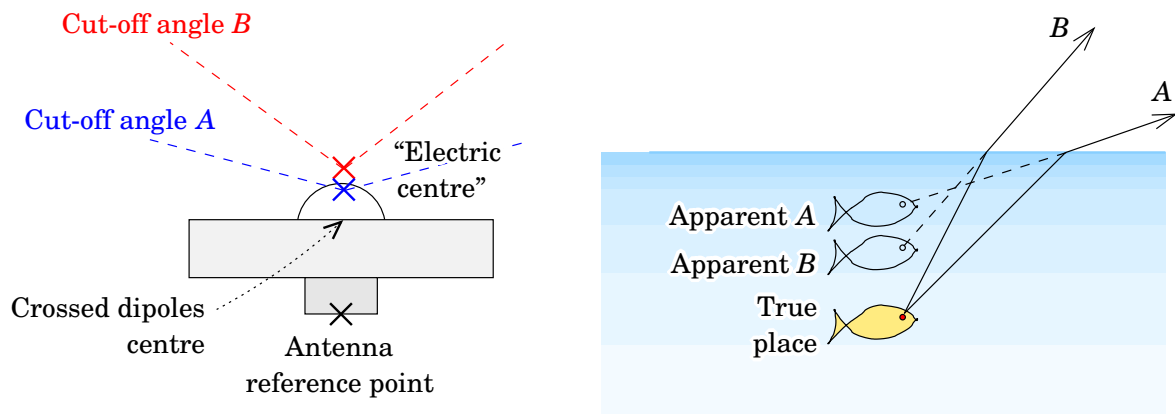
vaiheen
kulkuviive


In highly precise geodynamic deformation measurements it is nowadays the practice to calibrate, not just antenna *types*, but *individual antennas*.

The radio waves transmitted by GPS satellites are clockwise *circularly polarised*. Upon reflection, the polarisation direction reverses to anti-clockwise. The antenna — in the example in figure 12.11, a cross dipole — is built so that it transfers only the clockwise polarised signal on to the receiver. In this way, the harm caused by reflections is minimised.

Technological development goes into the direction of greater integration.





 FIGURE 12.12. The electric centre of an antenna is not a self-evident thing!


Today's **GPS** positioning instruments including those for geodetic use are so small that they are integrated with the antenna. Because the instruments are very autonomous, they no longer even have a proper display screen.

ohjelmisto-
pohjainen
vastaanotin

Another development is the onwards march of *software-defined receivers*: today's personal computers are beginning to be powerful enough to do the digital processing work that in today's receivers is done on the hardware level. Then one only needs, in addition to a general-purpose **PC**, a "dumb" analogue radio device with antenna (**Lázaro, 2012**).

12.7 Observables of GPS

A geodetic **GPS** receiver stores the observations it makes into its memory as a long table containing many numbers. The numbers in the table represent *distances* between the receiver and the different satellites observed. It is easy to understand why the number count becomes so large: if, for example, the time between measurements is 30 seconds, and five satellites are visible in the local sky, and with a single-frequency receiver we observe both the **C/A code** and the **P code**, then the number of

 TABLEAU 12.3. Start of a **RINEX** file. The device collects *five observation types*: carrier-phase observations and **P code** observations on both frequencies L_1 and L_2 , as well as **C/A code** observations on frequency L_1 . The observations are stored at intervals of *30 seconds*. There are eleven satellites on the first epoch, *1 January 2000 0 : 00 : 00*, and also on the second epoch, *0 : 00 : 30*. They are all **GPS** satellites (G). The observation station is DGAR, Diego Garcia in the Indian Ocean (**Vine, 2011**). →



```

2.00      OBSERVATION DATA      G (GPS)      RINEX VERSION / TYPE
teqc 19990ct8      gpsops      20000103 21:13:37UTCPGM / RUN BY / DATE
OSF1 V4.0 564|Alpha|cc 4.4.18.4|+=|=|      COMMENT
DGAR      MARKER NAME
30802M001      MARKER NUMBER
GNOG      JPL      OBSERVER / AGENCY
T341U      AOA SNR-8000 ACT      3.3.32.3      REC # / TYPE / VERS
250      AOAD/M_T      ANT # / TYPE
1916269.8405 6029977.3167 -801720.2273      APPROX POSITION XYZ
0.0814      0.0000      0.0000      ANTENNA: DELTA H/E/N
1 1      WAVELENGTH FACT L1/2
5 L1 L2 P1 P2 C1      # / TYPES OF OBSERV
30.0000      INTERVAL
This data is provided as a public service by NASA/JPL.      COMMENT
No warranty is expressed or implied regarding suitability      COMMENT
for use. For further information, contact:      COMMENT
Dave Stowers, NASA/JPL m/s 238-600      COMMENT
4800 Oak Grove Drive, Pasadena CA 91109 USA      COMMENT
2000 1 1 0 0 0.0000000 GPS      TIME OF FIRST OBS
END OF HEADER
00 1 1 0 0 0.0000000 0 11G21G23G17G30G 1G31G29G22G15G25G 3
-6078127.503 4 -4736200.435 4 23397694.178 23397698.378 23397695.030
-2556364.753 4 -1991958.142 4 24025055.814 24025059.840 24025056.373
-8073501.747 5 -6291024.522 5 22565280.025 22565283.337 22565280.587
18247234.140 4 14218628.480 4 24610505.696 24610508.842 24610505.143
-1299479.831 4 -1012581.476 4 24824108.761 24824113.289 24824108.748
-5233446.124 4 -4077998.775 4 24175634.461 24175638.438 24175635.537
16878293.604 4 13151917.927 4 24427189.279 24427193.024 24427188.034
-13489828.171 5 -10511530.918 5 22792735.726 22792739.295 22792736.451
-4494062.929 4 -3501865.147 4 23961699.555 23961704.148 23961699.834
-21099958.763 9 -16441519.960 9 20331187.861 20331190.408 20331187.808
-15215098.740 5 -11855903.290 5 22202394.742 22202398.143 22202394.948
00 1 1 0 0 30.0000000 0 11G21G23G17G30G 1G31G29G22G15G25G 3
-6132427.986 4 -4778512.477 4 23387361.455 23387365.218 23387361.703
-2586441.342 4 -2015394.448 4 24019332.425 24019336.115 24019332.872
-7990741.587 5 -6226536.097 5 22581028.707 22581032.399 22581029.091
18274808.415 4 14240114.880 4 24615752.673 24615756.082 24615752.310
-1317133.094 4 -1026337.267 4 24820749.907 24820754.243 24820749.907
-5259685.471 4 -4098444.960 4 24170641.011 24170645.465 24170641.809
16938081.982 4 13198506.304 4 24438566.644 24438570.267 24438567.293
-13548930.874 5 -10557584.968 5 22781488.710 22781492.153 22781489.545
-4600217.585 4 -3584583.008 4 23941499.061 23941503.148 23941500.039
-21083529.873 9 -16428718.216 9 20334314.137 20334316.731 20334314.115
-15210926.086 5 -11852651.873 5 22203188.835 22203192.175 22203189.231

```



observations collected within one minute already amounts to $(60/30) \cdot 5 \cdot 2 = 20$. In an hour, this means 1200 observations; if these are stored into the memory as ordinary double precision real numbers¹⁴ (eight bytes per number), we need 9.4 kilobytes of storage space. In 24 hours, the storage requirement would then be 225kB.

An often-used format in international GNSS data exchange is RINEX, Receiver-Independent EXchange format, Gurtner and Estey (2007). This is a text format, human-readable, in which observational data from most geodetic receiver types can be transferred, read, and processed, in a way that is independent from the manufacturer of the receiver. See the example in tableau 12.3 from the Diego Garcia station (Vine, 2011) in the Indian Ocean.



12.7.1 Pseudo-ranges as observables

etuliite Why does one speak of *pseudo*-ranges? The prefix “pseudo” is because the observable’s value is affected by, in addition to the geometric distance between satellite and receiver, the clock errors or *offsets* Δt and ΔT , as well as the propagation delays caused by the ionospheric and tropospheric media.¹⁵ Thus the *observation equation for pseudo-range* is obtained:

$$p = \rho + c(\Delta T - \Delta t) + d_{\text{ion}} + d_{\text{trop}}, \quad (12.1)$$

in which

p pseudo-range

ρ natural (geometric) distance. According to Pythagoras

$$\rho = \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2},$$

in which

$\begin{bmatrix} x & y & z \end{bmatrix}^T$ location of the satellite in space

$\begin{bmatrix} X & Y & Z \end{bmatrix}^T$ location of the receiver in space

c speed of light in a vacuum

ΔT offset of receiver clock from GPS time (clock correction = $-\Delta T$)

¹⁴The “real” numbers of computers are in fact rational numbers. . .

¹⁵More precisely: the effects of free electrons and neutral molecules, respectively, on the propagation of radio waves. Therefore, the stratosphere also causes “tropospheric delay”.



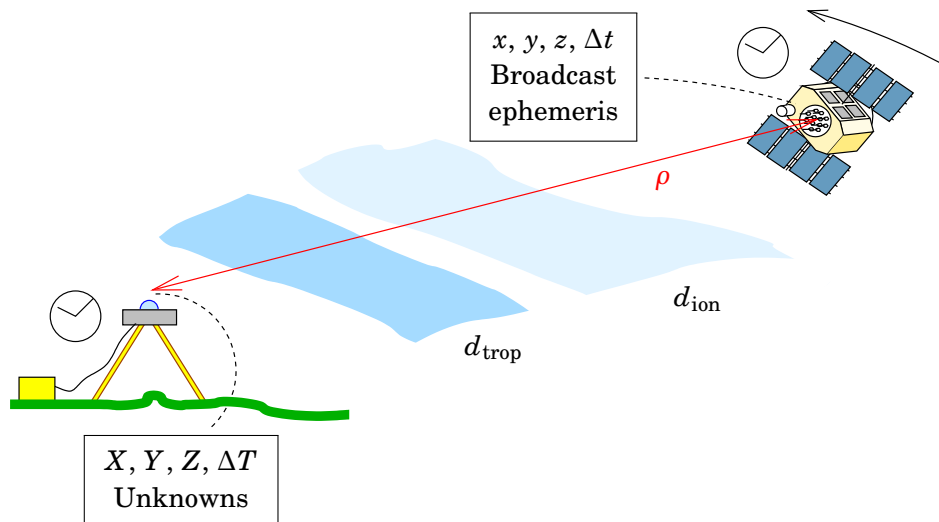


FIGURE 12.13. Pseudo-range observation.

Δt offset of satellite clock from GPS time

d_{ion} “ionospheric” — meaning caused by the ionised atmosphere — propagation delay

d_{trop} “tropospheric” — meaning caused by the neutral atmosphere — delay.


The clock offset Δt of the satellite is included in the *broadcast ephemeris* transmitted by the satellites. The receiver clock offset ΔT on the other hand remains unknown, and it will have to be estimated as one unknown together with the co-ordinates. Thus there are *four* unknowns for every GPS receiver: three co-ordinates X, Y , and Z , and the clock offset ΔT — if we forget for a moment about the atmospheric unknowns d_{ion} and d_{trop} . Four pseudo-range observations, to *four different satellites*, are enough to determine four unknowns. See figure 12.13.

12.7.2 The carrier wave’s phase angle as observable

The “chip rate” of the C/A code, the number of bits transmitted per second, is 1.023MHz, corresponding to a “wavelength” of 293m, when the corresponding number for the P code, 10.23MHz, means a wavelength of 29.3m. If, instead of a modulation, we use the *carrier wave* itself, the relevant wavelength becomes 19cm (L_1) or 24.4 cm (L_2), a distance that is two orders of magnitude shorter. *Geodetic GPS positioning*, where dual-frequency receivers observe the phase of the carrier waves transmitted by the GPS satellites, is based on this. Electronic phase measurement is relatively easy and precise, but, as the market is small and specialised,

vaihe



 TABLE 12.4. Properties of GPS carrier waves. “Factor” denotes the multiple of the base frequency (10.23 MHz).

Carrier wave	Frequency (MHz)	Wavelength (cm)	Factor
L_1	1575.42	19.0	154×
L_2	1227.60	24.4	120×

the prices of the devices are nevertheless as high as for other geodetic instruments, upwards of thousands of euros.

As is also the case with electronic range-finders (distance measurement devices), phase measurement is always afflicted by an *ambiguity problem*. *Measured* phase values ϕ may be expressed as values within the single-cycle interval $[0, 2\pi)$, and the corresponding pseudo-range between satellite and receiver can be determined only “modulo an integer number of wavelengths”. If a certain pseudo-range value P is compatible with a measurement done on wavelength λ , then the pseudo-range values $P + \lambda$, $P - \lambda$, $P + 2\lambda$, $P - 2\lambda$, ... are also compatible.

16 The *observation equation for the carrier phase* is, as a phase difference angle,¹⁶ in radian units

$$\bar{\phi} \stackrel{\text{def}}{=} \phi + 2\pi N = 2\pi \frac{\rho + c(\Delta T - \Delta t) + D_{\text{ion}} + D_{\text{trop}}}{\lambda}$$

or as a distance, unit metres

$$\bar{P} \stackrel{\text{def}}{=} \lambda \frac{\bar{\phi}}{2\pi} = \lambda \left(\frac{\phi}{2\pi} + N \right) = \rho + c(\Delta T - \Delta t) + D_{\text{ion}} + D_{\text{trop}}. \quad (12.2)$$

The left-hand side of the observation equation is either the phase angle $\bar{\phi} = \phi + 2\pi N$ or the equivalent pseudo-range $\bar{P} \stackrel{\text{def}}{=} P + \lambda N$, see figure 12.14. Both contain the correct integer number N of wavelengths making up the range between satellite and receiver. The symbol ϕ without overbar, on the other hand, designates the “raw” phase difference measurement. It is often assumed that $\phi \in [0, 2\pi)$ at least at the moment of signal acquisition.

Other symbols used in the equation:

D_{ion} the propagation delay of the carrier wave caused by the ionosphere
(which is actually negative, $D_{\text{ion}} = -d_{\text{ion}}$)

¹⁶The measured quantity is actually the phase of the received radio wave subtracted from that of the receiver’s reference oscillator, representing the *delay* in transit from satellite to receiver. In the difference, the phase angle of the received wave enters with a negative algebraic sign.



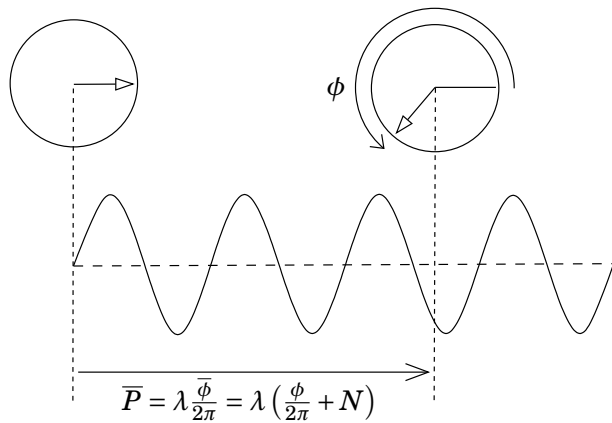


FIGURE 12.14. Measurement of the phase of the GPS signal's carrier wave. The measured phase difference angle is ϕ , initially $\phi \in [0, 2\pi)$, the metric pseudo-range, including whole wavelengths, is \bar{P} , and the ambiguity is N , in this example $N = 2$.



D_{trop} the delay caused by the troposphere, $D_{\text{trop}} = d_{\text{trop}}$

λ the wavelength of the received carrier wave according to table 12.4, including Doppler shift

N the integer unknown or *ambiguity*.

kokonaisluku-
tuntematon

The variables $\rho \left(= \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2} \right)$, Δt and ΔT are the same as those for code pseudo-range, see equation 12.1. There are again four geodetic, real-valued unknowns: X , Y , Z , and ΔT . But now the integer values N , one for each observation, also need to be determined in addition to the geodetic unknowns. This may seem impossible, but there are ways of doing it; see section 13.3.

In equations 12.1 and 12.2 we have used different symbols for the ionospheric and tropospheric delays of the carrier wave, than for the corresponding delays of code measurement, because *they are different*. One says that the ionosphere is *dispersive* for radio waves: the speed of propagation depends on the frequency, or equivalently, on the wavelength.

The travel speed of the code modulations is the *group velocity*, which is always less than the speed of light in a vacuum. The travel speed of the carrier phase is the *phase velocity*. In a dispersive medium these two speeds differ from each other.

12.7.3 Effects of the ionosphere and troposphere

The ionosphere is a *dispersive* medium to radio waves: different frequencies propagate at different velocities. As a consequence of dispersion,



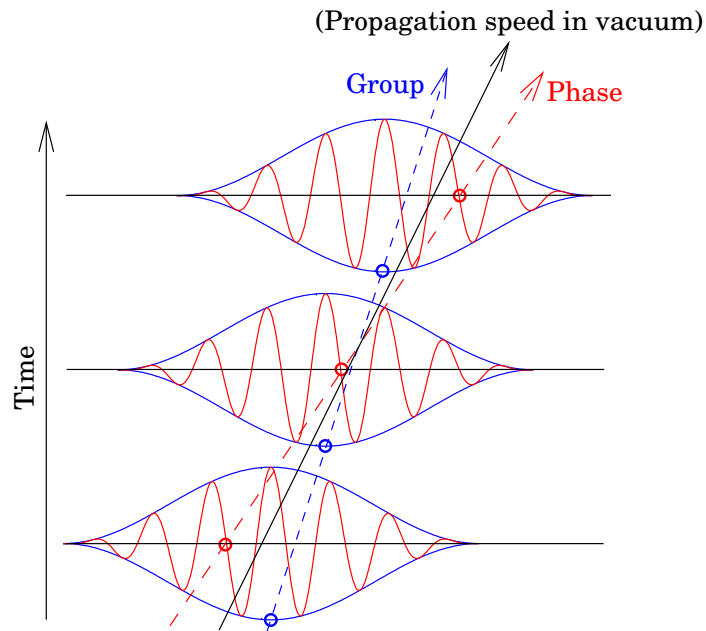


FIGURE 12.15. The propagation of a wave packet in a dispersive medium, phase and group velocity. The carrier wave travels with phase velocity, the modulations — also the pseudo-random codes of the GPS signal — travel at group velocity.



vaihetaitte-
kerroin

the phase and group velocities of propagation are different. For phase

$$n_p = \sqrt{1 - \frac{f_p^2}{f^2}} \approx 1 + \frac{c_2}{f^2} + \frac{c_4}{f^4} + \frac{c_6}{f^6} + \dots \quad (12.3)$$

In this equation the constants c_i and the plasma frequency f_p depend on the electron density, n_e , of the ionosphere. A good approximate formula, which explains 99.9% of the whole ionospheric propagation effect, is (Seeber, 1989):

$$n_p = 1 - \frac{C}{f^2}, \quad C = 40.3 n_e \text{ m}^3/\text{s}^2,$$

in which the electron density n_e is expressed in electrons per m^3 . Typical numbers are $10^8 - 10^{12} \text{ m}^{-3}$. Electron density varies between day and night — greater in the daytime — with the season — greater in summer — with solar activity, and of course with the latitude and height of the location.

ryhmä-
taitekerroin
17

The group propagation velocity, more precisely the group index of refraction, is obtained as the derivative with respect to the frequency¹⁷ of

¹⁷When the propagation velocity is $c' = c/n$, in which c is the speed of light in a vacuum,



$n_p f$,

$$n_g = \frac{d(n_p f)}{df} = 1 + \frac{C}{f^2}, \quad C \text{ like above.}$$

Because the index of refraction of the ionosphere is dependent upon the frequency f , it is possible to *eliminate* the effect of the ionosphere by combining measurements made on two different frequencies. This is the fundamental reason why the GPS uses two different frequencies L_1 and L_2 .

If we form the linear combination of code observations

$$p_3 \stackrel{\text{def}}{=} \frac{f_1^2 p_1 - f_2^2 p_2}{f_1^2 - f_2^2}$$

and correspondingly

$$n_{g,3} \stackrel{\text{def}}{=} \frac{f_1^2 n_{g,1} - f_2^2 n_{g,2}}{f_1^2 - f_2^2},$$

we obtain

$$n_{g,3} = \frac{f_1^2 (1 + C/f_1^2) - f_2^2 (1 + C/f_2^2)}{f_1^2 - f_2^2} = \frac{f_1^2 - f_2^2}{f_1^2 - f_2^2} = 1,$$

from which it is seen that the effect of the ionosphere has vanished.¹⁸ 18

The *troposphere* — more precisely, the neutral atmosphere, which also includes the stratosphere, and neutral fractions of higher layers — on the other hand is not dispersive to radio waves. Its speciality is, however, a strong dependence on *water vapour content* of the index of refraction. The equation, the same equation 7.6 that also applies for electronic distance measurement, is (Rüeger, 1996, 2002):

$$N_M = 10^6 \cdot (n_M - 1) = \frac{77.624 \text{ K/hPa}}{T} (p - e) + \frac{64.70 \text{ K/hPa}}{T} \left(1 + \frac{5748 \text{ K}}{T} \right) e,$$

in which n_M is the index of refraction for microwaves. Here, the unit of p and e is hPa (hectopascal), or millibar. T is the absolute temperature in kelvin. taitekerroin

it follows that the phase propagation velocity is greater than the speed of light. The carrier phase cannot however carry information, so the directionality of time according to thermodynamics is preserved. . . if it were possible to move information faster than light, it would, according to special relativity, also be possible to carry information back in time!

¹⁸The result and its proof are the same if we take, instead of the group index of refraction, the phase one, n_p .



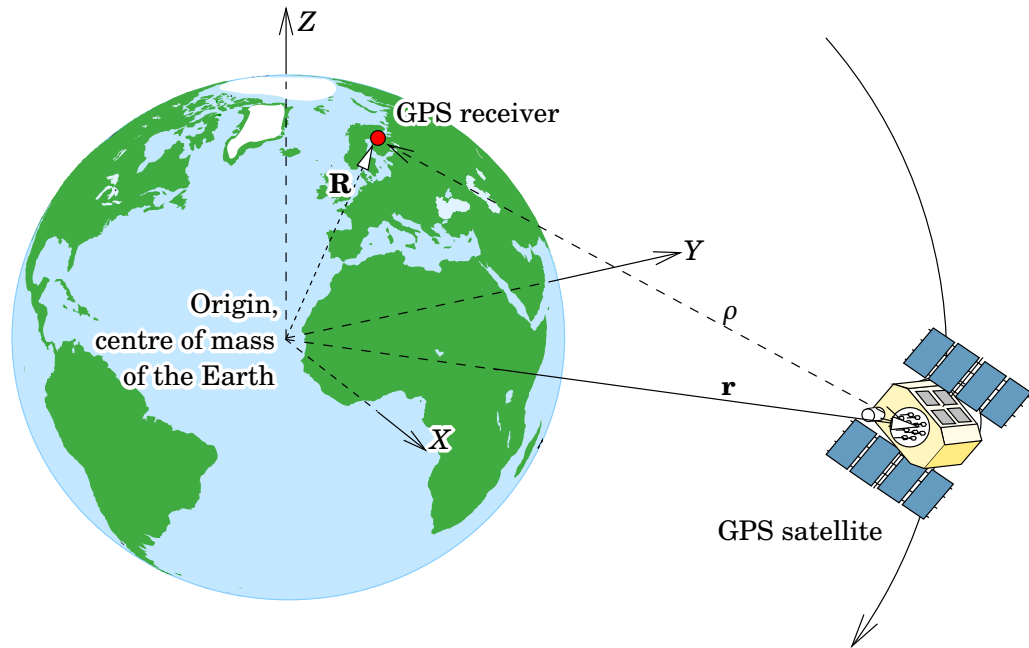


FIGURE 12.16. Geometry of GPS positioning.

The coefficient of the partial pressure of water vapour e in this equation is at ambient temperature as much as 17 times the coefficient for dry-air partial pressure $p - e$!

Because both the ionosphere and the troposphere affect the propagation of GPS radio waves, GPS measurements can be used for both ionospheric and tropospheric research. Meteorologists and climate researchers are very interested in this; see section 18.7.



12.8 GPS measurement geometry

See figure 12.16. Let the position vector of satellite S in a geocentric frame be $\mathbf{r} = \mathbf{r}^S$, the position vector of observation station A in the same frame be $\mathbf{R} = \mathbf{R}_A$, and let the distance between the two be $\rho = \rho_A^S$.

Then the following vector equation applies:

$$\mathbf{r}^S = \mathbf{R}_A + \mathbf{e}_A^S \rho_A^S,$$

in which \mathbf{e}_A^S is the *direction* (unit vector) to satellite S seen from observation station A . The task of GPS positioning is to compute \mathbf{R}_A when ρ_A^S is given to sufficiently many satellites S .

The Pythagoras theorem gives

$$\rho_A^S = \|\mathbf{r}^S - \mathbf{R}_A\| = \sqrt{(x^S - X_A)^2 + (y^S - Y_A)^2 + (z^S - Z_A)^2}, \quad (12.4)$$



in which

$$\mathbf{r}^S = x^S \mathbf{i} + y^S \mathbf{j} + z^S \mathbf{k}, \quad \mathbf{R}_A = X_A \mathbf{i} + Y_A \mathbf{j} + Z_A \mathbf{k}$$

are the geocentric position vectors of satellite S and observation station A . The unit vectors \mathbf{i} , \mathbf{j} , and \mathbf{k} are aligned with the co-ordinate axes: $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ forms an *orthonormal basis* in Euclidean space.

ortonormaalikanta

Often one writes the vectors as column vectors of their components, the co-ordinates of location:

$$\bar{\mathbf{r}}^S \stackrel{\text{def}}{=} \begin{bmatrix} x^S \\ y^S \\ z^S \end{bmatrix}, \quad \bar{\mathbf{R}}_A \stackrel{\text{def}}{=} \begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix}.$$

From the observations, however, no true distances ρ are obtained, but *pseudo-ranges* p , equation 12.1, or \bar{P} , equation 12.2.

The atmospheric propagation delays d_{ion} , d_{trop} , D_{ion} , and D_{trop} must also be taken into account in some way. The alternatives are

- Elimination from the observation equations — as described in subsection 12.7.3 for the ionospheric effects d_{ion} and D_{ion} .
- Computation using a good, externally provided atmospheric model.
- Modelling of the atmospheric effect using unknown parameters, to be estimated from the same observation equations — as we will describe for the tropospheric effects d_{trop} and D_{trop} in subsection 18.7.1.



12.9 Measurement geometry and sensitivity of observations

Above it was shown that the observables of GPS are *pseudo-ranges*, the observation equation of which looks like this:

$$p = \rho + c(\Delta T - \Delta t) + d_{\text{ion}} + d_{\text{trop}}.$$

Let us leave out the effect of the atmosphere, and assume also that the satellite's orbit — and thus, the momentaneous position vector of the satellite in space as computed from the orbit and clock time — and the satellite's clock offset Δt are *known*, in other words already taken into account:

$$p = \rho + c \Delta T,$$

in which ρ is the geometric distance between satellite and receiver, and ΔT is the clock offset of the receiver.



We can write equation 12.4 more simply:

$$\rho_A^S = \rho = \sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2},$$

in which $\begin{bmatrix} x & y & z \end{bmatrix}^\top$ is the known position vector of the satellite in space, computed from the orbit information, also called the ephemeris. $\begin{bmatrix} X & Y & Z \end{bmatrix}^\top$ is the receiver position vector. Now

$$p = \sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2} + c \Delta T,$$

in which there are *four unknowns*, X , Y , Z , and ΔT .

In what follows, we shall use instead the expression $c \Delta T$ as the unknown, as it has the same dimension, length, as the co-ordinate unknowns.

To solve four unknowns, it is sufficient to have observations to four satellites. If the number of useable satellites is larger, we have redundancy and an *adjustment problem*.

häiriö Question In what way do small “perturbances” in the location of the receiver impact a certain measurement p ?

Answer Look at the *place* of the satellite *in the sky*. See figure 12.17. Let the satellite’s direction vector as seen from the observation site be \mathbf{e} . This is a unit vector: its length is

$$\|\mathbf{e}\| = \sqrt{e_1^2 + e_2^2 + e_3^2} = 1.$$

If the satellite’s place in the sky is azimuth α , elevation η , then

$$\mathbf{e} \stackrel{\text{def}}{=} e_1 \cdot \mathbf{N} + e_2 \cdot \mathbf{E} + e_3 \cdot \mathbf{U}$$

and the column vector of components is

$$\bar{\mathbf{e}} \stackrel{\text{def}}{=} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} \cos \alpha \sin \zeta \\ \sin \alpha \sin \zeta \\ \cos \zeta \end{bmatrix} = \begin{bmatrix} \cos \alpha \cos \eta \\ \sin \alpha \cos \eta \\ \sin \eta \end{bmatrix} = \frac{1}{\rho} \begin{bmatrix} x - X \\ y - Y \\ z - Z \end{bmatrix}.$$

Here, $\{\mathbf{N}, \mathbf{E}, \mathbf{U}\}$ is the orthonormal basis of unit vectors in the local-horizon system (“north, east, up”). $\eta = 90^\circ - \zeta$ is the satellite’s elevation angle above the horizon.

We carry out a *sensitivity analysis*. In what way do small coordinate shifts ΔX , ΔY , or ΔZ in the location of the observation site influence the observable p ?



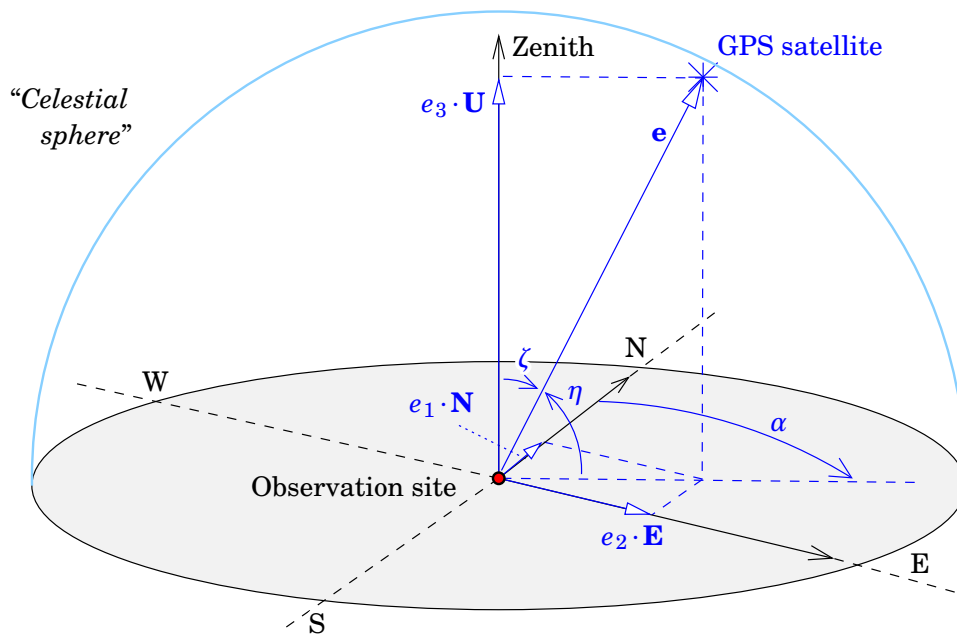


FIGURE 12.17. The geometry between a GPS satellite and an observation site. $\{\mathbf{N}, \mathbf{E}, \mathbf{U}\}$ is the local-horizon orthonormal basis (“north, east, up”).



- If the co-ordinate shift influences highly, then the observation p will help in determining the unknown co-ordinate in question.
- If the co-ordinate shift does not influence at all, then the unknown in question cannot be determined using observation p .
- The greater the *sensitivity* of the observations, the better the precision of the solution of the unknowns.


Intuitively observation p is most affected, in a ratio of one-on-one, by a shift in observation site location along the satellite’s direction vector \mathbf{e} . Shifts in observation site location perpendicular to this direction vector have no effect at all.

Equation:

$$\text{“Effect”} = -\langle \Delta \mathbf{R} \cdot \mathbf{e} \rangle = -(\Delta X e_1 + \Delta Y e_2 + \Delta Z e_3). \quad (12.5)$$

This intuitive result may also be derived more formally by *linearisation*. See tableau 12.5 and section 14.6.



 TABLEAU 12.5. A more rigorous derivation of the influence formula 12.5 by means of linearisation.

Choose for the observation station $[X \ Y \ Z]^T$ an approximate location $[X_0 \ Y_0 \ Z_0]^T$, and an approximate clock offset $\Delta T = 0$. Then we can construct an approximate pseudo-range observation

$$p_0 = \sqrt{(x - X_0)^2 + (y - Y_0)^2 + (z - Z_0)^2}.$$

Carry out a Taylor series expansion in the neighbourhood of the approximate location, around the value p_0 . The first, linear terms yield

$$\begin{aligned} p &\approx p_0 + \frac{\partial p}{\partial X} (X - X_0) + \frac{\partial p}{\partial Y} (Y - Y_0) + \frac{\partial p}{\partial Z} (Z - Z_0) + c\Delta T \\ \implies \Delta p = p - p_0 &\approx \frac{\partial p}{\partial X} \Delta X + \frac{\partial p}{\partial Y} \Delta Y + \frac{\partial p}{\partial Z} \Delta Z + c\Delta T. \end{aligned} \quad (12.6)$$

Here, the coefficients, *partial derivatives*, are obtained as follows:

$$\frac{\partial p}{\partial X} = -\frac{x - X}{\rho}, \quad \frac{\partial p}{\partial Y} = -\frac{y - Y}{\rho}, \quad \frac{\partial p}{\partial Z} = -\frac{z - Z}{\rho}. \quad (12.7)$$

The values for the coefficients are evaluated at the approximate location $[X_0 \ Y_0 \ Z_0]^T$ instead of at the true but unknown location. This suffices, because in equation 12.6 the values $(\Delta p, \Delta X, \Delta Y, \Delta Z)$ are *small differences* between the true (p, X, Y, Z) and approximate values (p_0, X_0, Y_0, Z_0) .

The partial derivatives 12.7 are *direction cosines*, the direction, apart from the algebraic sign, to the satellite as seen from the observation station, projected onto the co-ordinate axes X , Y , and Z .

12.9.1 DOP quantities and observation equations

The measure of the quality of the measurement geometry of the GPS satellites in the local sky is DOP, dilution of precision. Using the above geometric sensitivity analysis, we may calculate various variants of DOP. A larger DOP number means a *poorer* measurement geometry!

DOP characterises the quality of the satellite *geometry*: how much worse or better one may get the co-ordinates resolved from “standard quality” measurements due to the poorer or better geometry of the satellites on offer in the sky. This is valuable information in the planning phase of measurement work.

One rule of thumb is that the measurement geometry is acceptable if GDOP < 7...10, depending on the intended use.



If we now describe the “variation” of the co-ordinates of the observation site by a (small) difference quantity $\Delta\bar{\mathbf{R}} = \begin{bmatrix} \Delta X & \Delta Y & \Delta Z \end{bmatrix}^T$, and the “variation” of the receiver clock by an (also small) difference quantity $c\Delta T$, we may express the dependence of the observable $p^{(i)}$ on these altogether four unknowns like this:

$$\begin{aligned} \Delta p^{(i)} &= \begin{bmatrix} e_1^{(i)} & e_2^{(i)} & e_3^{(i)} & 1 \end{bmatrix} \begin{bmatrix} -\Delta X \\ -\Delta Y \\ -\Delta Z \\ c\Delta T \end{bmatrix} = \\ &= \begin{bmatrix} \cos\alpha^i \cos\eta^i & \sin\alpha^i \cos\eta^i & \sin\eta^i & 1 \end{bmatrix} \begin{bmatrix} -\Delta X \\ -\Delta Y \\ -\Delta Z \\ c\Delta T \end{bmatrix}, \end{aligned}$$

in which α^i and η^i are the azimuth and elevation of satellite i in the local sky. This way of writing is called *linearisation*.

This equation may be understood as an *observation equation*. If the equation is written symbolically, as is the practice in geodesy, in the form¹⁹

$$\underline{\ell} + \underline{v} = A\hat{\mathbf{x}}, \quad (12.8)$$

then the elements of the vector of unknowns $\hat{\mathbf{x}}$, the estimators of the unknowns, are $-\Delta\hat{X}$, $-\Delta\hat{Y}$, $-\Delta\hat{Z}$, and $c\Delta\hat{T}$, the vector of observations $\underline{\ell}$ is made up of values $\Delta p^{(i)}$, and the *design matrix* is

$$A = \begin{bmatrix} \cos\alpha^1 \cos\eta^1 & \sin\alpha^1 \cos\eta^1 & \sin\eta^1 & 1 \\ \cos\alpha^2 \cos\eta^2 & \sin\alpha^2 \cos\eta^2 & \sin\eta^2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \cos\alpha^i \cos\eta^i & \sin\alpha^i \cos\eta^i & \sin\eta^i & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \cos\alpha^n \cos\eta^n & \sin\alpha^n \cos\eta^n & \sin\eta^n & 1 \end{bmatrix}. \quad (12.9)$$

This is in fact the linearised version of the original observation equation 12.1.

This design matrix contains *everything we know* about the GPS measurement geometry.²⁰ From this, all DOP quantities may be calculated,²⁰

¹⁹The residuals \underline{v} are needed to reconcile the observations $\underline{\ell}$, which contain measurement uncertainty, with each other when there are more observations than unknowns. See section 14.4.

²⁰As does also a *sky plot*, for example Borre (2009).





TABLE 12.6. Variants of the DOP quantity.

Acronym	Name	Quantity characterised
GDOP	Geometric DOP	Place and time
PDOP	Position DOP	Place
HDOP	Horizontal DOP	Horizontal location
VDOP	Vertical DOP	Height
TDOP	Time DOP	Time

without using a single real observation — it suffices that the places of the satellites in the sky can be computed. The size of the matrix is $n \times 4$: n rows and four columns, where n is the number of satellites available for use.

rekognosointi

The situation is the same as in the case of reconnaissance of a terrestrial geodetic network: the quality of the network can already be judged based on point locations and planned measurement geometry, before even a single measurement has been carried out. This is a great tool for planning.

We will discuss more about observation equations and least-squares adjustment in section 14.4. Here we do not even try to compute a least-squares solution. We only look into the *precision* of the four unknowns $-\Delta X$, $-\Delta Y$, $-\Delta Z$, and $c \Delta T$ to be computed!

We assume, for this computation, that *all observations are equally precise* — their precision may be assumed 1 — and that *they are statistically independent of each other*. Then, the following simple calculation is valid. It gives a picture of the role of the GPS measurement geometry in the final precision of the measurement results. Other factors, like the technical capability of the receiver and antenna used, duration of measurement and the atmosphere, can be looked at separately.



12.9.2 Error ellipsoids for presentation of measurement precision

From the design matrix A we may construct the *normal matrix*, or *weight matrix of the unknowns*, as follows:

$$N = P_{xx} \stackrel{\text{def}}{=} A^T A. \quad (12.10)$$

Expressed in satellite positions in the sky, given as azimuths α^i and elevation angles η^i , the result is according to equation 12.11. The summation \sum is understood to be over the azimuths and elevations all satellites, $\alpha^i, \eta^i, i = 1, \dots, n$.



$$P_{xx} = \begin{bmatrix} \sum \cos^2 \alpha \cos^2 \eta & \sum \sin \alpha \cos \alpha \cos^2 \eta & \sum \cos \alpha \sin \eta \cos \eta & \sum \cos \alpha \cos \eta \\ \sum \sin \alpha \cos \alpha \cos^2 \eta & \sum \sin^2 \alpha \cos^2 \eta & \sum \sin \alpha \sin \eta \cos \eta & \sum \sin \alpha \cos \eta \\ \sum \cos \alpha \sin \eta \cos \eta & \sum \sin \alpha \sin \eta \cos \eta & \sum \sin^2 \eta & \sum \sin \eta \\ \sum \cos \alpha \cos \eta & \sum \sin \alpha \cos \eta & \sum \sin \eta & n \end{bmatrix}. \quad (12.11)$$

The inverse of this weight matrix P_{xx} , $Q_{xx} \stackrel{\text{def}}{=} P_{xx}^{-1}$, is the *weight-coefficient matrix*:

$$Q_{xx} = \begin{bmatrix} q_{xx} & q_{xy} & q_{xz} & q_{xt} \\ q_{yx} & q_{yy} & q_{yz} & q_{yt} \\ q_{zx} & q_{zy} & q_{zz} & q_{zt} \\ q_{tx} & q_{ty} & q_{tz} & q_{tt} \end{bmatrix}.$$

This matrix, like the weight matrix P_{xx} or the design matrix A , still describes exclusively the *geometry of the measurement site and satellites*, and nothing else.

Now, the *variance matrix* of the vector of unknowns, or *solution*,

$$\hat{x} = \begin{bmatrix} -\Delta\hat{X} & -\Delta\hat{Y} & -\Delta\hat{Z} & c\Delta\hat{T} \end{bmatrix}^T$$

is

$$\Sigma_{xx} = \sigma^2 Q_{xx} = \sigma^2 \begin{bmatrix} q_{xx} & q_{xy} & q_{xz} & q_{xt} \\ q_{yx} & q_{yy} & q_{yz} & q_{yt} \\ q_{zx} & q_{zy} & q_{zz} & q_{zt} \\ \hline q_{tx} & q_{ty} & q_{tz} & q_{tt} \end{bmatrix}.$$

The constant σ^2 is called the (*a priori*) *variance of unit weight*. Its square root, the *mean error of unit weight* σ , is the mean error, assumed constant, of a single observable, meaning one pseudo-range.

painoyksikön
keskivirhe

The variance matrix of the co-ordinate solution is a 3×3 element sized submatrix Σ_{rr} of the variance matrix of the unknowns:

$$\Sigma_{rr} = \sigma^2 \begin{bmatrix} q_{xx} & q_{xy} & q_{xz} \\ q_{yx} & q_{yy} & q_{yz} \\ q_{zx} & q_{zy} & q_{zz} \end{bmatrix}. \quad (12.12)$$

The **DOP** quantities are calculated directly from the weight-coefficient matrix Q_{xx} :

$$\text{PDOP} = \sqrt{q_{xx} + q_{yy} + q_{zz}}, \quad \text{HDOP} = \sqrt{q_{xx} + q_{yy}}, \quad \text{VDOP} = \sqrt{q_{zz}}.$$



The time t comes along in the quantities **TDOP** and **GDOP**. Remember, however, that the unknown to be estimated is $c\widehat{\Delta T}$ rather than $\widehat{\Delta T}$, and q_{tt} corresponds to this quantity:

$$\text{TDOP} = \sqrt{q_{tt}}, \quad \text{GDOP} = \sqrt{q_{xx} + q_{yy} + q_{zz} + q_{tt}} = \sqrt{\text{PDOP}^2 + \text{TDOP}^2}.$$

The mean errors of the co-ordinates are obtained as the square roots of the diagonal elements of the variance matrix **12.12**:

$$\sigma_X = \sigma \sqrt{q_{xx}}, \quad \sigma_Y = \sigma \sqrt{q_{yy}}, \quad \sigma_Z = \sigma \sqrt{q_{zz}}.$$

Also

$$\sigma_{c\Delta T} = \sigma \sqrt{q_{tt}}.$$

The familiar point mean error in the plane is now related directly to the **HDOP** quantity:

$$\underbrace{\sigma_P}_{\text{precision}} \stackrel{\text{def}}{=} \sqrt{\sigma_X^2 + \sigma_Y^2} = \underbrace{\sigma}_{\text{instrumental technology, etc.}} \cdot \underbrace{\text{HDOP}}_{\text{geometry}}.$$

The co-ordinate variance matrix can be graphically represented by a three-dimensional *error ellipsoid*. The error ellipsoid around a measurement point visualises the uncertainty of the location based on the above definitions. A similar **DOP** ellipsoid is obtained by leaving off the constant σ^2 : it has the same shape as the error ellipsoid, but it has no metric size, as the elements of the matrix Q_{xx} , like those of the design matrix A , are dimensionless.

In the general case it is not so simple to calculate the parameters of the ellipsoid from the matrix elements. Let us look at a simpler special case, which nevertheless is practically relevant. If the measurement geometry is symmetric, meaning that the satellites and their elevation angles are evenly distributed by azimuth, around the horizon, it follows that the **DOP** ellipsoid will be oriented along the co-ordinate axes: $q_{xy} = q_{yx} = q_{yz} = 0$ (and $q_{xx} = q_{yy}$!), and the matrix is

$$\Sigma_{\mathbf{rr}} = \sigma^2 \begin{bmatrix} q_{xx} & & \\ & q_{yy} & \\ & & q_{zz} \end{bmatrix}.$$

In this special case, the axes of the **DOP** ellipsoid are pointing along the local co-ordinate axes, and the longest axis points in the vertical direction. In a practical measurement situation, the variance matrix



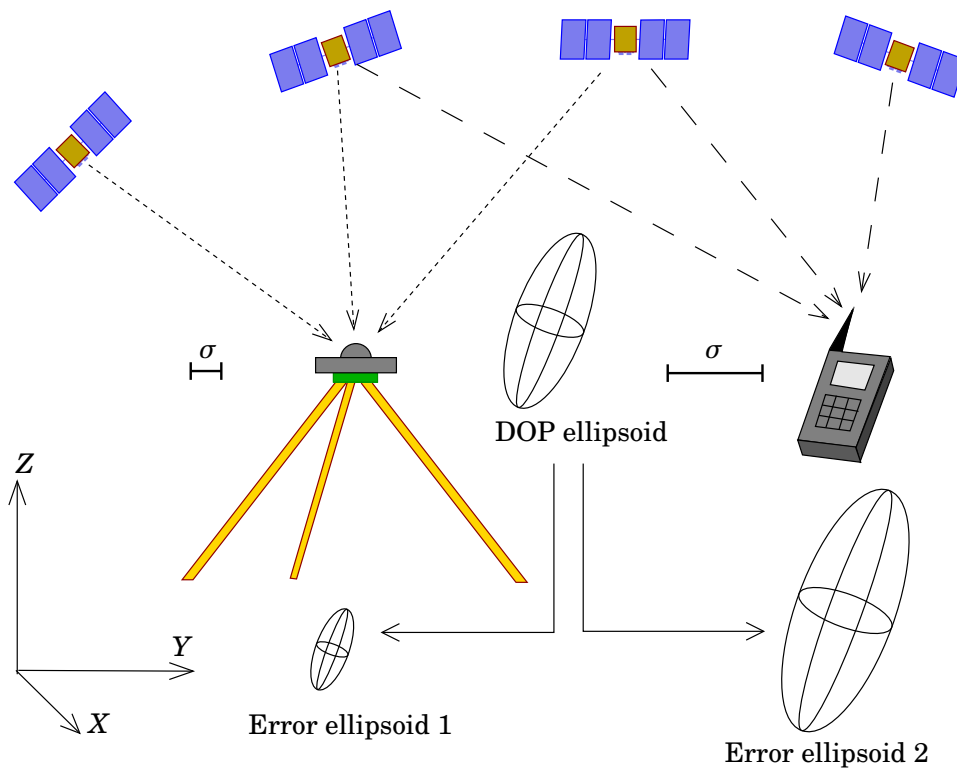


FIGURE 12.18. The connection between the DOP ellipsoid, and the error ellipsoid, and the mean error of unit weight σ . The DOP ellipsoid only characterises the effect of the geometry, whereas the error ellipsoid also depends on the precision of measurement, meaning the device type.

of the co-ordinates is often close to this. The longest axis of the error ellipsoid is almost always close to the vertical, which tells us that the height is more weakly determined than the horizontal location.²¹ In this ²¹

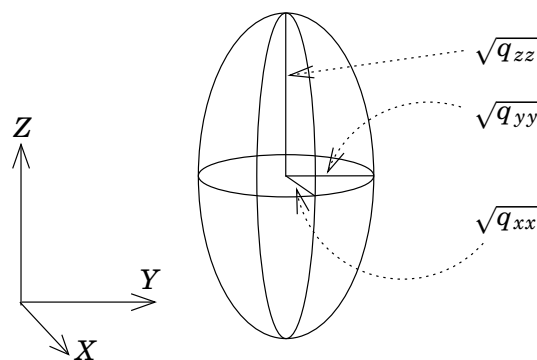


FIGURE 12.19. The DOP ellipsoid of GPS positioning, assuming its principal axes are in the same directions as the co-ordinate axes.

²¹Reasons for this are:



case the co-ordinate mean errors are

$$\sigma_X = \sigma_Y = \sigma \cdot \frac{1}{2} \sqrt{2} \cdot \text{HDOP}, \quad \sigma_Z = \sigma \cdot \text{VDOP},$$

akselin puolikas based on the above definitions. The lengths of the semi-axes of the **DOP** ellipsoid are $\sqrt{q_{xx}}$, $\sqrt{q_{yy}}$ and $\sqrt{q_{zz}}$. In the symmetric case, we thus have

$$q_{xx} = q_{yy} = \frac{1}{2} \text{HDOP}^2$$

in the horizontal plane, and always

$$q_{zz} = \text{VDOP}^2$$

in the vertical direction.



12.9.3 DOP and measurement planning

In general it can be said that the *measurement geometry* is better if **HDOP** and **VDOP**, thus also **PDOP**, are smaller. This again requires that

- A sufficient number of satellites is above the horizon and observable from the measurement site.
- They are positioned evenly around the sky.

This of course works only if one has a sufficiently free view from the measurement site up to the sky. In practice, there will always be obstacles, which should be mapped on a *horizon plot* for planning the measurement. Many planning programmes let the user draw a horizon plot and take it into account when calculating **DOP**.

The situation improves if the receivers to be used can make use of both the **GPS** and **GLONASS** systems, meaning more satellites are seen, and good measurement geometry is more easily achieved.

Of course, the **GPS** positioning geometry is only one factor among others. Other factors affecting measurement precision include

-
- Only satellites above the horizon contribute to the determination of height, so this is *extrapolation*. In the horizontal plane again, there are satellites in the east and in the west, in the north and in the south, which contribute to the positioning: *interpolation*.
 - Moreover, because of the same non-symmetry, the vertical location unknown Z and the clock unknown ΔT “compete” for the same information when they are estimated together from the same observation data.
 - The uncertainty in the signal delays caused by the atmosphere also affects mostly in the vertical direction, whereas — also due to the horizontal stratification of the atmosphere — the situation in the horizontal plane is more symmetric.



- The capabilities of receiver and antenna.
- Local disturbances, like *multipath*: reflections off the ground and objects nearby.
- The activity of the Sun, ionospheric conditions.
- The measurement mode: static or kinematic, absolute (for example “precise point positioning”, PPP) or relative.
- In relative GPS measurement, the distance between measurement points or from the base station. Geodetic measurements are nearly always relative and are carried out as network measurements.
- In static GPS measurement, the duration of measurement, the number of measurement epochs. Geodetic base-network measurements are always static, despite the method being time-consuming, because of its robustness. Only in local measurements, like detail surveys, is the faster kinematic technique (RTK, real-time kinematic) used.

runkomittaus
kartoitus-
mittaus

 12.9.4 Example 1: an azimuthally symmetric geometry

Assume that the satellites are evenly distributed around the sky, according to azimuth α , for every elevation angle η . Then, in the above weight matrix 12.11:

- Almost all non-diagonal elements vanish, because they contain either $\sin \alpha$ or $\cos \alpha$ or even $\sin \alpha \cos \alpha$. Only $\sum \sin \eta$ does not vanish.
-

$$\sum_{i=1}^n \cos^2 \alpha^i \cos^2 \eta^i = \frac{1}{2} \sum_{i=1}^n \cos^2 \eta^i, \quad \sum_{i=1}^n \sin^2 \alpha^i \cos^2 \eta^i = \frac{1}{2} \sum_{i=1}^n \cos^2 \eta^i.$$

Therefore, the weight or normal matrix $P_{xx} = A^T A$ becomes an almost diagonal matrix that would be relatively easy to invert — although we will not even try:

$$P_{xx} = A^T A = \begin{bmatrix} \frac{1}{2} \sum_{i=1}^n \cos^2 \eta^i & 0 & 0 & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^n \cos^2 \eta^i & 0 & 0 \\ 0 & 0 & \sum_{i=1}^n \sin^2 \eta^i & \sum_{i=1}^n \sin \eta^i \\ 0 & 0 & \sum_{i=1}^n \sin \eta^i & n \end{bmatrix}.$$



Transform now observation equation 12.8 and design matrix 12.9 in the following way:

$$\underline{\ell} + \underline{v} = A\Lambda\Lambda^{-1}\widehat{\underline{x}} = \widetilde{A}\widetilde{\underline{x}},$$

in which

$$\widetilde{\underline{x}} = \begin{bmatrix} -\Delta\widehat{X} \\ -\Delta\widehat{Y} \\ -\Delta\widehat{Z} \\ c\Delta\widehat{T} \end{bmatrix} = \Lambda^{-1}\widehat{\underline{x}} = \begin{bmatrix} -\Delta\widehat{X} \\ -\Delta\widehat{Y} \\ -\Delta\widehat{Z} \\ \left(\frac{1}{n}\sum\sin\eta\right)(-\Delta\widehat{Z}) + c\Delta\widehat{T} \end{bmatrix} \quad (12.13)$$

and

$$\widetilde{A} = A\Lambda = \begin{bmatrix} \cos\alpha^1\cos\eta^1 & \sin\alpha^1\cos\eta^1 & \sin\eta^1 - \frac{1}{n}\sum\sin\eta & 1 \\ \cos\alpha^2\cos\eta^2 & \sin\alpha^2\cos\eta^2 & \sin\eta^2 - \frac{1}{n}\sum\sin\eta & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \cos\alpha^i\cos\eta^i & \sin\alpha^i\cos\eta^i & \sin\eta^i - \frac{1}{n}\sum\sin\eta & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \cos\alpha^n\cos\eta^n & \sin\alpha^n\cos\eta^n & \sin\eta^n - \frac{1}{n}\sum\sin\eta & 1 \end{bmatrix},$$

²² because²²

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{n}\sum\sin\eta & 1 \end{bmatrix}, \quad \Lambda^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & +\frac{1}{n}\sum\sin\eta & 1 \end{bmatrix}.$$

Now we get a clean diagonal matrix:

$$\begin{aligned} \widetilde{P}_{xx} &= \widetilde{A}^T\widetilde{A} = \\ &= \begin{bmatrix} \frac{1}{2}\sum_{i=1}^n\cos^2\eta^i & 0 & 0 & 0 \\ 0 & \frac{1}{2}\sum_{i=1}^n\cos^2\eta^i & 0 & 0 \\ 0 & 0 & \sum_{i=1}^n\left(\sin\eta^i - \frac{1}{n}\sum_{j=1}^n\sin\eta^j\right)^2 & 0 \\ 0 & 0 & 0 & n \end{bmatrix}. \end{aligned}$$

Let us write the equation for the error or visual ellipsoid of \widetilde{P}_{xx} :

$$\widehat{\underline{x}}^T\widetilde{P}_{xx}\widehat{\underline{x}} = 1,$$

²²Verify that $\Lambda\Lambda^{-1} = \Lambda^{-1}\Lambda = I$!



in which \tilde{x} is as in equation 12.13. The result is

$$p_{11} \Delta \hat{X}^2 + p_{22} \Delta \hat{Y}^2 + \tilde{p}_{33} \Delta \hat{Z}^2 + p_{44} \Delta \hat{T}^2 = \frac{\Delta \hat{X}^2}{q_{xx}} + \frac{\Delta \hat{Y}^2}{q_{yy}} + \frac{\Delta \hat{Z}^2}{\tilde{q}_{zz}} + \frac{\Delta \hat{T}^2}{\tilde{q}_{tt}} = 1,$$

in which are immediately seen the elements of the variance matrix of the unknowns

$$Q_{xx} = P_{xx}^{-1} = (A^T A)^{-1}$$

according to the definition:

$$q_{xx} = q_{yy} = \frac{2}{\sum_i \cos^2 \eta^i},$$

from which

$$\text{HDOP} = \sqrt{q_{xx} + q_{yy}} = \frac{2}{\sqrt{\sum_i \cos^2 \eta^i}}.$$

Similarly, after a little reorganising,²³

23

$$\tilde{q}_{zz} = \frac{1}{\tilde{p}_{33}} = \frac{1}{\sum_i \left(\sin \eta^i - \frac{1}{n} \sum_j \sin \eta^j \right)^2} = \frac{n}{n \sum_i \sin^2 \eta^i - \left(\sum_i \sin \eta^i \right)^2},$$

and VDOP is its square root.²⁴

24

12.9.5 Example 2: a singular case

Look again at design matrix A , expression 12.9, and write it into the form

$$A = \begin{bmatrix} e_1^{(1)} & e_2^{(1)} & e_3^{(1)} & 1 \\ e_1^{(2)} & e_2^{(2)} & e_3^{(2)} & 1 \\ \vdots & & & \vdots \\ e_1^{(n)} & e_2^{(n)} & e_3^{(n)} & 1 \end{bmatrix} = \begin{bmatrix} (\bar{e}^{(1)})^T & 1 \\ (\bar{e}^{(2)})^T & 1 \\ \vdots & \vdots \\ (\bar{e}^{(n)})^T & 1 \end{bmatrix},$$

²³Because

$$\begin{aligned} \sum_i \left(\sin \eta^i - \frac{1}{n} \sum_j \sin \eta^j \right)^2 &= \sum_i \sin^2 \eta^i - \sum_i \frac{2}{n} \sin \eta^i \overbrace{\left(\sum_j \sin \eta^j \right)}^{\text{constant}} + \sum_{i=1}^n \frac{1}{n^2} \overbrace{\left(\sum_j \sin \eta^j \right)^2}^{\text{constant}} \\ &= \sum_i \sin^2 \eta^i - \frac{2}{n} \left(\sum_i \sin \eta^i \right) \left(\sum_i \sin \eta^i \right) + \frac{1}{n} \left(\sum_i \sin \eta^i \right)^2 \\ &= \sum_i \sin^2 \eta^i - \frac{1}{n} \left(\sum_i \sin \eta^i \right)^2. \end{aligned}$$

²⁴If η is constant — so, all $\eta^i, i = 1, \dots, n$ are the same — then the denominator vanishes! So, the determination of height by GPS requires that there are satellites on different elevations in the sky.



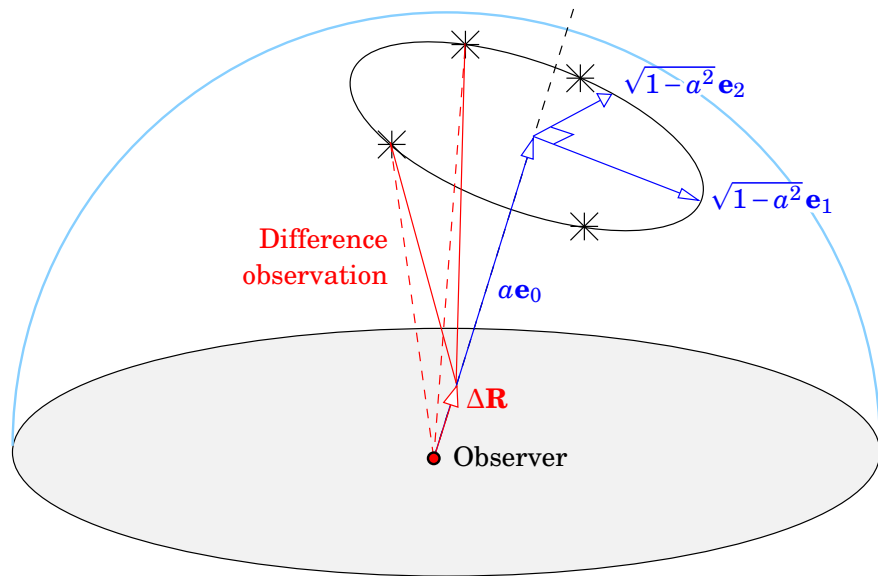


FIGURE 12.20. The circle singularity or “dangerous circle” for GPS.

assuming there are n satellites. If the satellites are all on the same circle, then the unit direction vector in the direction of satellite i is

$$\mathbf{e}^{(i)} = e_1^{(i)} \mathbf{i} + e_2^{(i)} \mathbf{j} + e_3^{(i)} \mathbf{k} = a \mathbf{e}_0 + b^{(i)} \mathbf{e}_1 + c^{(i)} \mathbf{e}_2$$

in which the values $b^{(i)}$ and $c^{(i)}$ satisfy the condition $(b^{(i)})^2 + (c^{(i)})^2 = 1 - a^2$, for all satellites $i = 1, \dots, n$. Here, $a \mathbf{e}_0$ is the vector from the observer to the centre of the circle. Thus there are only three independent vectors $\mathbf{e}^{(i)}$ when four are needed.

See figure 12.20. The situation is also geometrically clear: if the observation site is shifted along the direction of the vector \mathbf{e}_0 , any difference between the pseudo-ranges from two different satellites will remain unchanged. What is happening here is that the receiver’s clock unknown ΔT and the component of the observation site’s location in the \mathbf{e}_0 direction (in other words the *projection* onto the \mathbf{e}_0 direction) *cannot be separated from each other* in this geometry.

taaksepäin
leikkaus

This is the circle singularity or “dangerous circle” in the case of GPS positioning: compare section 6.2. GPS positioning is in fact a three-dimensional resection!



12.9.6 Calculation example for DOP quantities

See figure 12.21.

Let one satellite be in the zenith ($\eta = 90^\circ$) and three satellites at an elevation angle $\eta = 30^\circ$ at azimuths $\alpha = 0^\circ, 120^\circ, 240^\circ$. Compute first



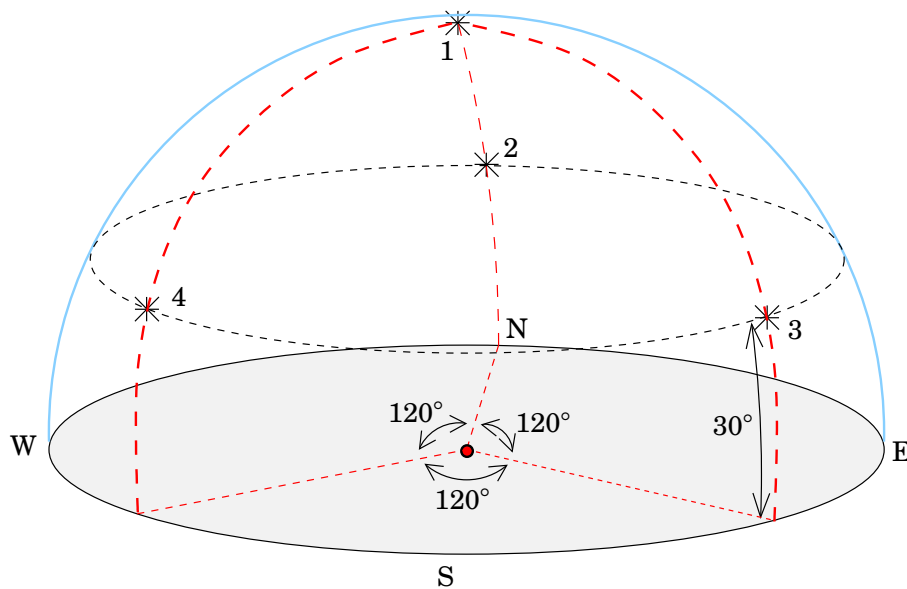


FIGURE 12.21. Calculation example of DOP quantities.

design matrix A according to equation 12.9. It is given that

$$\begin{aligned} \alpha_1 &= 0^\circ, & \eta_1 &= 90^\circ, \\ \alpha_2 &= 0^\circ, & \eta_2 &= 30^\circ, \\ \alpha_3 &= 120^\circ, & \eta_3 &= 30^\circ, \\ \alpha_4 &= -120^\circ, & \eta_4 &= 30^\circ. \end{aligned}$$

Numerical values are obtained by remembering that

$$\begin{aligned} \sin(90^\circ) &= 1, & \cos(90^\circ) &= 0, \\ \sin(30^\circ) &= \frac{1}{2}, & \cos(30^\circ) &= \frac{1}{2}\sqrt{3}, \\ \sin(120^\circ) &= -\sin(-120^\circ) = \frac{1}{2}\sqrt{3}, & \cos(120^\circ) &= \cos(-120^\circ) = -\frac{1}{2}. \end{aligned}$$

The result is

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ \frac{1}{2}\sqrt{3} & 0 & \frac{1}{2} & 1 \\ -\frac{1}{2} \cdot \frac{1}{2}\sqrt{3} & (\frac{1}{2}\sqrt{3})^2 & \frac{1}{2} & 1 \\ -\frac{1}{2} \cdot \frac{1}{2}\sqrt{3} & -(\frac{1}{2}\sqrt{3})^2 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ \frac{1}{2}\sqrt{3} & 0 & \frac{1}{2} & 1 \\ -\frac{1}{4}\sqrt{3} & \frac{3}{4} & \frac{1}{2} & 1 \\ -\frac{1}{4}\sqrt{3} & -\frac{3}{4} & \frac{1}{2} & 1 \end{bmatrix}.$$

Next we compute the weight matrix of the unknowns, or normal matrix, equation 12.10:

$$P_{xx} = A^T A = \begin{bmatrix} \frac{9}{8} & 0 & 0 & 0 \\ 0 & \frac{9}{8} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{3}{2} \\ 0 & 0 & \frac{3}{2} & 4 \end{bmatrix}.$$



Inverting this matrix would yield Q_{xx} , the weight-coefficient matrix of the unknowns. Here, we invert the matrix *only partially*:

$$Q_{xx} = P_{xx}^{-1} = \begin{bmatrix} \frac{8}{9} & 0 & 0 & 0 \\ 0 & \frac{8}{9} & 0 & 0 \\ 0 & 0 & \begin{bmatrix} \frac{3}{4} & \frac{3}{2} \\ \frac{3}{2} & 4 \end{bmatrix}^{-1} \\ 0 & 0 & \begin{bmatrix} \frac{3}{2} & 4 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} q_{xx} & & & \\ & q_{yy} & & \\ & & q_{zz} & q_{zt} \\ & & q_{tz} & q_{tt} \end{bmatrix}. \quad (12.14)$$

From this we read directly that the weight coefficients of the co-ordinates X and Y are

$$q_{xx} = q_{yy} = \frac{8}{9} = 0.888\dots,$$

and thus

$$\text{HDOP} = \sqrt{q_{xx} + q_{yy}} = \sqrt{\frac{16}{9}} = \frac{4}{3} = 1.333\dots$$

From equation 12.14 we see that the \hat{Z} co-ordinate and the clock unknown $c\hat{\Delta T}$ are “entangled” with each other ($q_{zt} \neq 0$) and calculating their weight coefficients is not attempted here.²⁵



12.10 Orbits of the GPS satellites

The orbit of a GPS satellite in the Earth’s gravitational field is approximately an ellipse satisfying Kepler’s laws. In practice, it is almost a circle, the radius²⁶ of which is 26560 km and the orbital period 11^h58^m. The tilt angle of the orbital planes with respect to the equator, the *inclination*, is $i = 55^\circ$, meaning that at the latitude of Finland, the GPS satellites will

²⁵In fact, the sub-matrix to be inverted is

$$\begin{bmatrix} \frac{3}{4} & \frac{3}{2} \\ \frac{3}{2} & 4 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{16}{3} & -2 \\ -2 & 1 \end{bmatrix},$$

and thus

$$Q_{xx} = \begin{bmatrix} \frac{8}{9} & & & \\ & \frac{8}{9} & & \\ & & \frac{16}{3} & -2 \\ & & -2 & 1 \end{bmatrix}.$$

This is now the weight-coefficient matrix (and, up to the variance of unit weight, the variance matrix) of the vector of unknowns $[-\hat{X} \quad -\hat{Y} \quad -\hat{Z} \quad c\hat{\Delta T}]^T$.

In this result is also seen, how \hat{Z} and $\hat{\Delta T}$ “compete” for the same information: $\text{VDOP} = \sqrt{q_{zz}} = \sqrt{\frac{16}{3}} \approx 2.309$, when, without the clock unknown, it would be $\text{VDOP} = \sqrt{\frac{4}{3}} \approx 1.155$.

²⁶So: the distance from the Earth’s surface is about 26560 km – 6378 km = 20182 km, using 6378 km for the Earth’s radius.



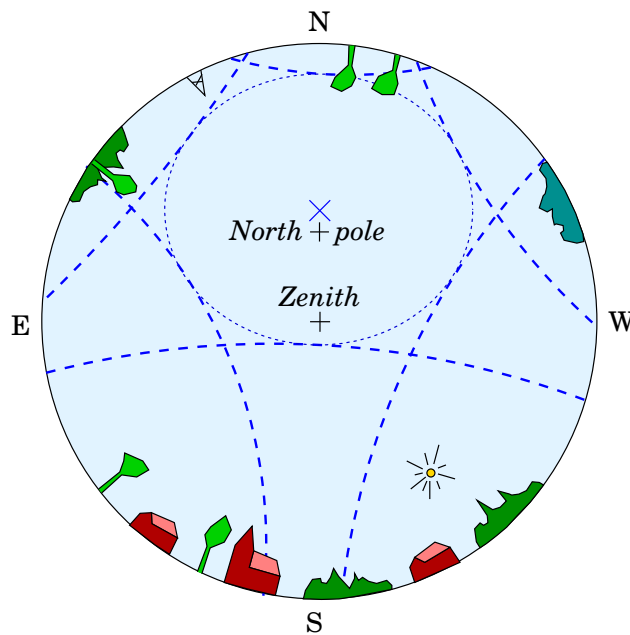


FIGURE 12.22. The six orbital planes of GPS satellites in the Helsinki sky. GPS satellites will never be found inside the oval around the zenith, although they are also visible in the northern sky, though very close to the horizon.

never pass through the zenith. However, due to their great height, the satellites are also visible “over the North Pole” in the northern half of the sky, albeit very low. See figure 12.22.

We need *six orbital elements* to describe a satellite orbit. As orbital elements we may choose, for example, the three co-ordinates of position

$$\mathbf{r}(t_0) = x(t_0) \cdot \mathbf{i} + y(t_0) \cdot \mathbf{j} + z(t_0) \cdot \mathbf{k}$$

and the three components of velocity

$$\left. \frac{d\mathbf{r}}{dt} \right|_{t=t_0} = \dot{\mathbf{r}}(t_0) = \dot{x}(t_0) \cdot \mathbf{i} + \dot{y}(t_0) \cdot \mathbf{j} + \dot{z}(t_0) \cdot \mathbf{k}$$

at a certain time t_0 , using Newton’s dot notation for the time derivative.²⁷ ²⁷

The vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ form an orthonormal basis. Position and velocity,

$$\mathbf{r}(t) = x(t) \cdot \mathbf{i} + y(t) \cdot \mathbf{j} + z(t) \cdot \mathbf{k}, \quad \dot{\mathbf{r}}(t) = \dot{x}(t) \cdot \mathbf{i} + \dot{y}(t) \cdot \mathbf{j} + \dot{z}(t) \cdot \mathbf{k},$$

can be calculated from these equations for some later moment t , by just calculating, step by small step, forward in time, correcting both the

²⁷This dot notation for the derivative of time, fluxion, was introduced by Newton in 1665.



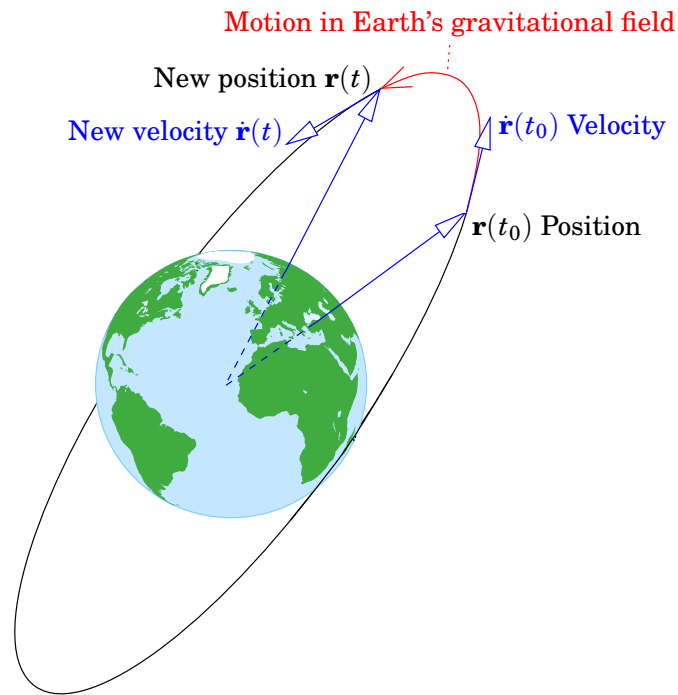


FIGURE 12.23. Satellite orbital motion described by position and velocity vectors.

velocity using the gravitation equation and the position using the velocity, figure 12.23. We know the attraction field of the Earth as an equation: the acceleration caused by the attraction is computable when we know the position in space of the satellite.

The geometry of a satellite orbit is normally described using the *six Kepler orbital elements*,²⁸ Ω , i , ω , a , e , and v , see figure 17.14. More details are given in Poutanen (2017) section 5.1, Hofmann-Wellenhof et al. (2001) subsection 4.2.1, and in section 17.7. There is a one-on-one correspondence between the Kepler elements and the position and velocity vector representation described above:

$$\text{Kepler: } \{\Omega, i, \omega, a, e, v\} \longleftrightarrow \{\mathbf{r}, \dot{\mathbf{r}}\}.$$

This means that, from the given Kepler elements, we may *calculate* the *position* of the satellite in space, as well as its *velocity*. All GPS computation software packages know how to do this.

²⁸So every satellite has six Kepler orbital elements that describe the shape, size and orientation of the orbit of that satellite, as well as the location of the satellite in its orbit, in space.





12.10.1 Navigation message transmitted by the satellites

As already noted, all GPS satellites transmit a navigation message modulated upon the carrier wave of their radio signal. The modulation frequency of the navigation message is 50 Hz, so every second contains 50 bits. The whole navigation message consists of 25 “packets” (*frames*), each of which contains 1500 bits and transmission of which lasts 30 seconds. Thus, the total length is 37500 bits, and the duration of the transmission is 12.5 minutes.

When a GPS receiver is switched on for the first time, the search for satellites starts. Immediately when the first satellite is “caught” (*lock-on*), the reading of the navigation message starts. Lock-on may easily last for several minutes, especially if the approximate location fed to the receiver is seriously wrong, or the instrument has been transported between continents. After that, however, finding the other satellites generally proceeds apace.

The navigation message is regularly uploaded to the satellites by the control segment. After that, the information is, as part of the signal of the GPS satellites, available to all users of the GPS. The navigation message consists of three parts.

- Information related to time keeping, like the clock corrections of the satellites, the “health” information on the satellites, the quality of the positioning signal and orbital data transmitted by the satellite, and the freshness of the navigation message.
- The satellite’s orbital information (broadcast ephemeris), disseminated by the satellite by radio. These ephemeris are computed by the US military authorities and are based on observation data continuously produced by a global network of tracking stations. The orbital elements of all satellites computed from the observations are uploaded to the satellites by the control segment, typically once per 24 hours. They are then transmitted from the satellite’s memory, modulated onto the radio signal, as a bit stream to all users.

The orbital elements are the Kepler elements augmented by coefficients used to calculate the perturbations caused by the Earth’s flattening, both secular (growing linearly with time, in the mean motion, orbital inclination²⁹ and right ascension of the ascending node) and short-period (half the satellite orbital period), in the or-

keskiliike

29

rektaskensio

²⁹Not actually secular: rather, long-period.



nouseva solmu

bital inclination, the radius and the angle between ascending node and satellite (the argument of latitude). There are nine coefficients, which for GPS satellites need to be taken into account. The origin of the theory used is the classical article by Yoshihide Kozai (Kozai, 1959).

Broadcast ephemeris are used in navigation applications and in real-time positioning. It is also practical to use them in GPS surveying and relative positioning in relatively small areas.

From the ephemeris, every satellite's position in space at the moment of observation is computed, so that they may be used as "beacons" for the determination of the location of the ground station. From the ephemeris, also the velocity of the satellite is computed.³⁰ More is said about the computing methods used in chapter 5 of Poutanen (2017) and in chapter 4 of Hofmann-Wellenhof et al. (2001).

- 31 ○ The *almanac* for all satellites.³¹ The purpose of the almanac is to provide approximate orbital elements for all satellites, sufficient for planning measurement campaigns and helping the receiver find satellites. The almanac is valid for many weeks. The almanac also contains a crude global ionosphere model.



12.10.2 Precise ephemeris

Precise ephemeris, precise orbital elements, may be obtained by the user some time afterwards directly from the Internet. They are computed and distributed by a service called the International GNSS Service (IGS), see section 12.11. A standard format, SP3 ("Standard Product 3") is used for data distribution, originally designed by the US National Geodetic Survey. See tableau 12.7.

The SP3 ephemeris file contains the orbital data in the form of three-dimensional component vectors of position and velocity,

$$\begin{bmatrix} x(t_i) \\ y(t_i) \\ z(t_i) \end{bmatrix}, \begin{bmatrix} \dot{x}(t_i) \\ \dot{y}(t_i) \\ \dot{z}(t_i) \end{bmatrix},$$

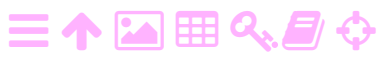
³⁰Knowing the velocity of the satellite would not be necessary for this, but is needed to calculate the *Doppler shift* of the signal frequency. The receiver must know the Doppler shift of every satellite in order to lock on, and remain locked on, to the satellite signal.

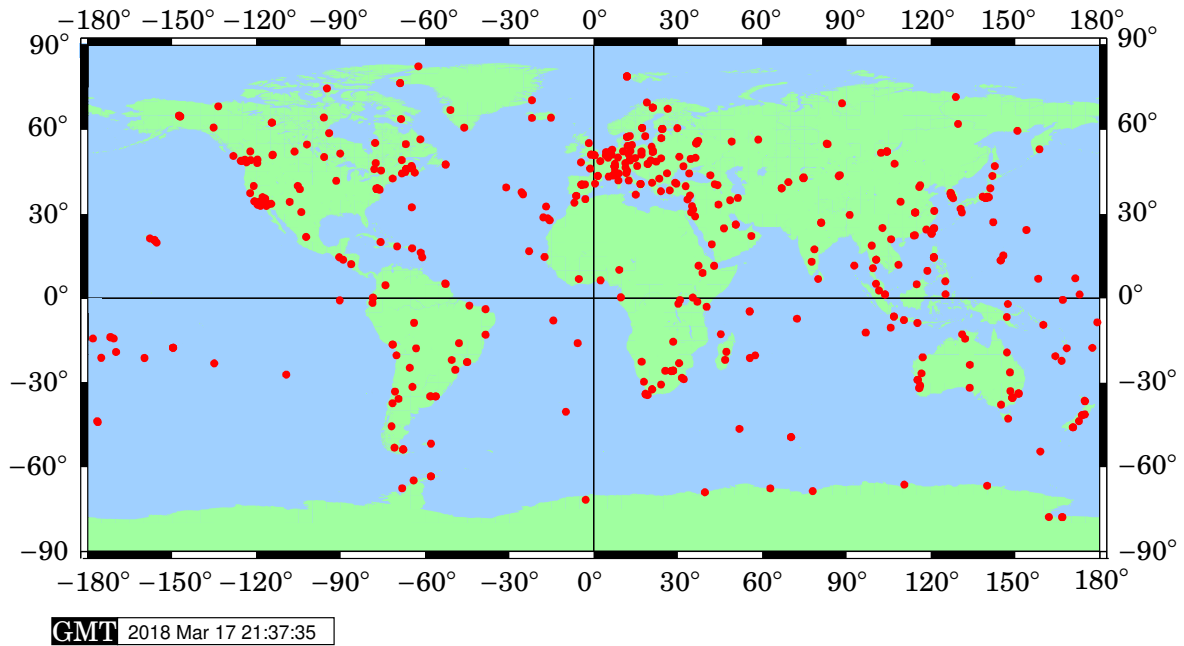
³¹*Almanac* is apparently not a word of Arabic origin, although it looks like it is.




TABLEAU 12.7. Precise ephemeris in the original SP3 format. Satellite numbers, position vectors, velocity vectors, clock correction and clock drift, date and time, etc. Start of example file © US National Geodetic Survey.

```
#aV1994 12 17 0 0 0.00000000 96 d ITR92 FIT NGS
## 779 518400.00000000 900.00000000 49703 0.00000000000000
+ 25 1 2 4 5 6 7 9 12 14 15 16 17 18 19 20 21 22
+ 23 24 25 26 27 28 29 31 0 0 0 0 0 0 0 0 0
+ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
+ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
++ 7 6 5 5 5 5 5 5 5 6 5 5 5 5 6 5 5
++ 5 5 6 5 5 5 5 5 0 0 0 0 0 0 0 0 0
++ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
++ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
++ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
%C cc cc ccc ccc cccc cccc cccc cccc ccccc ccccc ccccc ccccc
%C cc cc ccc ccc cccc cccc cccc cccc ccccc ccccc ccccc ccccc
%f 0.0000000 0.000000000 0.00000000000 0.000000000000000
%f 0.0000000 0.000000000 0.00000000000 0.000000000000000
%i 0 0 0 0 0 0 0 0 0 0
%i 0 0 0 0 0 0 0 0 0 0
/* CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
/* CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
/* CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
/* CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
* 1994 12 17 0 0 0.00000000
P 1 16258.524750 -3529.015750 -20611.427050 -62.540600
V 1 -6560.373522 25605.954994 -9460.427179 -0.024236
P 2 -21998.652100 -8922.093550 -12229.824050 -131.326200
V 2 -9852.750736 -12435.176313 25738.634180 -0.029422
P 4 -26019.547600 4809.810900 -2508.578200 3.544600
V 4 2559.038002 -3340.527442 -31621.490838 0.016744
*
*
*
P 29 -1638.431050 -24391.479200 10455.312650 3.690300
V 29 5754.005457 -12065.761570 -27707.056273 0.003537
P 31 6265.255800 -25687.986950 -753.359000 70.830800
V 31 3053.344058 -63.091750 31910.454757 0.033749
* 1994 12 17 0 15 0.00000000
P 1 15716.820135 -1169.850490 -21281.578766 -62.542746
V 1 -5439.955846 26738.341429 -5409.793390 -0.023226
P 2 -22813.261065 -9927.616864 -9816.490189 -131.328686
V 2 -8178.974330 -9924.329320 27813.754308 -0.025238
*
*
```





 FIGURE 12.24. The tracking stations of the IGS, situation in 2018 (data © IGS).

for epochs t_i . The velocity vector is optional. The data is tabulated at time intervals of 15 minutes: $t_{i+1} - t_i = 15^m$. From these, the position $\mathbf{r}(t)$ and velocity $\dot{\mathbf{r}}(t)$ are interpolated to the moment of measurement t by Lagrange³² interpolation. In addition, the ephemeris contain information characterising the behaviour of the satellite clock and the accuracy of the orbital information, among other things.

The current version of the format is SP3-c, which also allows the distribution of orbital information on GLONASS satellites. It is a textual format and human-readable.

The best-known source from as early as 1992 has been the International GNSS Service. The precise orbits produced by them are published on the Internet a couple of weeks after the time of observation.

In addition to precise ephemeris, “rapid orbits” are also being produced. These are almost as precise as precise ephemeris, but are turned out faster.

The newest “ultra-rapid” solutions are satellite orbital predictions twenty-four hours into the future, which can thus be used in real-time applications.

³²Joseph-Louis Lagrange (1736–1813) was a French mathematician, astronomer, developer of classical mechanics, one of the 72 names on the Eiffel Tower, Eiffel Tower, 72 names.

Precise ephemeris are, unlike broadcast ephemeris, disseminated over the Internet and not via GPS satellites. The organisations producing them are independent from the United States military authorities.

Unlike broadcast ephemeris, precise ephemeris are very close to the *true* orbits of the satellites, where they actually *were* at that moment. Broadcast ephemeris are *predictions* and therefore less accurate.



12.11 The International GNSS Service IGS

The IAG (International Association of Geodesy) established the International GNSS Service (IGS) in 1990, and it became an official service of the IAG in 1994. The main purpose of the service is to produce precise orbital ephemeris in support of *geodynamics* research, like the study of the motions of the solid Earth. However, its products are used much more broadly, in many fields of geophysics.

The activities of the IGS are led by a Central Bureau, currently at the JPL (Jet Propulsion Laboratory) in the United States. In 2015 the IGS used observations made globally by some 500 GNSS stations to compute its orbital ephemeris. The number has grown only slowly over recent years.

The computation as such is carried out by seven different computing centres, and the orbital data is available for use a couple of weeks after the time of measurement. The computed data also include the clock correction parameters for the satellites. The Earth's orientation parameters (EOP) polar motion and variations in the length of the day (LoD) are published separately. See IGS Central Bureau.

vuorokauden
pituus



Self-test questions

1. Explain how a hyperbolic positioning system like Decca functions.
2. Why does GPS broadcast on two different carrier frequencies?
3. How does the densest part of the Earth's atmosphere, the troposphere, affect the propagation of GPS radio waves? Why are meteorologists interested in this?
4. How is it possible that all GPS satellites broadcast on the same frequencies? How does the receiver separate out the signals from the different satellites?





TABLEAU 12.8. DOP calculation script.

```

% Dilution of precision (DOP):
% Part 1. Run the program. What does the error message tell you?
% What can you say about the value of VDOP?
% Places of satellites, azimuth A and elevation h in the sky (below).
% Part 2. Change the elevation of the first satellite 30 -> 60 degrees.
% Run again. Note down the HDOP and VDOP values.
% Why is the run now successful?
A1 = 0; h1 = 30;
A2 = 90; h2 = 30;
A3 = 180; h3 = 30;
A4 = 270; h4 = 30;
conv = pi/180;
% Design matrix A (below)
% Part 3. Modify the program to include a fifth satellite, place
% in the sky A5 = 0, h5 = 45.
A = [cos(A1*conv)*cos(h1*conv), sin(A1*conv)*cos(h1*conv), sin(h1*conv), 1;
cos(A2*conv)*cos(h2*conv), sin(A2*conv)*cos(h2*conv), sin(h2*conv), 1;
cos(A3*conv)*cos(h3*conv), sin(A3*conv)*cos(h3*conv), sin(h3*conv), 1;
cos(A4*conv)*cos(h4*conv), sin(A4*conv)*cos(h4*conv), sin(h4*conv), 1];
N = A'*A;
Ninv = inv(N)
HDOP = sqrt(Ninv(1,1) + Ninv(2,2));
VDOP = sqrt(Ninv(3,3));
% Part 4. Add to the program the evaluation and output of PDOP.
% PDOP = Position dilution of precision. See lecture notes
% for definition.
fprintf(1, 'HDOP = %20.10f\n', HDOP);
fprintf(1, 'VDOP = %20.10f\n\n', VDOP);

% Part 5. Play around with the five satellite places in the sky,
% in order to minimise PDOP.

```

5. Describe how the pseudo-random codes modulated on the GPS carrier wave make it possible to measure pseudo-ranges. Where does correlation come in?
6. What is multipath, and what methods are available to minimise its influence?
7. Explain the concept of DOP, dilution of precision.
8. Why are geodesists interested in measuring the phase of the carrier wave of the GPS signal, even though it is harder than code measurement?
9. What are broadcast ephemeris, who generates them, and how does



the user acquire them?

10. What are precise ephemeris, who generates them, and how does the user acquire them?
11. What are in tableau 12.3 the units of the observables $L1$, $L2$, $P1$, $P2$, and $C1$? Do not look up the RINEX format definition!



Exercise 12–1: Calculation of DOP quantities

In this exercise you are going to write code to calculate the various DOP quantities of a GPS measurement geometry, when the places of the satellites in the sky, their azimuths α and elevation angles η , are given.

You may use your preferred rapid prototyping language: MATLAB™, GNU Octave, Scilab (Scilab Enterprises), R (The R Project for Statistical Computing), even Excel™. And do not bother with inputting from files, just put the satellite positions into the source.

1. For an arbitrary number of satellites, write, or adapt, code to build the design matrix A and the normal or weight matrix $P_{xx} = A^T A$.
One could also build, from the vector of observations $\underline{\ell}$, the right-hand side of the system of normal equations, $A^T \underline{\ell}$, but we *do not need it here*. The beauty of DOP is that we can use it in measurement planning, before any real observations are available.
2. Make your software output to the screen the various DOP quantities.
3. Now assume you have *five satellites*, one more than the minimum to make positioning possible. Play with the satellite positions (α_i, η_i) , $i = 1, \dots, 5$, in order to minimise one of the DOPs, for example the PDOP. What is the best geometry of five satellites you find?
4. You cheated, didn't you? You cannot observe GPS satellites that are below the horizon. So, introduce the constraint $\eta > 5^\circ$.
5. After PDOP, try to minimise HDOP, and VDOP.

What did you learn?





Processing GPS observations

13

[...] everything is related to everything else, but near things are more related than distant things. [...]

First law of geography, by Waldo R. Tobler (1970)



13.1 Forming difference observations

In geodetic GPS measurement and GPS surveying applications, commonly one wishes to measure the difference in location between two points, like a point pair in a local geodetic network. The distance between the points may be in the order of 100–1000km. This is a much shorter distance than that to the GPS satellites, which orbit at a height of some 20000km. See figure 13.1.

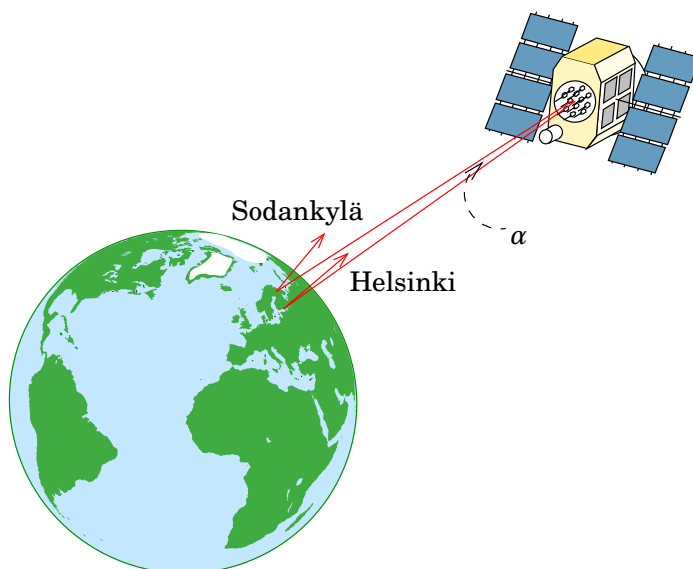


FIGURE 13.1. The “common-mode” error assumption.

kellopoikkeama

Seen from the satellite, the angle α separating the observation sites is very small, in the example case (Helsinki and Sodankylä) only 2° ! For this reason, many errors will be partly common to the two sites, similar and approximately equal in magnitude. The effect of the satellite clock error or *clock offset* is even identical at both observation sites. The effect of orbit error is, due to the geometry, roughly the same; the errors caused by ionosphere and troposphere are also similar, due both to the similarity in geometry and the long-range *spatial correlation*¹ of atmospheric conditions.

luotiviiva
korkeuskulma

On the other hand, however, one should remember that the *difference in directions of the plumb line* between Helsinki and Sodankylä is already 7° , so the difference in elevation angle of a satellite above the local horizon may amount to this much because of that alone.

Based on the “common-mode” error assumption, we form *differences* between observations from two sites to one satellite. In these differences, many errors vanish entirely or are materially reduced. Forming the difference is straightforward: subtract two simultaneously made raw observations from each other, each lifted from an observation file looking like tableau 12.3.

The differences can be *single* — either between two receivers or between two satellites, in which cases one can use the visually appropriate symbols Δ or ∇ — *double*, or *triple*, between successive measurement epochs, symbol δ . See figure 13.2.

The influence of forming the various difference types on the magnitude of errors — the interesting thing here! — has been catalogued in table 13.1.



13.1.1 Single differences

We explain with equations how a single difference is calculated from original observations. We shall see how some systematic errors are eliminated altogether while others are substantially reduced.

One receiver (observer) A, two satellites S, T

$$P_A^{ST} \stackrel{\text{def}}{=} P_A^T - P_A^S, \quad P_A^{ST} \stackrel{\text{def}}{=} P_A^T - P_A^S.$$

¹By this is meant that conditions change only slowly with place. Helsinki and Sodankylä lie practically in the same climate zone, and if there is a high or low pressure zone over Northern Europe, it will undoubtedly affect both places. The *synoptic scale* (Wikipedia, [Synoptic scale in meteorology](#)) of weather phenomena is of the order of 1000 km.



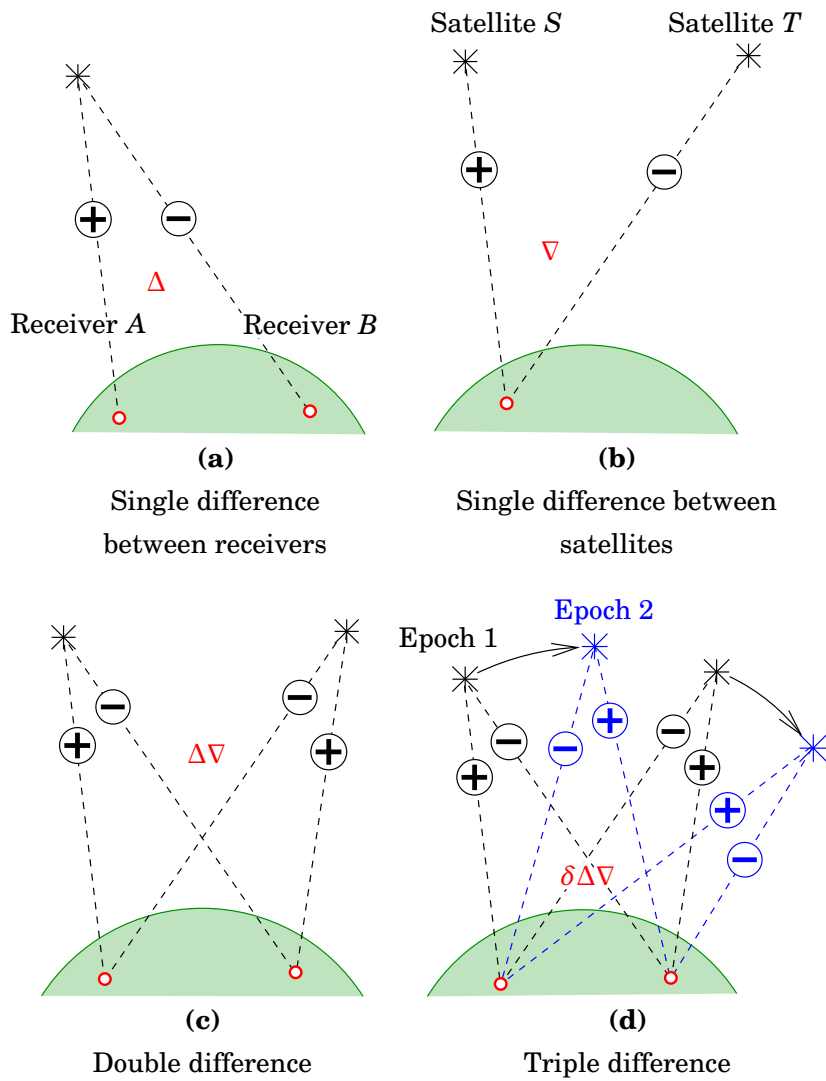


FIGURE 13.2. Forming difference observations of various types and symbols used for them.

Write the observation equation 12.1 for pseudo-range in this extended notation:

$$p_A^S = \rho_A^S + c (\Delta T_A - \Delta t^S) + d_{\text{ion},A}^S + d_{\text{trop},A}^S,$$

$$p_A^T = \rho_A^T + c (\Delta T_A - \Delta t^T) + d_{\text{ion},A}^T + d_{\text{trop},A}^T.$$

It has been taken into account that, of the clock offsets, Δt is *satellite-specific*, ΔT again is *observer-*, that is, *receiver-*, specific. Subtraction yields the difference quantity

$$p_A^{ST} = \rho_A^{ST} - c \Delta t^{ST} + d_{\text{ion},A}^{ST} + d_{\text{trop},A}^{ST},$$



TABLE 13.1. Effect of forming difference observations from GPS observations on the magnitude of various errors.

Error source	Type of difference			
	∇^{ST}	Δ_{AB}	$\Delta_{AB}\nabla^{ST}$	$\delta_{12}\Delta_{AB}\nabla^{ST}$
Satellite orbit $\{\mathbf{r}, \dot{\mathbf{r}}\}$	-	↓	↓	↓↓
Satellite clock Δt	-	0	0	0
Receiver clock ΔT	0	-	0	0
Ionosphere $d_{\text{ion}}, D_{\text{ion}}$	-	↓	↓	↓↓
Troposphere $d_{\text{trop}}, D_{\text{trop}}$	-	↓	↓	↓↓
Ambiguities N	-	-	-	0*

↓ The error is reduced substantially, especially for short distances between measurement points.

↓↓ The error is diminished even more strongly.

0 The error is completely eliminated.

0* The error is eliminated, unless there is a *cycle slip*.

in which the definitions apply

$$\begin{aligned}\rho_A^{ST} &\stackrel{\text{def}}{=} \rho_A^T - \rho_A^S, & d_{\text{ion},A}^{ST} &\stackrel{\text{def}}{=} d_{\text{ion},A}^T - d_{\text{ion},A}^S, \\ \Delta t^{ST} &\stackrel{\text{def}}{=} \Delta t^T - \Delta t^S, & d_{\text{trop},A}^{ST} &\stackrel{\text{def}}{=} d_{\text{trop},A}^T - d_{\text{trop},A}^S.\end{aligned}$$

Here, the clock offset of receiver A , ΔT_A , has vanished, because, being a receiver property, it is the same for different satellites and cancels out in calculating the difference observation between satellites.

käynti

This is important in practice, because receiver clocks are usually based on inexpensive quartz oscillators, the *drift* of which is so large that it needs to be taken into account.

kantaaallon
vaihe

A similar observation equation derived from equation 12.2 also applies for the raw carrier-phase observable:

$$\begin{aligned}P_A^{ST} &= \lambda \frac{\phi}{2\pi} = \bar{P}_A^{ST} - \lambda N_A^{ST} = \\ &= \rho_A^{ST} - c\Delta t^{ST} + D_{\text{ion},A}^{ST} + D_{\text{trop},A}^{ST} - \lambda N_A^{ST},\end{aligned}$$

in which we may assume $\phi \in [0, 2\pi)$, and

$$N_A^{ST} \stackrel{\text{def}}{=} N_A^T - N_A^S$$

kokonaisluku-
tuntematon
vaihekatko

is the difference of the ambiguities between satellites S and T . Without a *cycle slip* occurring, it will be constant in time.



Two receivers A, B , one satellite S

$$p_{AB}^S \stackrel{\text{def}}{=} p_B^S - p_A^S, \quad P_{AB}^S \stackrel{\text{def}}{=} P_B^S - P_A^S. \quad (13.1)$$

Here, in the same way, the satellite clock offset Δt drops out: it is a property of the satellite, not the receiver, and vanishes when one calculates the difference quantity between two different receivers with the same satellite.

Moreover, the influence of orbit errors, ionosphere and troposphere is reduced substantially: *for short distances AB between receivers* it holds that

$$|\rho_{AB}^S| \stackrel{\text{def}}{=} |\rho_B^S - \rho_A^S| \ll |\rho_A^S| \approx |\rho_B^S|$$

and for $AB \rightarrow 0$:

$$\begin{aligned} d_{\text{ion},AB}^S &\stackrel{\text{def}}{=} d_{\text{ion},B}^S - d_{\text{ion},A}^S \rightarrow 0, \\ d_{\text{trop},AB}^S &\stackrel{\text{def}}{=} d_{\text{trop},B}^S - d_{\text{trop},A}^S \rightarrow 0, \\ D_{\text{ion},AB}^S &\stackrel{\text{def}}{=} D_{\text{ion},B}^S - D_{\text{ion},A}^S \rightarrow 0, \\ D_{\text{trop},AB}^S &\stackrel{\text{def}}{=} D_{\text{trop},B}^S - D_{\text{trop},A}^S \rightarrow 0, \end{aligned}$$

because


- The measurement geometry is almost the same at point A as at point B , see figure 13.1.
- Atmospheric conditions do not change much between points A and B : the measurement rays traverse nearly the same air mass.
- The elevation angle in the sky of satellite S seen from point A is nearly the same as that seen from point B .

13.1.2 Other difference quantities

In the same way, by combining the operations described above, we may also calculate double and triple differences. The formulas look complicated but the process is straightforward, see the summary in tableau 13.2. The equations in the tableau are directly derived from the original observation equations 12.1 and 12.2 by addition and subtraction.

Double differences are much-used in geodetic software packages for processing static GPS network measurements. The double differences are constructed in the office from measurements simultaneously collected in several locations from common satellites.



 TABLEAU 13.2. Summary of GPS observables and difference quantities. The notations $d = d_{\text{ion}} + d_{\text{trop}}$, $D = D_{\text{ion}} + D_{\text{trop}}$ are used in the equations. Note the consistent use of super- and subscripts.

Pseudo-range p	Carrier phase ϕ , equivalent pseudo-range P
$p_A^S = \rho_A^S + c(\Delta T_A - \Delta t^S) + d_A^S$	$P_A^S = \rho_A^S + c(\Delta T_A - \Delta t^S) + D_A^S - \lambda N_A^S$
Single difference, between satellites:	
$p_A^{ST} = \rho_A^{ST} - c\Delta t^{ST} + d_A^{ST}$	$P_A^{ST} = \rho_A^{ST} - c\Delta t^{ST} + D_A^{ST} - \lambda N_A^{ST}$
Single difference, between receivers:	
$p_{AB}^S = \rho_{AB}^S + c\Delta T_{AB} + d_{AB}^S$	$P_{AB}^S = \rho_{AB}^S + c\Delta T_{AB} + D_{AB}^S - \lambda N_{AB}^S$
Double difference:	
$p_{AB}^{ST} = \rho_{AB}^{ST} + d_{AB}^{ST}$	$P_{AB}^{ST} = \rho_{AB}^{ST} + D_{AB}^{ST} - \lambda N_{AB}^{ST}$
Triple difference:	
$\delta_{12} p_{AB}^{ST} = \delta_{12} \rho_{AB}^{ST} + \delta_{12} d_{AB}^{ST}$	$\delta_{12} P_{AB}^{ST} = \delta_{12} \rho_{AB}^{ST} + \delta_{12} D_{AB}^{ST} - \lambda \cdot (\text{cycle slips})$

Triple differences, again, calculated by subtracting double differences for successive time epochs from each other, are almost exclusively used for cleaning up carrier-phase observations, as they are uniquely able to detect “cycle slips”: sudden changes in integer ambiguities caused by interruption of the radio connection between satellite and receiver.

13.2 Relative (static) GPS

The difference in location, or *vector*, $\mathbf{R}_{AB} = \mathbf{R}_B - \mathbf{R}_A$ between two observation sites A and B close to each other may be solved more precisely than the absolute location of either site \mathbf{R}_A or \mathbf{R}_B with respect to the centre of mass of the Earth. The reason for this is the cancellation or partial cancellation from simultaneous observations of various error sources, the impacts of which are similar in both points. *Difference observations* $p_{AB} = p_B - p_A$, $P_{AB} = P_B - P_A$ between the two observation sites, in which this cancellation already occurs, are used for computation. As shown earlier in figure 13.1, the places of the satellites in the sky are nearly the same seen from observation sites A and B , and the atmospheres above A and B are also likely rather similar.

How much, for example, does the *orbit error* of satellites S and T affect the determination of the vector AB ? See figure 13.3. An order-of-magnitude rule of thumb says that the positioning error δ caused by orbit



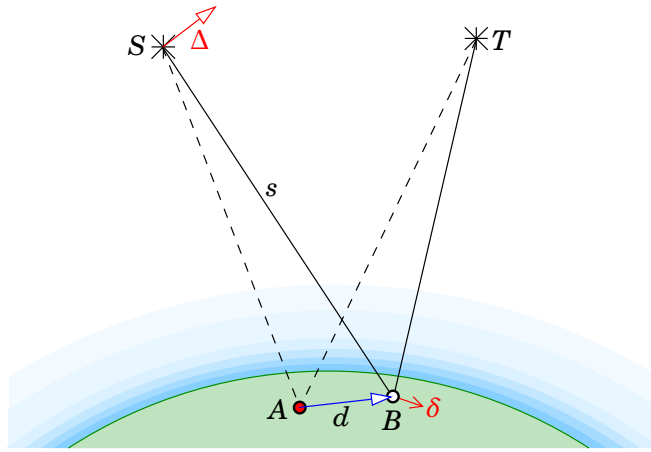


FIGURE 13.3. Double difference, short distance between GPS receivers.

error in point B relative to point A is

$$\delta \approx \frac{d}{s} \Delta,$$

in which Δ is the assumed orbit error. This is only a crude estimate. We know that $s \gtrsim 20000$ km.

The values given in table 13.3 for the orbit error, 1 m and 2 cm, correspond to the accuracies of today's *broadcast*² and *precise* ephemeris. The conclusion is that



TABLE 13.3. Approximate relation between orbit error, length of vector, and positioning error.

Vector length d (km)	Orbit error Δ (m)	Positioning error δ (mm)
1	1	0.05
10	1	0.5
100	1	5
1000	1	50
1	0.02	0.001
10	0.02	0.01
100	0.02	0.1
1000	0.02	1

²This is a crude estimate. The quality of broadcast ephemeris has improved since the early days of GPS, slowly but surely. Other GNSS systems, like GLONASS, Galileo and BeiDou, perform at about the same level, or perhaps a little less well (Montenbruck et al., 2015).

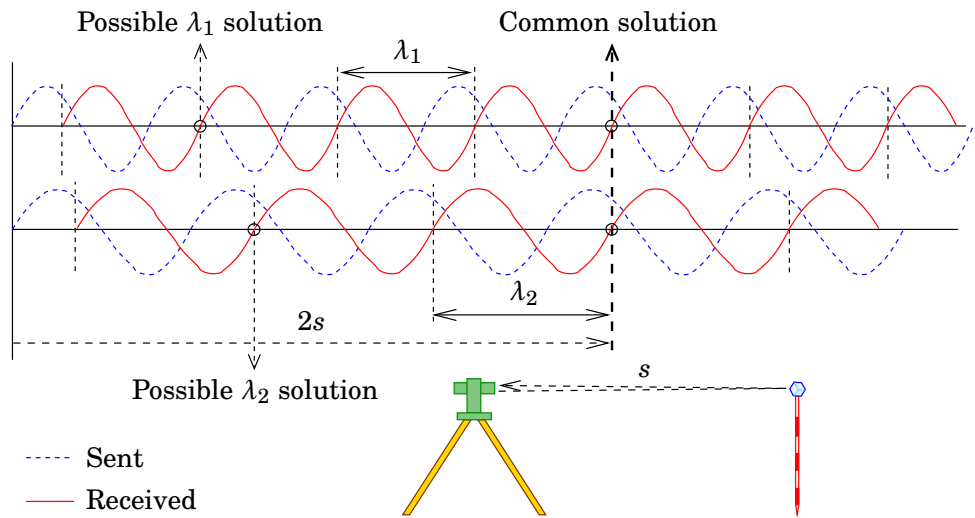


FIGURE 13.4. One-dimensional ambiguity resolution in the case of a distance measurement instrument. See also figure 7.10. Depicted is the situation where the transmitted phases cross zero in the positive direction. It is assumed that at that moment the received phases are all measured to be zero too, see figure 7.9.



In GPS surveying work in a small area ($\lesssim 100$ km) the satellite orbits may be assumed known.

In geodetic work, first one computes double differences P_{AB}^{ST} from observations at points A and B . As we are dealing with carrier-phase observations, the ambiguities or integer unknowns N_{AB}^{ST} must first be resolved. After that, a *vector*

$$\bar{\mathbf{R}}_{AB} = \begin{bmatrix} X_{AB} \\ Y_{AB} \\ Z_{AB} \end{bmatrix} = \begin{bmatrix} X_B - X_A \\ Y_B - Y_A \\ Z_B - Z_A \end{bmatrix}$$

between the points is computed from the observations. This is where the term *relative GPS* (or *GNSS measurement*) comes from.

A generalisation of this vector solution is the measurement and processing of observations from a number of points, a *geodetic network*.



13.3 Fixing ambiguities

Resolving the integer unknowns or *ambiguities* is a precondition for using **kantaaalto** GPS carrier-wave observations. There are several methods for this.

- *Distance measurement equipment* resolves the integers by measuring at several different wavelengths. Figure 13.4, as earlier on



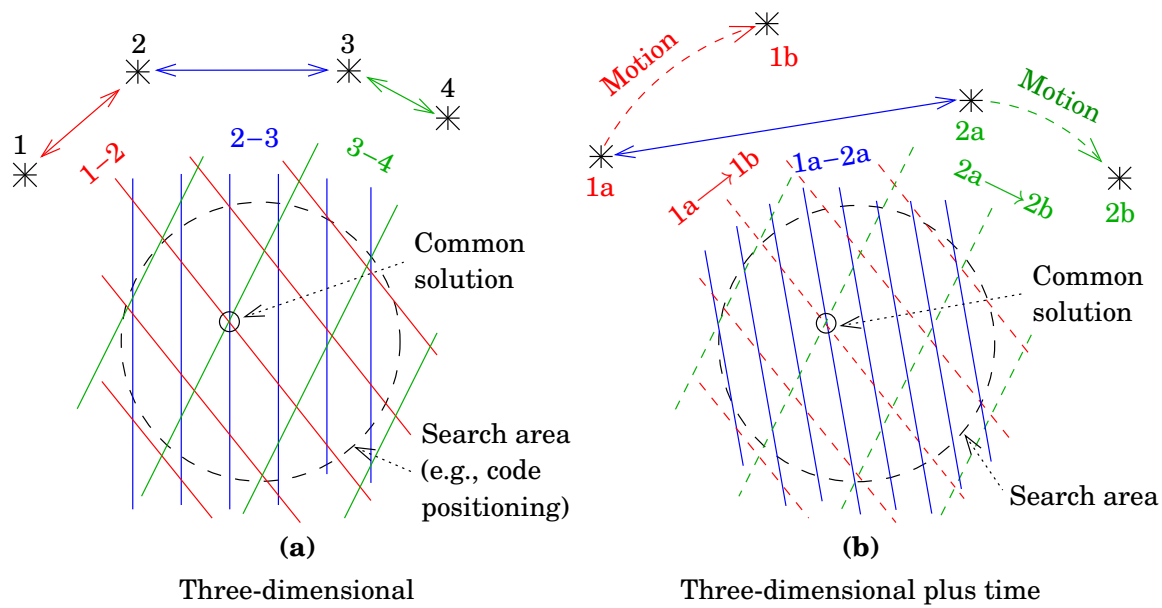


FIGURE 13.5. Various ambiguity-resolution methods used in GPS computation. The coloured line bundles — actually cut-outs of bundles of hyperboloids, or in three dimensions, hyperboloids of revolution — indicate the sets of relative receiver position solutions that are compatible with double-difference observations obtained using either a pair of satellites, or a single satellite at two different times.



figure 7.10, shows how this puzzle can be solved.

A GPS satellite transmits on two frequencies L_1 and L_2 . When there are two frequencies, they can be combined in a way which makes it easier to resolve the ambiguities over short distances. Calculate the phase difference $\phi_w = \phi_1 - \phi_2$ between the phase measurements at L_1 and L_2 . It is like using a carrier wave with a frequency $f_w = f_1 - f_2 = 347.82\text{MHz}$, and a corresponding “wavelength” $\lambda_w = c/f = 86\text{cm}$. This method is called *wide-laning*.

leveäkuja-
ratkaisu

The wide-lane solution only works over short distances, because otherwise the unknown difference in ionospheric influence between the two measurement sites grows too large.

The code observation already yields a pseudo-range at the metre precision level, after which the wide-laning method yields the ambiguities, and the phase difference ϕ_w yields an “ambiguity-free” pseudo-range at a precision level of centimetres.

vaihe-ero
pseudoetäisyys

- Use many satellites. At the moment, over 30 GPS satellites orbit



the Earth. At almost any moment, 6–10 of these are in the local sky above the horizon. This is a more complicated method, because, unlike in distance measurement, the geometry is three-dimensional. Efficient algorithms for this exist.

- Use the same satellites for a longer time. Because the GPS measurement geometry has time to change, we obtain more conditions.

See figure 13.5.

taajuusjako-
kanavointi
kokonaisluku-
tuntemattomien
ratkaisu
3

ohjelmisto-
pohjainen
GNSS-
vastaanotin

In recent years, other global positioning systems have appeared besides GPS. The Russian GLONASS — in which every satellite has its own carrier transmission frequency, *frequency division multiple access* (FDMA), which complicates ambiguity resolution³ — is operating, after a long period of decay, again with a full constellation of 24 satellites, the European Galileo system is approaching completion, and the Chinese are coming with their Beidou-3 system. Joint use of the systems in the same receiver is technically complicated but promises a very fast and reliable resolution of ambiguities. Relief may come from “*software-defined GNSS receivers*”, in which all processing work after the antenna and analogue electronics is implemented digitally in software on an off-the-shelf computer.



13.4 Real-time positioning

jälkilaskenta

The static method described above is based on *post-processing*. For geodetic use, this is usually unproblematic. The use of precise satellite orbits — essential if one wants geodetic precision for long vectors — also imposes a certain waiting time: a couple of weeks in the case of precise ephemeris.

tosiaikainen
4

Sometimes, however, we need the co-ordinates of new points immediately, and there are also situations in which this would be useful or handy. Then, we speak of *real-time*⁴ positioning. *Navigation* is a broad field of application.

In precise GPS positioning, real-timeness can be implemented by transferring the observations made at known point *A* on the fly, for example by radio, to unknown point *B* for joint processing.

³In the new GLONASS-K satellites, CDMA, code division multiple access, like in GPS, is also being offered.

⁴The formal definition of real-timeness is *a guaranteed latency*. It may be long, as long as it is guaranteed. The latency or response time is the time that elapses from a measurement event to the availability for use of the measurement values.



13.4.1 Differential GPS (DGPS)

Differential GPS is a real-time positioning method based on *code observations*, which makes use of a reference or *base station*. It is thus a *relative* measurement between base station *A* and moving receiver, or *rover*, *B*. As always with measurements from two receivers close to each other, the error sources are the same or nearly the same in both receivers: the orbit errors and clock offsets of the satellite, as well as the effect of the atmosphere, cancel out in the difference measurement between *A* and *B*, either completely or nearly so.

tukiasema

In the DGPS method, it is not the raw observations at point *A* that are transferred to point *B* — that would be an unnecessary amount of information to transfer. Instead, the *difference* between the measured and the computed pseudo-range is first calculated from the observations. For every observation p_A , the geometric distance $\rho_A^{(0)}$ between observation site *A* and satellite *S* can be computed. This “reference distance” is based on the known location of point *A* and the orbital data transmitted by the satellite. Then, one obtains the *pseudo-range offset* for each satellite in the sky:

pseudoetäisyys-poikkeama

$$dp_A \stackrel{\text{def}}{=} p_A - \rho_A^{(0)}.$$

The orbital data from the same satellite are also available to the moving receiver *B*, which may itself compute from these the same $\rho_A^{(0)}$ — after all, the location of point *A* is known. So, the *information content* of the offsets dp_A is the same as that of the full measurements p_A , and the offsets may replace them in the dissemination.

liikkuva vastaanotin

Using pseudo-range *offsets* has the following advantages:

- The numerical values are much smaller. The offsets were of the order of $\pm 100\text{m}$ back when selective availability (*SA*, an artificial reduction in accuracy of the disseminated orbital and clock information) was still on. When *SA* was switched off in 2000, the magnitude of the offsets dropped to the level of $\pm 5\text{m}$. Both orders of magnitude are fractions of the size of the observables themselves, tens of thousands of kilometres.
- The values change more slowly. They crawl over the course of hours in a way which looks random. Extrapolation over several seconds or minutes into the future works better than with raw observations.

For both reasons, the communications bandwidth needed is much less, and the following channels are sufficient:



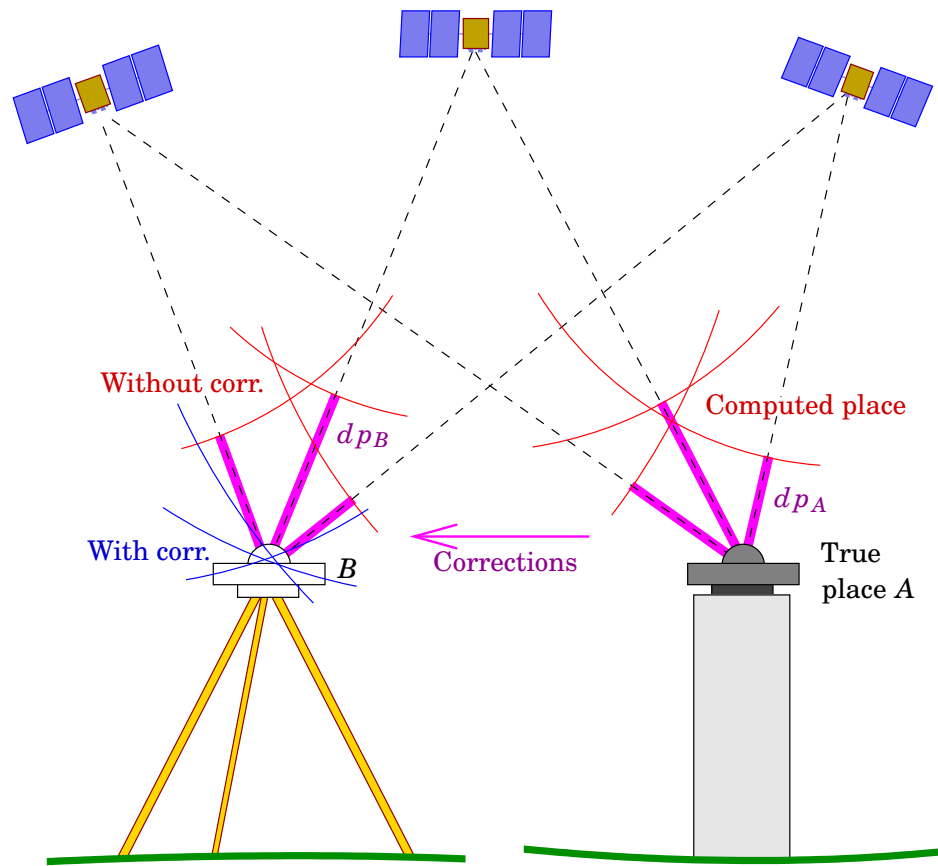


FIGURE 13.6. Principle of operation of the DGPS method, somewhat simplified.

- Mobile telephony. Modern network data connections (3G, 4G) are fast indeed. The “mobile Internet”.
- A radio modem (short distances).
- For navigation at sea, long-wave radio.

The values of the offsets also vary slowly as functions of *place*. Therefore we may, if the distance AB is suitably short, 100–1000 km, write with sufficient accuracy

$$dp_B \approx dp_A + c(\Delta T_B - \Delta T_A),$$

in which ΔT_A and ΔT_B are the receiver clock offsets.

Instead of the original, single observations, we construct *difference observations* p_B^{ST} between two satellites S and T :

$$\left. \begin{aligned} dp_B^S &\approx dp_A^S + c(\Delta T_B - \Delta T_A) \\ dp_B^T &\approx dp_A^T + c(\Delta T_B - \Delta T_A) \end{aligned} \right\} \implies dp_B^{ST} = dp_B^T - dp_B^S \approx dp_A^T - dp_A^S,$$

a difference between disseminated quantities, from which the clock offsets of both receivers are eliminated. Now, we may calculate the single



difference in point B of the ranges to the two satellites:

$$\rho_B^{T(0)} - \rho_B^{S(0)} = \rho_B^{ST(0)} = p_B^{ST} - dp_B^{ST},$$

a purely geometric quantity between the location of point B and the locations in space of the satellites S and T computed from the orbital data, all known to the user. From these, the unknown location of point B may be solved for, without any influence from the satellites' orbit errors or clock offsets. Three difference observations, four satellites, are sufficient.

By differential GPS positioning and navigation ("DGPS") is commonly meant precisely this real-time method using the modulations of the GPS signal, the C/A code and P code.



13.4.2 Kinematic real-time positioning, "RTK"

If we use carrier-phase measurements in real time, we speak of RTK, real-time kinematic, measurement. With this technique, precision is much better, even if only over short distances. Unlike DGPS, this is an *incremental* method: the measurement must start from a *point with known co-ordinates*, and preferably also close on a known point — just in case. The co-ordinates of the points measured in between are obtained in relation to these known points.

See figure 13.7. When using *double differences*, between two satellites and the base station and rover, the only unknowns are the rover's three co-ordinates X , Y , and Z , its location in three-dimensional space \mathbb{R}^3 . All possible locations of the rover that are compatible with the double-difference observation now form a *bundle of hyperboloids of revolution*. The distance separating the hyperboloids corresponds to one wavelength of the observable. In the figure, the hyperboloids are drawn in cross-section as curves of different colours, corresponding to two different satellite pairs. As always with GPS measurement, the minimum number of satellites is four.

pyörähdys-
hyperboloidi

On known point A is installed a permanent (or semi-permanent) GPS reference or base station. Let the momentaneous location of the rover be B . The double differences are

$$P_{AB}^{ST} = P_B^T - P_A^T - P_B^S + P_A^S.$$

Substituting equation 12.2 into this yields

$$P_{AB}^{ST} = \rho_{AB}^{ST} + D_{\text{ion},AB}^{ST} + D_{\text{trop},AB}^{ST} - \lambda \cdot N_{AB}^{ST}.$$



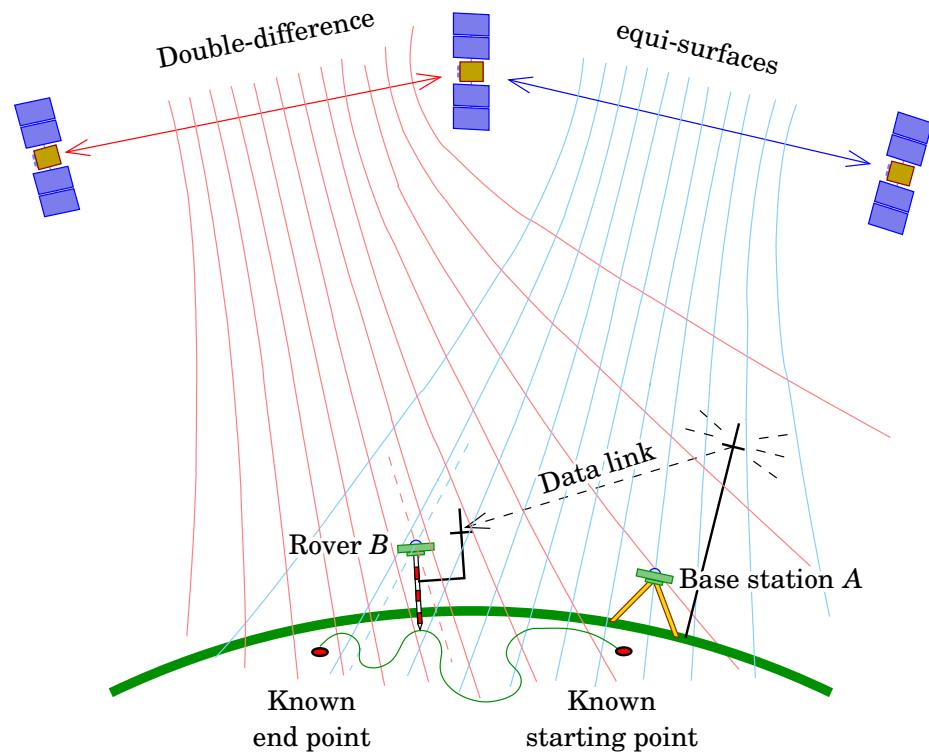


FIGURE 13.7. Principle of operation of the RTK method, simplified to two dimensions. The hyperbolas are the rover locations which are compatible with phase double differences corresponding to a round number of whole wavelengths. The dashed hyperbola segments again are locations compatible with the *observed* double differences. Compare this with the Decca figure 12.1!



Let us forget the atmospheric terms for a moment:

$$P_{AB}^{ST} = \rho_{AB}^{ST} - \lambda \cdot N_{AB}^{ST}.$$

- ⁵ If the *rover's* antenna is set up on a point with known co-ordinates ⁵ B , all co-ordinates are known:
- The co-ordinates of satellites S and T can be computed from the orbital ephemeris — even from broadcast ephemeris, if the distance AB is not very long.
 - The co-ordinates of stations A and B are assumed known.

Therefore, the *geometric* double difference for that moment, $\rho_{AB}^{ST}(t_0)$, is ⁶ *computable*. After that, the *ambiguities* ⁶ may also be resolved from the

⁵Or the co-ordinates of the rover's starting point are *determined from observations* before starting to move, the on-the-fly method. See the next subsection.



observations:

$$N_{AB}^{ST} = - \frac{P_{AB}^{ST}(t_0) - \rho_{AB}^{ST}(t_0)}{\lambda}.$$

After this, we take the rover away from the known point and go and measure. We move in the terrain and measure a suitable number of unknown points — but *in such a way that the connection to the satellites is not interrupted*. Then namely the values of the ambiguities N_{AB}^{ST} do not change either, and we may straightforwardly calculate the *geometric* double difference

$$\rho_{AB}^{ST}(t_P) = P_{AB}^{ST}(t_P) + \lambda \cdot N_{AB}^{ST}$$

from the measurements $P_{AB}^{ST}(t_P)$ at terrain point P . This is the essence of *kinematic* GPS positioning. The correct place is obtained immediately, even with millimetre-precision: of course, only in relation to the reference station, not absolutely. Therefore, the precise geodetic determination of the location of the reference or base station is essential.

RTK works best over short distances, from hundreds of metres to tens of kilometres. The real-time nature requires use of a data communications link between base station and rover. The possible data communications solutions are in principle the same as in the case of DGPS.

tosiaikaisuus



13.4.3 RTK, technical considerations

Above it was assumed that the moving receiver, the rover, starts from a point with co-ordinates known to geodetic precision. Such a point may however also be *created* “on the fly”, by staying on the starting point for so long that the satellite geometry changes and the resolution, the fixing to integer values, of the ambiguities becomes successful. The link with the base station has to be open, and the signal connection with those satellites that are also visible from the base station must be uninterrupted.

lennossa

Four satellites are generally sufficient for solving the location with GPS. The RTK technique’s “on-the-fly” initialisation, however, requires at least a fifth satellite, in order to resolve the ambiguities. Real-time quality control of the measurements also requires this. Without redundancy,

⁶These values must thus be integers or close to integers, in which case they may be rounded, resulting in the “fix” solution. If they are not near integer values, they may *not* be rounded. Then one obtains the, weaker, “float” solution. This may happen if, for example, there is an error in the given co-ordinates of A or B . Or if the effect of the atmosphere is too strong after all, or the distance AB too long.



the ambiguities N_{AB}^{ST} would be freely chooseable, and one could compute, from the three values ρ_{AB}^{ST} thus obtained, a fantasy vector solution $\bar{\mathbf{R}}_{AB}$, without any contradictions being generated!

The more satellites are available, the quicker the ambiguity resolution will proceed. This is why instruments have appeared on the market that are able to use simultaneously the signals from both the **GPS** and, for example, the **GLONASS** satellites.

Over short distances — less than a kilometre — the number of satellites effectively doubles, because the difference between the L_1 and L_2 frequencies, the “wide lane”, can be used, the effective wavelength of which is 86 cm.

Today’s **RTK** instruments are able to intelligently use many known points around a measurement area. Before and after the survey, these points are visited, and the instrument forms, using the known and measured point locations, a local *transformation formula*. Using the formula, all measured points are transformed to the same system in which the known points have been given. This is a handy but also dangerous property: the accuracy of the transformed co-ordinates cannot be better than the interior accuracy of this local system. If it is, for example, the old **KKJ** system based on traditional measurement techniques (subsection 3.2.1), one loses the major advantage of **GPS** surveying, its superior geometric accuracy!



13.4.4 Network-mode real-time services

In recent years, many *base-station networks* have been built, in order to offer both differential **GPS** (**DGPS**) and real-time kinematic (**RTK**) support services. The main problem with using a single base station is that the corrections disseminated by it are good close to the base station, but deteriorate quickly as the distance of the rover from the base station grows.

It is intuitively clear that the corrections change *only slowly*, and almost linearly, with place. The correction for the satellite clock offset is even constant. The satellites are so high that even over an area the size of Europe the geometry is almost the same in different parts of the continent. Moreover, the atmosphere is usually rather similar everywhere within a small area. This opens up the possibility of *interpolation*, and the base-station networks do precisely this.

tukiasema-
verkko

Realistic alternatives for disseminating correction signals:



- Geostationary communications satellites. Their advantage is the homogeneous coverage of large areas, their disadvantage, the satellites' low elevation angle at Finnish latitudes.
- Using the mobile Internet through the mobile telephony network. The Internet is not real-time, but in practice fast enough for it to often not matter. Advantages are
 - an easy way to *charge* for the service
 - the possibility to supply, to the location of the receiver, *tailored* corrections — the “virtual base station” idea
 - a large data transfer capacity at little cost.

virtuaali-
tukiasema

Using the services in network mode also requires software support. The geometric aspect of interpolating the corrections is easy; the problem is formed by the *modelling* of the propagation delay by the *atmosphere*. In order to achieve good accuracy, the base-station network used for the model computation must be sufficiently dense. This complex problem field is the subject of active research, for example [Koivula et al. \(2018\)](#).

kulkuviive



13.5 SBAS, satellite based augmentation systems

The term wide-area differential [GPS \(WADGPS\)](#) is also used for this type of system. These systems use geostationary satellites to disseminate differential corrections for [GPS](#) positioning. Globally, three intercompatible systems exist:

- [WAAS](#) (Wide Area Augmentation System, USA)
- [EGNOS](#) (European Geostationary Navigation Overlay System)
- [MSAS](#) (Multi-functional Satellite Augmentation System, Japan).

These are already in widespread use, although their development continues. The services are continent-wide and are based on the simultaneous use of many [GPS](#) base stations in network mode. In this way, accurate differential corrections are obtained for the area covered by the base stations. The signal structure and frequencies used are the same as for the [GPS](#) satellites, which makes it relatively easy to modify an existing [GPS](#) receiver design for [SBAS](#) use.

A core application is [GPS integrity monitoring](#), which sounds an alarm if the quality of positioning cannot be guaranteed. This is important in safety critical applications — [SoL](#), “Safety of Life” — like positioning aircraft during approach and landing.

eheyden
valvonta

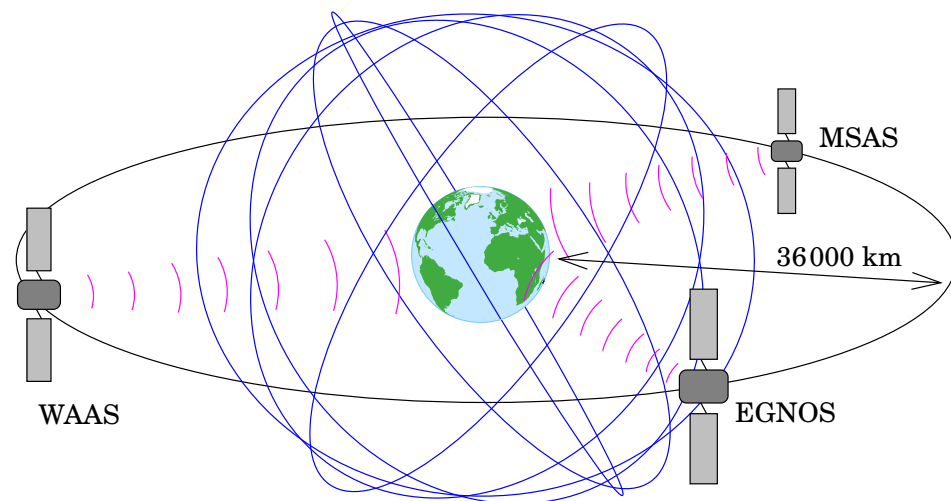


FIGURE 13.8. Satellite-based augmentation systems (SBAS).



13.6 Real-time satellite positioning support services in Finland



13.6.1 The DGNSS service of the Finnish Transport Agency

This service, **Väylä, Radionavigaatiopalvelut**, operated by the Finnish Transport Infrastructure Agency (earlier the Finnish Transport Agency, still earlier the Finnish Maritime Administration) sends correction messages by radio on long wavelengths, frequencies 287.5–314.5 kHz. The service covers the sea areas of the Baltic Sea, and in Finland also the lake area of Saimaa. The user community consists of seafarers. The service is free of charge.



13.6.2 Trimnet VRS

Geotrim, Trimnet is the real-time kinematic (RTK) network service maintained by Geotrim Oy, originally planned for use by the Finnish National Land Survey. The technology used is called network RTK, more precisely, virtual reference station RTK (VRS-RTK), and is based on generating, for every user, a computational “virtual” base station close to them. Correction data for this base station are generated in the RTCM-SC104 format, a standard format (“RTCM format”) useable by devices of any brand.

The network covers all of Finland with over a hundred base stations (as of 2019) and also supports the use of **GLONASS**. The corrections are disseminated commercially over the mobile Internet: a cluster of servers is located at Geotrim headquarters in Vantaa, to which the user can log in over the Internet. The corrections obtained are tailored for each user



separately. The precision obtainable is of the order of a centimetre in the horizontal plane, a little poorer vertically.

13.6.3 HxGN SmartNet

HxGN SmartNet ([Hexagon Geosystems, HxGN SmartNet](#)), operated and maintained by Hexagon Oy, is a real-time kinematic (RTK) base-station network offering corrections as a commercial service over the mobile Internet. In Finland there are currently (2017) over a hundred base stations.

13.6.4 The experimental service of the National Land Survey

This service of the Finnish National Land Survey's Geospatial Research Institute, [Finnish National Land Survey, About positioning services](#), which is at the moment experimental and free of charge but which requires registration, offers both differential GNSS and network RTK. The base stations are the twenty new FinnRef stations.

13.6.5 Archive data service

These network RTK service providers, in Finland and abroad, usually also supply under agreement archived data in the RINEX⁷ format for post-processing. Typically, the data is thinned out, from a data rate of one Hz — one measurement event per second from all visible satellites — to, say, $\frac{1}{30}$ Hz, one measurement event every 30 seconds.

13.7 The GDGPS system

[JPL, The Global Differential GPS System](#). This is a global real-time DGPS service using the Internet as its distribution channel. The system was developed and is being operated by NASA's Jet Propulsion Laboratory (JPL). As part of the service, it offers an online precise point positioning (PPP) service using RINEX files uploaded by the user.

⁷RINEX, Receiver-INdependent EXchange format, is a practical, device independent text format that also a human being can read. [Gurtner and Estey \(2007\)](#). There is a conversion program for every receiver brand. See tableau [12.3](#).



Self-test questions

1. What is the “common-mode” error assumption?
2. What kinds of difference observations are there, and what error sources do they eliminate or reduce?
3. What is differential [GPS \(DGPS\)](#)?
4. What is real-time kinematic ([RTK](#)) positioning?
5. For what purpose have [SBAS](#) (satellite-based augmentation systems) been built?
6. What is [RINEX](#)?



Exercise 13–1: Geodetic GPS positioning

The task is to determine a precise vector between two stations by processing “in the cloud”. [GNSS](#) data from many hundreds of continuously operating stations are available online, as are services for using these data for geodetic position computation. We are going to exploit this for a test computation.

1. Get the [RINEX](#) data files from the Sodankylä (SODA) and Metsähovi (METS) stations; see figure [13.1](#). As sources you may use the web-site of the [FGI \(Finnish National Land Survey, RINEX service\)](#), which however requires registration, or the [SOPAC web-site \(SOPAC, GPS Explorer\)](#) of the Scripps Institution of Oceanography in San Diego, California, or the [EUREF data centre \(ROB, EUREF Permanent GNSS Network\)](#) at the Royal Observatory in Brussels, Belgium. Download data for a single day⁸ — the same for both stations — of which the size should not exceed 5 MB.
2. The data downloaded may be *Hatanaka compressed* ([SOPAC, Hatanaka file compression](#)). This is a compression technique specifically for [RINEX](#) data, exploiting the similarity of the data for successive epochs. You need to get the conversion program [CRX2RNX](#) as a binary for your operating system, and convert the data into uncompressed [RINEX](#). The latter is human-readable and even clear. The Hatanaka compressed file is also human-readable, and the compression technique used is clearly visible.

Remember that the data may also be compressed by a standard

⁸Suggestion: your last year’s birthday!

method, like zip or gzip or Unix compress (*.Z) . . . which you first have to expand. On Windows, the utility 7-zip may be useful.

Also, on Windows, you may run into the line-ending problem: Unix text files end their lines with a line-feed (LF) only, Windows uses a carriage-return (CR) followed by a line-feed (LF). The text editor Notepad++ may be useful.

3. Now, you can upload your [RINEX](#) files to the cloud. There are two alternatives — choose one for this exercise:
 - [AUSPOS – Online GPS Processing Service](#), an Australian government service. It tends to be a little slow: often the result comes overnight, in the form of an extensive report. Publicly available data from nearby [IGS](#) stations is used in the positioning computation, as shown in a map.
 - [GDGPS APPS](#), a US government service run by the Jet Propulsion Laboratory [JPL](#) — the headquarters of the International [GNSS](#) Service [IGS](#) (subsection 12.11) — and yes, they too use data from the [IGS](#) network (figure 12.24) for fixing the reference frame in the computations.
You need to upload METS and SODA separately, one file at a time.
4. The [AUSPOS](#) results come by email.
 - (a) Read the results carefully. What other stations were included in the computation, and in what reference frame is the result expressed?
 - (b) The geocentric Cartesian (rectangular) co-ordinate solution X , Y , Z .
 - (c) The geodetic co-ordinates, and their precisions (“Positional Uncertainty”).
 - (d) Other interesting stuff. How is the tropospheric propagation delay modelled? Yes, they estimate dry and wet tropospheric zenith propagation delays as well as horizontal gradients!
 - (e) Note the use of a geoid model, for obtaining heights over mean sea level. How good do you think it is?
 - (f) Ambiguity resolution.
5. The [APPS](#) results appear online.
For [APPS](#), go to the summary file (*.sum) and look up the following



things:

- (a) The geocentric Cartesian co-ordinate solution X , Y , Z , and the co-ordinate uncertainties (“sigmas”). How does this precision concept differ from that of [AUSPOS](#)?
- (b) The geodetic co-ordinates Lat, East_Lon and Height, and their sigmas. Compare the height sigma with the others.
- (c) The other interesting headers. How is the troposphere modelled here? Compare with [AUSPOS](#).





Adjustment calculus in geodesy

14

[...] Mais comment vous décrire mon admiration et mon étonnement, en voiant se metamorphoser mon correspondant estimé M. Leblanc en cette illustre personnage, qui donne un exemple aussi brillant de ce que j'aurois peine de croire. Le goût pour les sciences abstraites en général et surtout pour les mysteres des nombres est fort rare : on ne s'en étonne pas ; les charmes enchanteurs de cette sublime science ne se decellent dans toute leur beauté qu'à ceux qui ont le courage de l'approfondir. Mais lorsqu'une personne de ce sexe, qui, par nos mœurs et par nos préjugés, doit rencontrer infiniment plus d'obstacles et de difficultés, que les hommes, à se familiariser avec ses recherches epineuses, sait neansmoins franchir ces entraves et pénétrer ce qu'elles ont de plus caché, il faut sans doute, qu'elle ait le plus noble courage, des talens tout à fait extraordinaires, le génie supérieur. [...]

Letter from Gauss to Sophie Germain, 1807 ([Friedelmeyer, 2014](#)). See also [Wikipedia, Sophie Germain](#).



14.1 Why adjustment?

In geodesy, as in science in general, we know that *all measurements are wrong*. Therefore, we collect always more measurements than the strict minimum, so as to be able to judge at least somewhat realistically the uncertainties in the measurement results. This practice is called *redundancy*.

A good example is measurements in a geodetic network. In a triangulation network, figure [14.1](#), directions from the triangle points to other [kolmiomittaus](#)

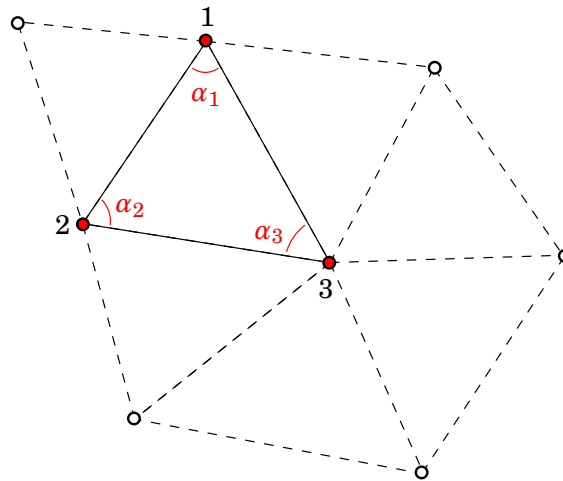


FIGURE 14.1. Triangulation network, triangle.

triangle points are measured.

From the direction measurements to neighbouring points made at points 1, 2, and 3, angles α_1 , α_2 , and α_3 are calculated. For the angles,¹ the *triangle condition*¹ applies:

$$\alpha_1 + \alpha_2 + \alpha_3 = 180^\circ.$$

The triangle condition makes the following checks possible:

karkea virhe

1. The measurements from which the angles α_1 , α_2 , and α_3 were calculated do not contain *gross* errors. For example, if we obtain for the sum of angles $\underline{\alpha}_1 + \underline{\alpha}_2 + \underline{\alpha}_3 = 173^\circ.6742$, we may immediately conclude that there must be at least one gross error in the measurement set, because the accuracies of the instruments used are fractions of a degree.
2. The amount by which the sum obtained differs from the theoretical value of 180° allows us to infer the precision of the measurement. For example, if we obtain as the sum $\underline{\alpha}_1 + \underline{\alpha}_2 + \underline{\alpha}_3 = 179^\circ.9958$, then the *closing error* $\underline{\Delta} \stackrel{\text{def}}{=} \underline{\alpha}_1 + \underline{\alpha}_2 + \underline{\alpha}_3 - 180^\circ = -0^\circ.0042$, and the inference is that the precision of the measurement method used is several thousandths of a degree.

If we do more than the minimum number of measurements in this way, we need a method for removing the, small, contradictions between these measurements, reconciling them with each other. A brute-force trick

¹On the curved surface of the Earth, the sum is not exactly 180° but a little larger, the *spherical excess*. This is an example of non-Euclidean geometry.



would be to just throw measured value $\underline{\alpha}_3$ away, and compute a replacement value, guaranteed to be compatible, of $\underline{\alpha}_3 = 180^\circ - \underline{\alpha}_1 - \underline{\alpha}_2$. However, one may justifiably ask, why $\underline{\alpha}_3$ rather than $\underline{\alpha}_1$ or $\underline{\alpha}_2$? Such arbitrariness is unacceptable, and we should not just throw away the valuable information contained in the observation. The right solution is *network adjustment*.

verkkotasoitus

In the simple triangle case we divide the closing error *equally* among the angles: the adjusted angle values will be democratically

$$\hat{\alpha}_1 = \underline{\alpha}_1 - \frac{1}{3}\Delta, \quad \hat{\alpha}_2 = \underline{\alpha}_2 - \frac{1}{3}\Delta, \quad \hat{\alpha}_3 = \underline{\alpha}_3 - \frac{1}{3}\Delta,$$

after which $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 = 180^\circ$ exactly.

If we however know that, say, the angle α_3 was measured *twice* (and the value $\underline{\alpha}_3$ is the average of these measurements) but the angles α_1 and α_2 only once using the same instrument, we may take this into account by adjusting in the following way:

$$\hat{\alpha}_1 = \underline{\alpha}_1 - \frac{2}{5}\Delta, \quad \hat{\alpha}_2 = \underline{\alpha}_2 - \frac{2}{5}\Delta, \quad \hat{\alpha}_3 = \underline{\alpha}_3 - \frac{1}{5}\Delta,$$

in which still $\Delta = \underline{\alpha}_1 + \underline{\alpha}_2 + \underline{\alpha}_3 - 180^\circ$. In this case one speaks of *weighting* the measurements. The measurement $\underline{\alpha}_3$ is given a *double weight* — and thus a halved correction — compared to the measurements $\underline{\alpha}_1$ and $\underline{\alpha}_2$.

A large weight means a small correction, and vice versa, figure 14.2. The weight ratios are 1 : 1 : 2, the correction ratios are the reverse 2 : 2 : 1, and the sum of the correction ratios is 5. This is how the above correction coefficients (“weight coefficients”) $\frac{2}{5}$, $\frac{2}{5}$, and $\frac{1}{5}$ are obtained.

The adjustment of a realistic, complex triangulation network (or any geodetic network) is mathematically much more complicated, but this is the basic idea.

kolmiomittausverkko

It is commonly assumed that the random measurement errors of the observations are distributed according to the bell curve named after C. F. Gauss (figure 2.5), in other words, that they are *normally distributed*. At least in that case, the theoretically best solution is given by least-squares adjustment. The simplest example of this is computing the *average*.

pienimmän neliosumman tasoitus



14.2 The average

Assume that the same quantity has been observed n times — so, we have a *stochastic quantity* ℓ — the observation values being ℓ_i , $i = 1, \dots, n$. The observations have a statistical expectancy μ , and they all have the same

odotusarvo



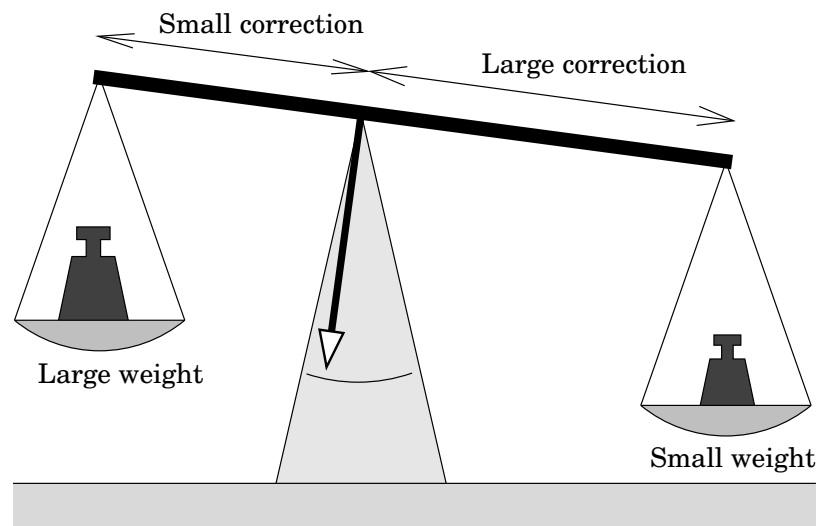


FIGURE 14.2. Metaphor: a large weight means a small correction and vice versa.

² standard deviation or *mean error*² σ . The average of the observations is³

$$\bar{\ell} = \frac{1}{n} (\ell_1 + \ell_2 + \dots + \ell_n) = \frac{1}{n} \sum_{i=1}^n \ell_i.$$

One can show that, as an estimator of the expectancy μ , this is the “best possible” linear combination of observations. One can also show that this linear combination minimises the sum of squares of the *residuals*⁴

jäännösvirhe⁴

$$v_i \stackrel{\text{def}}{=} \bar{\ell} - \ell_i, \quad i = 1, \dots, n,$$

$$\sum_{i=1}^n v_i^2 = v_1^2 + v_2^2 + \dots + v_n^2 = \min.$$

pienimmän
neliösumman
menetelmä

otoskeskivirhe

This property is the origin of the term “least-squares method”.

We may also *estimate* the standard deviation or mean error of a single observation, σ , by the equation for the sample mean error

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n v_i^2}.$$

²Expressed more theoretically, if the expectancy operator is $E\{\cdot\}$, we may write $\mu = E\{\ell\}$ and $\sigma^2 = E\{(\ell - \mu)^2\}$.

³Here, the values ℓ_i are written as *stochastic*, because the formation of the average may be *repeated*, to form different realisations of the stochastic quantity $\bar{\ell}$.

⁴The residual $\bar{\ell} - \ell_i$ of an observation is *not* the same as (the opposite of) the *error* $\ell_i - \mu$ of that observation! The residual is computable from the observations, the error is not.



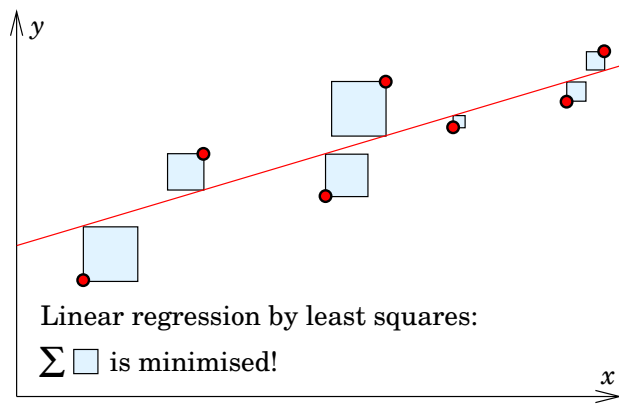


FIGURE 14.3. The idea of linear regression.

From this again, the quality, or uncertainty, measure for the average $\bar{\ell}$, its mean error (standard deviation) estimate follows, $\hat{\sigma}_n = \hat{\sigma} / \sqrt{n}$. This value, which describes the uncertainty of the average, thus becomes smaller the longer the series of measurement values is, in other words, the larger n .

If we know the mean error σ of a single observation *a priori*, we may also directly use the equation $\sigma_n = \sigma / \sqrt{n}$, the result of which is not an estimate but a computed value.

14.3 Linear regression

In linear regression, two parameters a and b are estimated when observations \underline{y} that depend linearly on the argument x are given:

$$\underline{y}_i + \underline{v}_i = \hat{a} + \hat{b}x_i.$$

Here, \underline{v}_i is again the *residual* of observation i . Parameters a and b describe a *straight line* that runs as well as possible — meaning, with as small residuals as possible — through the measured “point cloud” (x_i, \underline{y}_i) . See figures 14.3 and 14.4.

Linear regression is a *least-squares method*, figure 14.3. The sum of the squared residuals is minimised.

The least-squares solution is

$$\hat{b} = \frac{n \sum (xy) - \sum x \sum y}{n \sum (x^2) - (\sum x)^2}, \quad \hat{a} = \frac{\sum y - \hat{b} \sum x}{n},$$

in which the compact notation is used

$$\sum (\cdot) \stackrel{\text{def}}{=} \sum_{i=1}^n (\cdot),$$



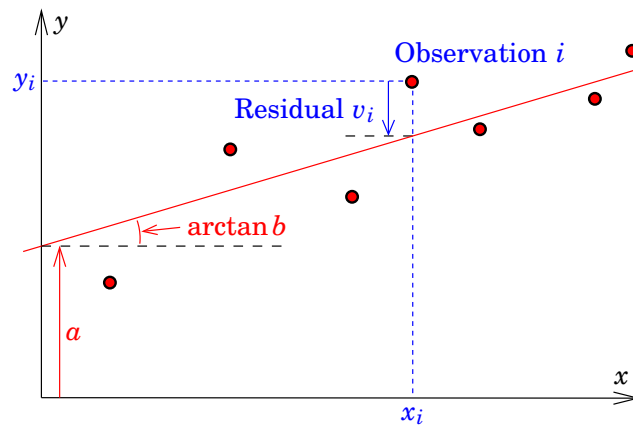


FIGURE 14.4. Linear regression, definitions of quantities.

a summation over all n points, or co-ordinate pairs, (x_i, y_i) . We can write the solution even more neatly in terms of *averages*:

$$\hat{b} = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2}, \quad \hat{a} = \langle y \rangle - \hat{b} \langle x \rangle,$$

if the average is written as

$$\langle \cdot \rangle \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n (\cdot)$$

The “hat notation” (\hat{a}, \hat{b}) is an often-used way to designate estimators.



14.4 Theory of least-squares adjustment



14.4.1 Calculating the solution from the observations

pienimmän
neliösumman
menetelmä

5

6

Presumably the first to use the method of least-squares was Carl Friedrich Gauss, although Adrien-Marie Legendre⁵ has also been claimed to be the inventor of the method. Gauss also carried out extensive geodetic network computations⁶ in Hannover using his method.

In astronomy, the first application of the method was to compute the orbits of asteroids and comets from observations. This, and the adjustment of geodetic networks, were special cases of situations that occur all the time in the life of an observer:

⁵Adrien-Marie Legendre (1752–1833) was a French mathematician, one of the 72 names on the Eiffel Tower ([Eiffel Tower, 72 names](#)).

⁶The numerical work of the network adjustment was carried out by an army of manual, human computers under Gauss’ command. Back then, a “computer” was a human being!



- We have available a body of observations, and we wish to compute from it certain interesting unknowns, in a way which
 - Treats all observations as equally valuable.
 - Makes the deviations of the computed values for the unknowns of interest from their “true values” as small as possible.
- In addition, it would still be desirable that
 - Any gross errors still hiding out in the observations are found and removed.

The *parametric form* of the least-squares adjustment method, which is based on the formation of *observation equations*, is used for this.

14.4.2 The observation equations

Forming the observation equations is done as follows. We write all observations as *linear*⁷ functions of all unknowns:

$$\begin{aligned}
 \underline{\ell}_1 + \underline{v}_1 &= a_{11}\hat{x}_1 + a_{12}\hat{x}_2 + \cdots + a_{1m}\hat{x}_m, \\
 \underline{\ell}_2 + \underline{v}_2 &= a_{21}\hat{x}_1 + a_{22}\hat{x}_2 + \cdots + a_{2m}\hat{x}_m, \\
 &\vdots \\
 \underline{\ell}_n + \underline{v}_n &= a_{n1}\hat{x}_1 + a_{n2}\hat{x}_2 + \cdots + a_{nm}\hat{x}_m,
 \end{aligned}
 \tag{14.1}$$

if there are n observations $\underline{\ell}_i$, n residuals \underline{v}_i , and m unknowns \hat{x}_j .

The system of equations can be conveniently written in the form of a matrix equation

$$\underline{\ell} + \underline{v} = A\hat{x}, \tag{14.2}$$

in which

$$\underline{\ell} = \begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ \vdots \\ \underline{\ell}_n \end{bmatrix}, \quad \underline{v} = \begin{bmatrix} \underline{v}_1 \\ \underline{v}_2 \\ \vdots \\ \underline{v}_n \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad \hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_m \end{bmatrix}.$$

The matrix A is rectangular: $n > m$, it is taller than it is wide. There are more observations, that is equations, than there are unknowns: *redundancy*. The observations $\underline{\ell}$, the unknowns \hat{x} and the residuals \underline{v} are *abstract vectors*, elements of an abstract vector space:

$$\underline{\ell}, \underline{v} \in \mathbb{R}^n, \quad \hat{x} \in \mathbb{R}^m.$$

jännös-
virheiden
vektori

⁷Often, the observation equations of real life are not linear. Then, *linearisation* is usually possible. See section 14.6.



Often, one may assume that all observations $\underline{\ell}_i$, $i = 1, \dots, n$ have the same mean error σ (and a similarly shaped statistical distribution), and that the observations are *statistically independent of each other*, which also means that they do not intercorrelate. This assumption is referred to as *i.i.d.* — “*independent, identically distributed*”.



14.4.3 The normal equations

pienimmän
neliösumman
ratkaisu

From the matrix equation 14.2 we compute the least-squares solution by first multiplying from the left with the matrix A^\top , the transpose of A :

$$A^\top A \hat{\mathbf{x}} = A^\top \underline{\ell} + A^\top \underline{\mathbf{v}}.$$

8 Set⁸

$$A^\top \underline{\mathbf{v}} = 0,$$

yielding for the least-squares solution $\hat{\mathbf{x}}$:

$$(A^\top A) \hat{\mathbf{x}} = A^\top \underline{\ell}. \quad (14.3)$$

This is a system of m equations and m unknowns in vector $\hat{\mathbf{x}}$: the coefficient matrix $A^\top A$ is square.



14.4.4 Solving the normal equations

The solution, or *estimator*, is obtained in for example the following way:

$$\hat{\mathbf{x}} = (A^\top A)^{-1} A^\top \underline{\ell}, \quad (14.4)$$

assuming that the matrix $A^\top A$ can actually be *inverted*, that it is not singular.

The equations 14.3 are known as the *normal equations*.

In the case that all observations have the same mean error σ (and thus the same *variance* σ^2), and that they do not intercorrelate, the solution 14.4 is *optimal* in the least-squares sense.

Of course, solving the system of equations by traditional means, without matrices, to find the elements of the solution vector $\hat{\mathbf{x}}$, being the unknowns,

⁸If one writes $\eta \stackrel{\text{def}}{=} A\xi$, the following holds:

$$\langle \eta \cdot \underline{\mathbf{v}} \rangle = \langle (A\xi) \cdot \underline{\mathbf{v}} \rangle = \xi^\top A^\top \underline{\mathbf{v}} = 0$$

for an *arbitrary vector* ξ ; we say that the sub-space of vectors $A\xi$ of the space of observables (the “solution space”, spanned by the columns of the matrix A) is *perpendicular* upon the sub-space of residuals. This is where the term “normal equations” comes from.



is also a readily useable method: write the normal equations 14.3 in the following way, which is well-suited to computer coding:

$$\sum_{i=1}^m \left(\sum_{j=1}^n a_{jk} a_{ji} \right) \hat{x}_i = \left(\sum_{j=1}^n a_{jk} \ell_j \right), \quad k = 1, \dots, m.$$

This is a system of m linear equations in m unknowns \hat{x}_i , for the solution of which numerical standard methods and software libraries are on offer. The greatest challenge is usually finding suitable observation equations 14.1 in a concrete measurement situation.

The solutions for the average and linear regression presented above, sections 14.2 and 14.3, are special cases of the general adjustment solution, as we shall show. tasoitus

14.4.5 Assessing the precision

When one computes the least-squares solution with the equation

$$\hat{\mathbf{x}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \underline{\ell},$$

one can also, with the *propagation law of variances*, obtain the precision of estimator $\hat{\mathbf{x}}$. varianssien
kasautumislaki

We reiterate the assumption that the variance matrix of the observations $\underline{\ell}$ is $\text{Var}\{\underline{\ell}\} = \sigma^2 \mathbf{I}$, so the observations do not correlate with each other and are all equally precise.

If we define the linear operator

$$\mathbf{L} \stackrel{\text{def}}{=} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top,$$

we obtain, based on the linear dependence,

$$\begin{aligned} \text{Var}\{\hat{\mathbf{x}}\} &= \Sigma_{\mathbf{xx}} = \mathbf{L} \text{Var}\{\underline{\ell}\} \mathbf{L}^\top = \\ &= (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \cdot \sigma^2 \mathbf{I} \cdot \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1} = \sigma^2 (\mathbf{A}^\top \mathbf{A})^{-1}. \end{aligned} \quad (14.5)$$

This interesting result tells us, that the matrix quantity $(\mathbf{A}^\top \mathbf{A})^{-1}$ represents the *propagation of the mean error* σ of the observations into the variances of the end result of the adjustment $\hat{\mathbf{x}}$.

The matrix $\mathbf{N} = P_{\mathbf{xx}} \stackrel{\text{def}}{=} \mathbf{A}^\top \mathbf{A}$ is called the *weight matrix of the unknowns* or *normal matrix*, and its inverse, the matrix $\mathbf{Q}_{\mathbf{xx}} \stackrel{\text{def}}{=} (\mathbf{A}^\top \mathbf{A})^{-1}$, is called the *weight-coefficient matrix* of the unknowns (Baarda, 1981).





14.5 Examples of the least-squares method

Both average and linear regression are great examples of least-squares adjustment methods: practical and still relatively simple. In the following, we go through them step by step, showing how they are special cases of the general least-squares method.



14.5.1 The average as least-squares adjustment

Observe the same quantity x directly n times:

$$\begin{aligned} \underline{x}_1 + \underline{v}_1 &= \hat{x} \\ \underline{x}_2 + \underline{v}_2 &= \hat{x} \\ &\vdots \\ \underline{x}_n + \underline{v}_n &= \hat{x} \end{aligned} \quad (14.6)$$

Here, $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ are individual observation values, \hat{x} is the estimator of the unknown quantity x , and $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$ are the residuals of the observations.

The secret to formulating a suitable adjustment procedure is: find the *standard form* of the system of observation equations,

$$\underline{\ell} + \underline{v} = A\hat{x}.$$

In the present case, success means choosing

$$\underline{\ell} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix}, \quad \underline{v} = \begin{bmatrix} \underline{v}_1 \\ \underline{v}_2 \\ \vdots \\ \underline{v}_n \end{bmatrix}, \quad A = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \left. \vphantom{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}} \right\} \begin{matrix} n \text{ times} \end{matrix}, \quad \hat{x} = \begin{bmatrix} \hat{x} \end{bmatrix}.$$

Verify that this really agrees with equations 14.6.

The normal equations are now

$$\overbrace{A^T A}^N \hat{x} = \overbrace{A^T \underline{\ell}}^b,$$

in which the normal matrix N is

$$N = A^T A = \overbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}}^{n \text{ times}} \left. \vphantom{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}} \right\} \begin{matrix} n \text{ times} \end{matrix} = \begin{bmatrix} n \end{bmatrix}$$



and the right-hand side vector

$$\underline{b} = A^T \underline{\ell} = \overbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}}^{n \text{ times}} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix} = \begin{bmatrix} n \\ \sum_{i=1}^n \underline{x}_i \end{bmatrix}.$$

So, the solution is obtained as follows:

$$\overbrace{\begin{bmatrix} N \\ n \end{bmatrix}}^N \hat{\underline{x}} = \overbrace{\begin{bmatrix} \underline{b} \\ \sum_{i=1}^n \underline{x}_i \end{bmatrix}}^n \implies \hat{\underline{x}} = \overbrace{\frac{1}{n}}^{N^{-1}} \overbrace{\sum_{i=1}^n \underline{x}_i}^{\underline{b}},$$

the classical equation for the average!

Deriving the *precision*, or mean error, of the solution is not hard either. The assumption that the mean errors of the observations \underline{x}_i , $i = 1, \dots, n$, are all the same σ and that the observations are uncorrelated — a precondition for using the equation for the average — means that the variance matrix of the observation vector $\underline{\ell}$ is

$$\text{Var}\{\underline{\ell}\} = \sigma^2 I = \sigma^2 \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$

a matrix of size $n \times n$. Then, according to equation 14.5, the variance matrix of the vector of unknowns⁹ is

$$\text{Var}\{\hat{\underline{x}}\} = \sigma^2 (A^T A)^{-1} = \frac{1}{n} \sigma^2,$$

and the mean error of the unknown is the square root of this:

$$\sigma_x = \frac{\sigma}{\sqrt{n}}.$$

14.5.2 Linear regression as least-squares adjustment

Write the observation equations of linear regression into the form

$$\underline{y}_i + \underline{v}_i = \hat{a} + \hat{b}x_i,$$

in which every pair (x_i, \underline{y}_i) , $i = 1, \dots, n$ is one observation, and the coefficients \hat{a} and \hat{b} of the straight line to be fitted are to be determined.

If we write the vector of observations, the vector of residuals, the vector of unknowns, and again (essential!) the design matrix as

rakennematriisi

⁹The “vector” $\hat{\underline{x}}$ has here only one element, as the “matrix” $\text{Var}\{\hat{\underline{x}}\}$ has, too.



$$\underline{\ell} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_n \end{bmatrix}, \quad \underline{v} = \begin{bmatrix} \underline{v}_1 \\ \underline{v}_2 \\ \vdots \\ \underline{v}_n \end{bmatrix}, \quad \hat{\underline{x}} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}, \quad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix},$$

one may write this system into the standard form of observation equations¹⁰

$$\underline{\ell} + \underline{v} = A\hat{\underline{x}}.$$

In the normal equations

$$\overbrace{A^T A}^N \hat{\underline{x}} = \overbrace{A^T \underline{\ell}}^{\underline{b}}$$

the normal matrix is

$$N = A^T A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

and the right-hand side vector

$$\underline{b} = A^T \underline{\ell} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n \underline{y}_i \\ \sum_{i=1}^n x_i \underline{y}_i \end{bmatrix}.$$

From this, the solution is obtained by inversion of the above 2×2 matrix. However, this may also be done simply by elimination and back-substitution. The system of equations is

$$\begin{aligned} n \cdot \hat{a} + \left(\sum_{i=1}^n x_i \right) \cdot \hat{b} &= \sum_{i=1}^n \underline{y}_i, \\ \left(\sum_{i=1}^n x_i \right) \cdot \hat{a} + \left(\sum_{i=1}^n x_i^2 \right) \cdot \hat{b} &= \sum_{i=1}^n x_i \underline{y}_i. \end{aligned}$$

Subtract the first equation from the second after multiplication by the factor $\frac{1}{n} \sum_{i=1}^n x_i$, yielding

$$\left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) \cdot \hat{b} = \sum_{i=1}^n x_i \underline{y}_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n \underline{y}_i,$$

¹⁰The use of the letter x may be confusing here, and has nothing to do with the x_i !





TABLE 14.1. Measurement results for linear regression.

	$i \rightarrow 1$	2	3	4	5	$\sum_{i=1}^5$
x_i	1.51	2.44	3.34	4.41	5.05	16.75
y_i	2.32	3.12	3.57	3.93	4.15	17.09
x_i^2	2.28	5.95	11.16	19.45	25.50	64.34
$x_i y_i$	3.50	7.61	11.92	17.33	20.96	61.32

from which

$$\hat{b} = \left(n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) / \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right).$$

Now \hat{a} is obtained by back-substitution:

$$\hat{a} = \frac{1}{n} \left(\sum_{i=1}^n y_i - \left(\sum_{i=1}^n x_i \right) \cdot \hat{b} \right).$$

The expressions found are equivalent to those given in section 14.3.



14.5.3 Calculation example of linear regression

The measurement results are given in table 14.1. A graphical presentation of the measurements is given in figure 14.5.

The needed sums are precomputed in the table:

$$\sum_{i=1}^n x_i, \quad \sum_{i=1}^n y_i, \quad \sum_{i=1}^n x_i^2, \quad \sum_{i=1}^n x_i y_i.$$

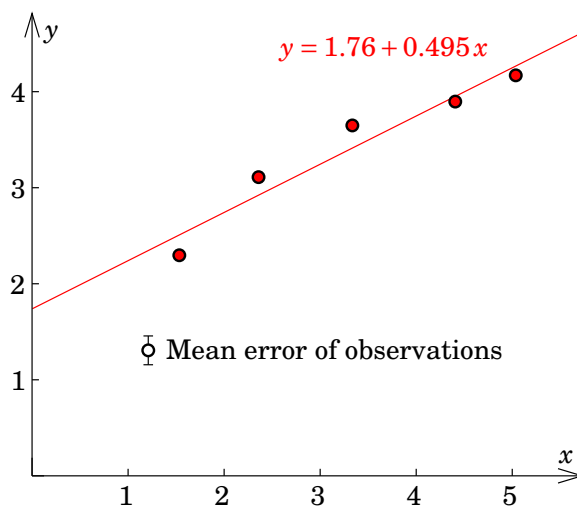


FIGURE 14.5. Calculation example of linear regression.



From this

$$\hat{b} = \frac{5 \cdot 61.32 - 16.75 \cdot 17.09}{5 \cdot 64.34 - 16.75^2} = \frac{20.3425}{41.1375} = 0.495,$$

$$\hat{a} = \frac{1}{5} (17.09 - 16.75 \cdot \hat{b}) = 1.76.$$

This solution has been plotted into figure 14.5.

The normal matrix, or *weight matrix of the unknowns*, is obtained as follows:

$$N = P_{xx} = A^T A = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} 5.00 & 16.75 \\ 16.75 & 64.34 \end{bmatrix}$$

and its inverse is

$$Q_{xx} = (A^T A)^{-1} = \begin{bmatrix} 1.5640 & -0.4072 \\ -0.4072 & 0.1215 \end{bmatrix},$$

the *weight-coefficient matrix*. The variance matrix

$$\Sigma_{xx} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} = \sigma^2 Q_{xx} = \sigma^2 \begin{bmatrix} 1.5640 & -0.4072 \\ -0.4072 & 0.1215 \end{bmatrix}$$

is obtained from this, from which are obtained the mean errors σ_a and σ_b as the square roots of the diagonal elements:

$$\hat{a} = 1.76 \pm 1.25 \sigma,$$

$$\hat{b} = 0.495 \pm 0.349 \sigma.$$

Here, σ is the *a priori* (given in advance) mean error of a single \underline{y} value, the *mean error of unit weight*.

painoyksikön
keskivirhe



14.6 Linearisation of geodetic models

In geodesy, as more generally in science, relationships are found between two quantities that behave *non-linearly*. Examples of this are the relationship between observables and unknowns, or between co-ordinates in two different co-ordinate frames.

Many theories, however, like the least-squares adjustment method, are based on *linear* equations, the mathematics of which is essentially simpler. The law of propagation of errors (variances) also applies only to linear relationships between quantities.

virheiden
kasautumislaki



In practice, a formally non-linear relationship, for example between point co-ordinates and the measured direction to a point, is often almost linear within the uncertainty area of the point location. Geodetic measurements are exceptionally precise: the location uncertainty of a point may be mere centimetres when the distance between points can be hundreds of metres or kilometres. In that case, instead of looking at the original quantities, one can look at the relationship between small variations or differences in these quantities — which will be almost linear. This will be demonstrated using a Taylor series expansion.

 14.6.1 The scalar case

Normally if we have two quantities between which exists a functional relationship

$$y = f(x),$$

we may linearise it by choosing an approximate value x_0 and expanding the function into a Taylor series in the neighbourhood of the approximate value. We obtain

$$y = f(x_0) + \left. \frac{df}{dx} \right|_{x=x_0} (x - x_0) + \dots$$

or

$$y - y_0 \approx a (x - x_0), \tag{14.7}$$

in which

$$y_0 \stackrel{\text{def}}{=} f(x_0), \quad a \stackrel{\text{def}}{=} \left. \frac{df}{dx} \right|_{x=x_0}.$$

This may be written as

$$\Delta y = a \Delta x,$$

which is often abbreviated to

$$y = ax,$$

as long as it is remembered that now x and y are “linearised quantities”, difference quantities $\Delta x \stackrel{\text{def}}{=} x - x_0$ and $\Delta y \stackrel{\text{def}}{=} y - y_0$ reckoned from the approximate values x_0 and y_0 .

 14.6.2 The vector case

If there are two vectorial quantities,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m,$$



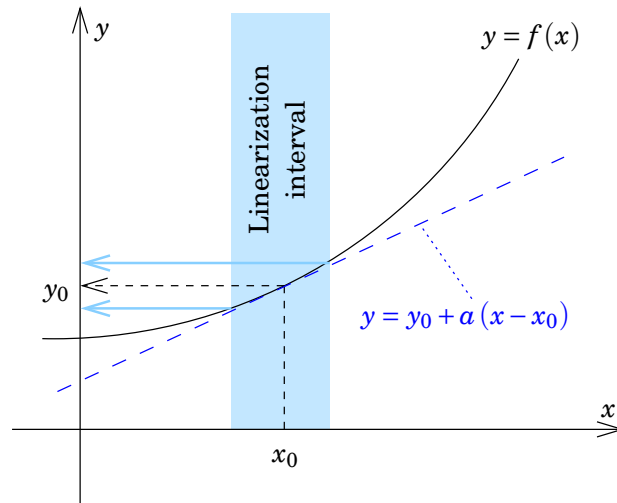


FIGURE 14.6. One-dimensional mapping and linearisation.

between which exists a functional relationship

$$y = F(x) = F(x_1, x_2, \dots, x_n)$$

or

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{bmatrix} = \begin{bmatrix} F_1(x_1, x_2, \dots, x_n) \\ F_2(x_1, x_2, \dots, x_n) \\ \vdots \\ F_m(x_1, x_2, \dots, x_n) \end{bmatrix},$$

likiarvojen vektori it becomes complicated. We can choose a *vector of approximate values*

$$x_0 \stackrel{\text{def}}{=} \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix},$$

and a corresponding vector of approximate values $y_0 \stackrel{\text{def}}{=} F(x_0)$, after which again

$$\begin{aligned} y = y_0 &+ \left. \frac{\partial F(x_1, x_2, \dots, x_n)}{\partial x_1} \right|_{x=x_0} (x_1 - x_1^{(0)}) + \\ &+ \left. \frac{\partial F(x_1, x_2, \dots, x_n)}{\partial x_2} \right|_{x=x_0} (x_2 - x_2^{(0)}) + \\ &\vdots \\ &+ \left. \frac{\partial F(x_1, x_2, \dots, x_n)}{\partial x_n} \right|_{x=x_0} (x_n - x_n^{(0)}) + \dots \end{aligned}$$



or

$$\begin{aligned}
 y_i = y_i^{(0)} &+ \left. \frac{\partial F_i(x_1, x_2, \dots, x_n)}{\partial x_1} \right|_{x=x_0} (x_1 - x_1^{(0)}) + \\
 &+ \left. \frac{\partial F_i(x_1, x_2, \dots, x_n)}{\partial x_2} \right|_{x=x_0} (x_2 - x_2^{(0)}) + \\
 &\vdots \\
 &+ \left. \frac{\partial F_i(x_1, x_2, \dots, x_n)}{\partial x_n} \right|_{x=x_0} (x_n - x_n^{(0)}) + \dots, \quad i = 1, \dots, m.
 \end{aligned}$$

In this equation there are m different rows, and in every row there are n different (linear) terms. As a summary of this system of equations, we write the following matrix equation:

$$y = y_0 + A(x - x_0) + \dots,$$

in which the matrix A is

$$A = \left[\begin{array}{cccc} \frac{\partial}{\partial x_1} F_1 & \frac{\partial}{\partial x_2} F_1 & \dots & \frac{\partial}{\partial x_n} F_1 \\ \frac{\partial}{\partial x_1} F_2 & \frac{\partial}{\partial x_2} F_2 & \dots & \frac{\partial}{\partial x_n} F_2 \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_1} F_m & \frac{\partial}{\partial x_2} F_m & \dots & \frac{\partial}{\partial x_n} F_m \end{array} \right]_{x=x_0}.$$

This matrix is the *matrix of Jacobi*¹¹ of the vector mapping $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ between the two abstract vector spaces \mathbb{R}^n and \mathbb{R}^m . The matrix describes *locally*, in a neighbourhood of point $x = x_0$, the way in which small “perturbances” in vector x propagate into vector y :

$$\Delta y = y - y_0 \approx A(x - x_0) = A \Delta x,$$

with definitions $\Delta x \stackrel{\text{def}}{=} x - x_0$ and $\Delta y \stackrel{\text{def}}{=} y - y_0$. So, the map between the difference quantities Δx and Δy is locally linear. This is referred to as *linearisation*.

In the general case, $m \neq n$. In the special case $m = n$ we may think that the mapping F has an inverse mapping $G = F^{-1}$, for which

$$\Delta x = G(\Delta y).$$

Locally, in the neighbourhood of the approximate point x_0 , we may say

¹¹Carl Gustav Jacob Jacobi, 1804–1851, was a Jewish German mathematician, professor at the University of Königsberg 1827–1842.



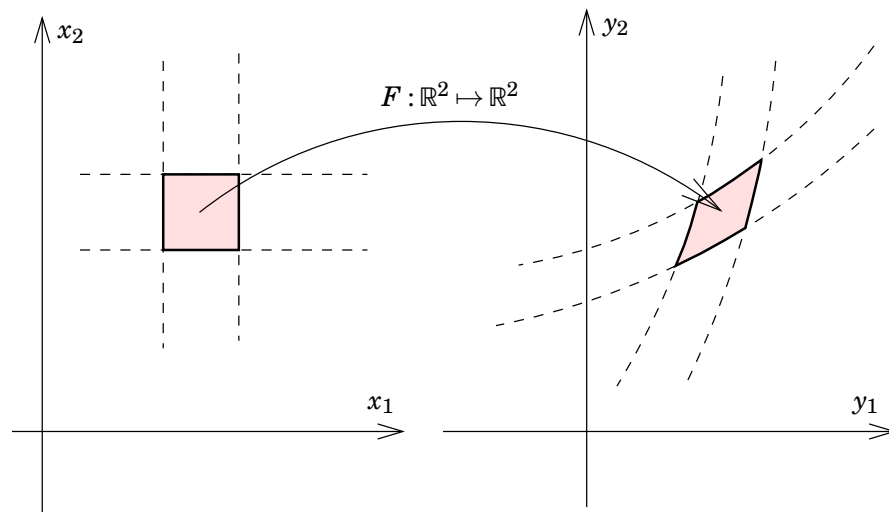


FIGURE 14.7. A two-dimensional mapping.

of this: if the matrix A is *singular*, meaning its determinant $\det A = 0$, in a neighbourhood of point x_0 , this means that the mapping F does not in that neighbourhood have an inverse mapping. This again means that there may be many (in fact, infinitely many) different values Δx having the same image $\Delta y = F(\Delta x)$. On the other hand, if $\det A \neq 0$ in the point x_0 , such an inverse mapping does (in a sufficiently small neighbourhood of approximate point y_0) exist.

Interpretation The determinant $\det A$ describes the way in which *volumes* are mapped under the vector mapping F .

suunnikas

If, for example, $n = m = 2$, it describes how the surface area of a small square in the \mathbb{R}^n space is mapped to the surface area of a parallelogram in the \mathbb{R}^m space. The determinant is the *ratio* of these two surface areas.

suuntaissärmiö

If $n = m = 3$, the determinant similarly equals the ratio of the volumes of a small cube in \mathbb{R}^n space and of its corresponding parallelepiped in \mathbb{R}^m space.

If the ratio is zero, then apparently the square is “squeezed” to a line segment, and the cube to a parallelogram, and the map is thus singular.



14.6.3 Linearisation of observation equations

Let us consider as an example the functional relationship between unknowns x and observables ℓ , which in a realistic observation geometry is rarely linear. We have to *linearise*: let the non-linear observation



equations be

$$\underline{\ell} + \underline{v} = F(\widehat{x}), \tag{14.8}$$

in which $F(\cdot)$ is a multidimensional non-linear “observation function”.

The equations are linearised by expanding them into a Taylor series around roughly estimated solution co-ordinates (“approximate values”), and using only the first-degree terms of the series. If the approximate co-ordinates used are not good enough, we end up *iterating*.

Choose the *vector of approximate values* x_0 and compatibly the vector ℓ_0 so that the following applies for them:

$$\ell_0 = F(x_0). \tag{14.9}$$

So, if the number of unknowns is m and the number of observations n :

$$\ell_i^{(0)} = F_i(x_1^{(0)}, x_2^{(0)}, \dots, x_m^{(0)}), \quad i = 1, \dots, n.$$

This is subtracted from equation 14.8, and we do a Taylor series expansion, retaining only the linear terms:

$$\begin{aligned} (\underline{\ell}_i - \ell_i^{(0)}) + \underline{v}_i &= F_i(\widehat{x}_1, \widehat{x}_2, \dots, \widehat{x}_m) - F_i(x_1^{(0)}, x_2^{(0)}, \dots, x_m^{(0)}) \approx \\ &\approx \sum_{j=1}^m \left. \frac{\partial F_i}{\partial x_j} \right|_{x=x_0} (\widehat{x}_j - x_j^{(0)}). \end{aligned}$$

Call

$$A_{ij} \stackrel{\text{def}}{=} \left. \frac{\partial F_i}{\partial x_j} \right|_{x=x_0}, \quad i = 1, \dots, n, \quad j = 1, \dots, m, \tag{14.10}$$

the elements of the *second-order design matrix*. The matrix itself is then **rakennematriisi**

$$A = \left[\begin{array}{cccc} \frac{\partial}{\partial x_1} F_1 & \frac{\partial}{\partial x_2} F_1 & \dots & \frac{\partial}{\partial x_m} F_1 \\ \frac{\partial}{\partial x_1} F_2 & \frac{\partial}{\partial x_2} F_2 & \dots & \frac{\partial}{\partial x_m} F_2 \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_1} F_n & \frac{\partial}{\partial x_2} F_n & \dots & \frac{\partial}{\partial x_m} F_n \end{array} \right] \Bigg|_{x_1=x_1^{(0)}, x_2=x_2^{(0)}, \dots, x_m=x_m^{(0)}}.$$

If we call

$$\Delta \ell \stackrel{\text{def}}{=} (\underline{\ell} - \ell_0) = (\underline{\ell} - F(x_0)), \quad \Delta \widehat{x} \stackrel{\text{def}}{=} (\widehat{x} - x_0)$$

“replacement” or “linearised” observables and unknowns), we obtain for the linearised observation equations

$$\Delta \underline{\ell} + \underline{v} = A \Delta \widehat{x}. \tag{14.11}$$



The least-squares solution to be computed here minimises the sum of squares of residuals $\underline{v}^T \mathbf{Q}_{\ell\ell}^{-1} \underline{v}$, wherefrom the name comes. The matrix $\mathbf{Q}_{\ell\ell}$ is the *weight-coefficient matrix*¹² characterising the relative precision and possible statistical interdependence (correlation) of the observations.

From equation 14.11, often the Δ are left off for the sake of writing convenience. The Δ quantities are typically much smaller than the “whole” quantities. Therefore the numerics work well even if the elements of the matrix A are not exact.

Equation 14.9 however must always be calculated *precisely*, and with a sufficient number of decimals.



14.7 Propagation of variances

If the stochastic quantity \underline{y} is a linear function of the stochastic quantity \underline{x} :

$$\underline{y} = \lambda \underline{x},$$

we may also write

$$\sigma_y = \lambda \sigma_x,$$

in which σ_x and σ_y are the mean errors of quantities x and y . We may also write

$$E\{\underline{y}\} = E\{\lambda \underline{x}\} = \lambda E\{\underline{x}\}$$

odotusarvojen kasautumislaki (“propagation of expectancies”), in which $E\{\cdot\}$ is the expectancy operator. The expectancy is a *linear operator*.

If we define the variance as follows:

$$\text{Var}\{\underline{x}\} = \sigma_x^2 \stackrel{\text{def}}{=} E\left\{(\underline{x} - E\{\underline{x}\})^2\right\},$$

it follows that

$$\sigma_y^2 = \lambda^2 \sigma_x^2.$$

This is the *law of propagation of variances* for simple stochastic quantities.

If the stochastic quantities

$$\underline{x} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_m \end{bmatrix}$$

¹²One writes $\text{Var}\{\underline{\ell}\} = \Sigma_{\ell\ell} = \sigma^2 \mathbf{Q}_{\ell\ell}$, in which $\Sigma_{\ell\ell}$ is the variance matrix of the observations, and σ the *mean error of unit weight*.



have several components — in other words, they are abstract vectorial quantities — it holds, if $\underline{y} = \Lambda \underline{x}$, that

$$E\{\underline{y}\} = \Lambda E\{\underline{x}\} \tag{14.12}$$

and

$$\text{Var}\{\underline{y}\} = \Lambda \text{Var}\{\underline{x}\} \Lambda^\top, \tag{14.13}$$

in which now Λ and the variances are *matrices*.

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda_{m1} & \lambda_{m2} & \cdots & \lambda_{mn} \end{bmatrix}$$

is an $m \times n$ size matrix.

$$\text{Var}\{\underline{x}\} = \Sigma_{xx} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_n} \\ \sigma_{x_2x_1} & \sigma_{x_2}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{x_nx_1} & \cdots & \cdots & \sigma_{x_n}^2 \end{bmatrix}$$

is a square matrix of size $n \times n$, and

$$\text{Var}\{\underline{y}\} = \Sigma_{yy} = \begin{bmatrix} \sigma_{y_1}^2 & \sigma_{y_1y_2} & \cdots & \sigma_{y_1y_m} \\ \sigma_{y_2y_1} & \sigma_{y_2}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{y_my_1} & \cdots & \cdots & \sigma_{y_m}^2 \end{bmatrix}$$

is a square matrix of size $m \times m$. Here, the variances are

$$\sigma_{x_i}^2 = \text{Var}\{x_i\} = E\left\{ (x_i - E\{x_i\})^2 \right\}$$

and the covariances

$$\sigma_{x_ix_j} = \text{Cov}\{x_i, x_j\} = E\left\{ (x_i - E\{x_i\})(x_j - E\{x_j\}) \right\},$$

and similarly for the components of \underline{y} .

Equation 14.13 is called the *general law of propagation of variances*. It is a generalisation of equation 2.4, already derived in subsection 2.4.4, for an arbitrary number of variables. The linearity property assumed here may be obtained by linearisation if needed, as discussed earlier.

varienssien
kasautumislaki



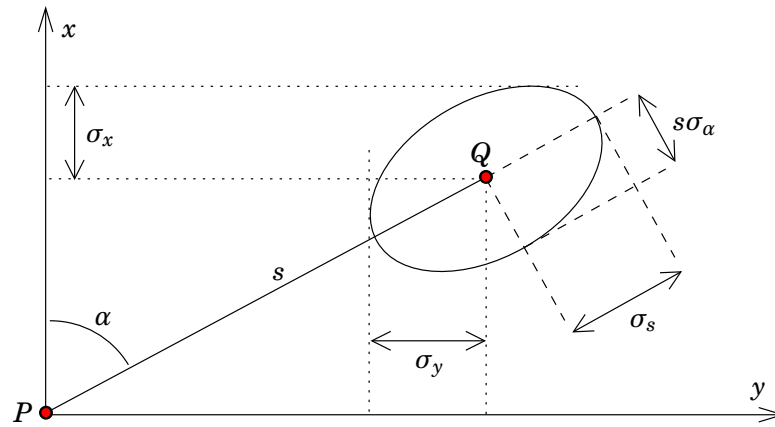


FIGURE 14.8. Quantities related to the error ellipse.



14.8 Example: the forward geodetic problem

geodeettinen
päätehtävä

As an application of the law of propagation of variances, we present the *forward geodetic problem* in the plane, in which the known uncertainties of direction and distance measurement are *propagated* into co-ordinate uncertainties of an unknown point.

Given measured quantities s and α as well as the co-ordinates (x_P, y_P) of the starting point P , the problem is determining the co-ordinates of the unknown point Q , figure 14.8:

$$x_Q = x_P + s \cdot \cos \alpha, \quad y_Q = y_P + s \cdot \sin \alpha.$$

The problem is solved in the following way. Choose approximate values $s^{(0)}, \alpha^{(0)}$:

$$s = s^{(0)} + \Delta s, \quad \alpha = \alpha^{(0)} + \Delta \alpha,$$

and write a Taylor series expansion:

$$\begin{aligned} x_Q &\approx x_P + s^{(0)} \cos \alpha^{(0)} + \Delta s \cos \alpha^{(0)} + s^{(0)} \left. \frac{\partial \cos \alpha}{\partial \alpha} \right|_{\alpha=\alpha^{(0)}} \Delta \alpha = \\ &= \underbrace{x_P + s^{(0)} \cos \alpha^{(0)}}_{x^{(0)}} + \overbrace{\left[\cos \alpha^{(0)} \quad -s^{(0)} \sin \alpha^{(0)} \right]}^{\Delta x} \begin{bmatrix} \Delta s \\ \Delta \alpha \end{bmatrix}, \end{aligned}$$

and in the same way

$$y_Q \approx \underbrace{y_P + s^{(0)} \sin \alpha^{(0)}}_{y^{(0)}} + \overbrace{\left[\sin \alpha^{(0)} \quad s^{(0)} \cos \alpha^{(0)} \right]}^{\Delta y} \begin{bmatrix} \Delta s \\ \Delta \alpha \end{bmatrix}.$$



Now, dropping, but remembering, the approximation labels (0), and turning the vectors x and y into random or *stochastic* quantities:

$$\underline{y} \stackrel{\text{def}}{=} \begin{bmatrix} \underline{\Delta x} \\ \underline{\Delta y} \end{bmatrix}, \quad \underline{x} \stackrel{\text{def}}{=} \begin{bmatrix} \underline{\Delta s} \\ \underline{\Delta \alpha} \end{bmatrix},$$

and

$$\Lambda = \begin{bmatrix} \cos \alpha & -s \sin \alpha \\ \sin \alpha & s \cos \alpha \end{bmatrix}, \quad \text{Var}\{\underline{x}\} = \begin{bmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_\alpha^2 \end{bmatrix},$$

the above equations may be written compactly as

$$\underline{y} = \Lambda \underline{x}.$$

The variance matrix is

$$\begin{aligned} \text{Var}\{\underline{y}\} &= \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} = \Lambda \text{Var}\{\underline{x}\} \Lambda^\top = \\ &= \begin{bmatrix} \cos \alpha & -s \sin \alpha \\ \sin \alpha & s \cos \alpha \end{bmatrix} \begin{bmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_\alpha^2 \end{bmatrix} \begin{bmatrix} \cos \alpha & \sin \alpha \\ -s \sin \alpha & s \cos \alpha \end{bmatrix} \end{aligned}$$

and the elements are

$$\begin{aligned} \sigma_x^2 &= \sigma_s^2 \cos^2 \alpha + \sigma_\alpha^2 s^2 \sin^2 \alpha, \\ \sigma_y^2 &= \sigma_s^2 \sin^2 \alpha + \sigma_\alpha^2 s^2 \cos^2 \alpha, \\ \sigma_{xy} &= \cos \alpha \sin \alpha (\sigma_s^2 - s^2 \sigma_\alpha^2), \end{aligned}$$

computed using the law 14.13 of propagation of variances.¹³

13

By substitution of

$$\cos \alpha = \frac{x_Q - x_P}{s}, \quad \sin \alpha = \frac{y_Q - y_P}{s},$$

¹³If we express the variance of direction α in gon, we may substitute into all equations

$$\sigma_\alpha^2 = \left(\frac{\sigma_\alpha [\text{g}]}{\rho} \right)^2,$$

in which ρ is the size of a radian in the degree unit in question, in this case $\rho = 63.661977236758$. Similarly if one uses seconds of arc: then

$$\sigma_\alpha^2 = \left(\frac{\sigma_\alpha ["]}{\rho} \right)^2,$$

in which now $\rho = 57.295779513 \times 60 \times 60 = 206264.806247$.



an alternative form is still obtained:

$$\begin{aligned}\sigma_x^2 &= \text{Var}\{\Delta x\} = \left(\frac{x_Q - x_P}{s}\right)^2 \sigma_s^2 + (y_Q - y_P)^2 \sigma_\alpha^2, \\ \sigma_y^2 &= \text{Var}\{\Delta y\} = \left(\frac{y_Q - y_P}{s}\right)^2 \sigma_s^2 + (x_Q - x_P)^2 \sigma_\alpha^2, \\ \sigma_{xy} &= \text{Cov}\{\Delta x, \Delta y\} = \left(\left(\frac{\sigma_s}{s}\right)^2 - \sigma_\alpha^2\right) (x_Q - x_P)(y_Q - y_P).\end{aligned}\quad (14.14)$$

This is how mean errors of the observations σ_s, σ_α are converted into co-ordinate mean errors σ_x, σ_y . The precision is affected both by the observational precisions σ_s, σ_α and by the geometry s, α .



14.8.1 The error ellipse and its eigenvalue problem

The *error ellipse* or uncertainty ellipse is the statistical area of uncertainty of the location solution (x, y) of a point. This can be used in statistical testing.

As a measure of point precision, there exists a suitable quantity that is independent of the directions of the co-ordinate axes. The error ellipse is really a *visual representation of the point variance matrix*. The variance matrix of the co-ordinates (x_Q, y_Q) of point Q in figure 14.8 — or equivalently, of the linearised co-ordinates $(\Delta x, \Delta y)$ — may be written as

$$\Sigma = \text{Var} \left\{ \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right\} = \begin{bmatrix} \text{Var}\{\Delta x\} & \text{Cov}\{\Delta x, \Delta y\} \\ \text{Cov}\{\Delta x, \Delta y\} & \text{Var}\{\Delta y\} \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}.$$

The *invariants* of this matrix are its eigenvalues and -vectors: the solutions of the eigenvalue problem $(\Sigma - \lambda I)\bar{\mathbf{x}} = 0$, $(\lambda_i, \bar{\mathbf{x}}_i)$, $i = 1, 2$. If we rotate the co-ordinate axes so, that they are oriented along the main axes of the ellipse, we obtain

$$\Sigma = \begin{bmatrix} s^2 \sigma_\alpha^2 & \\ & \sigma_s^2 \end{bmatrix}$$

and obviously $\lambda_1 = s^2 \sigma_\alpha^2$ and $\lambda_2 = \sigma_s^2$.

More generally, one solves the determinant equation

$$\det(\Sigma - I\lambda) = \det \begin{bmatrix} \Sigma_{11} - \lambda & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} - \lambda \end{bmatrix} = \det \begin{bmatrix} \sigma_x^2 - \lambda & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 - \lambda \end{bmatrix} = 0.$$

This is called the *characteristic polynomial*:

$$\det(\Sigma - I\lambda) = (\sigma_x^2 - \lambda)(\sigma_y^2 - \lambda) - \sigma_{xy}^2 = 0,$$



so

$$\lambda^2 - (\sigma_x^2 + \sigma_y^2) \lambda + (\sigma_x^2 \sigma_y^2 - \sigma_{xy}^2) = 0, \quad (14.15)$$

a standard-issue quadratic equation. The solutions or eigenvalues are ominaisarvo

$$\begin{aligned} \lambda_{1,2} &= \frac{1}{2} \left((\sigma_x^2 + \sigma_y^2) \pm \sqrt{(\sigma_x^2 + \sigma_y^2)^2 - 4(\sigma_x^2 \sigma_y^2 - \sigma_{xy}^2)} \right) = \\ &= \frac{1}{2} \left((\sigma_x^2 + \sigma_y^2) \pm \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2} \right) = \\ &= \frac{1}{2} (\sigma_x^2 + \sigma_y^2) \pm \sqrt{\frac{1}{4} (\sigma_x^2 - \sigma_y^2)^2 + \sigma_{xy}^2}, \quad (14.16) \end{aligned}$$

and the semi-major and semi-minor axes of the error ellipse are $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$. iso- ja pikkuakselin puolikkaat

The *directions* of the axes may also be determined: look at the linear combination of co-ordinates

$$z(\theta) = x \sin \theta + y \cos \theta,$$

which is a function of the direction angle θ .

The propagation law of variances yields

$$\text{Var}\{z\} = \sigma_x^2 \sin^2 \theta + \sigma_y^2 \cos^2 \theta + 2\sigma_{xy} \sin \theta \cos \theta.$$

The axis directions of the visual ellipse are *stationary values* of this function of direction angle θ ,

$$\frac{d}{d\theta} \text{Var}\{z\} = 0.$$

By differentiation

$$\begin{aligned} 2 \sin \theta \cos \theta (\sigma_x^2 - \sigma_y^2) + 2 (\cos^2 \theta - \sin^2 \theta) \sigma_{xy} &= 0 \\ \implies \sin 2\theta (\sigma_x^2 - \sigma_y^2) + 2 \cos 2\theta \cdot \sigma_{xy} &= 0, \end{aligned}$$

and

$$\begin{aligned} \theta &= \frac{1}{2} \arctan \left(-\frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2} \right) + k \cdot 100^g = \\ &= \arctan \frac{-\sigma_{xy}}{\frac{1}{2} (\sigma_x^2 - \sigma_y^2) + \sqrt{\frac{1}{4} (\sigma_x^2 - \sigma_y^2)^2 + \sigma_{xy}^2}} + k \cdot 100^g, \end{aligned}$$

using the half-angle formula for the arc tangent. ¹⁴ 14

One obtains still from equation 14.16:

$$\lambda_1 + \lambda_2 = \text{Var}\{x\} + \text{Var}\{y\} = \sigma_x^2 + \sigma_y^2 \quad (14.17)$$



15 and¹⁵

$$\lambda_1 \lambda_2 = \det \Sigma = \sigma_x^2 \sigma_y^2 - \sigma_{xy}^2. \quad (14.18)$$

The quantities 14.17 and 14.18 are *invariants* — so, always the same, no matter how the co-ordinate axes are oriented — and quantity 14.17, called the *point variance* of point P , σ_P^2 , is a particularly suitable measure of point precision:

$$\sigma_P^2 = \sigma_x^2 + \sigma_y^2.$$

The *point mean error* σ_P is the square root of this point variance.



14.9 Observables and observation equations in practice

vinoetäisyys

Here we present the observables of classical geodesy: measurements of horizontal directions, zenith angles and slant ranges, in the form of linearised observation equations. The GPS observables presented already above, equations 12.1 and 12.2, are special cases of slant-range measurement.

All the observables presented are of observations between two points, from one point to the other. The observation equations describe the dependence of the observables on the unknowns, in this case the three-dimensional co-ordinates of the points.

We only present the observation equations of single observations: in a real-life situation one must combine the observation equations for all observations, and all unknowns, into a system of equations before proceeding to solution.



14.9.1 Slant-range measurement

tähys Let the co-ordinates of instrument and signal be (x_1, y_1, z_1) and (x_2, y_2, z_2) . Then, the functional model, equation 14.8, is

$$s = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \quad (14.19)$$

¹⁴This is how one avoids division by zero in the edge case $\sigma_x^2 = \sigma_y^2$.

¹⁵Because the equation 14.15 can be written into the form

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1 \lambda_2 = 0,$$

from which $\lambda_1 + \lambda_2 = \sigma_x^2 + \sigma_y^2$, the *trace* of the matrix Σ , and $\lambda_1 \lambda_2 = \sigma_x^2 \sigma_y^2 - \sigma_{xy}^2$, its *determinant*.



Assume first that point 1, co-ordinates $\begin{bmatrix} x_1 & y_1 & z_1 \end{bmatrix}^\top$, is known. Then, unknowns are only the co-ordinates of point 2, $\begin{bmatrix} x_2 & y_2 & z_2 \end{bmatrix}^\top$.

Choose approximate values $x_2^{(0)}$, $y_2^{(0)}$, $z_2^{(0)}$ for these, and consistently

$$s^{(0)} \stackrel{\text{def}}{=} \sqrt{\left(x_2^{(0)} - x_1\right)^2 + \left(y_2^{(0)} - y_1\right)^2 + \left(z_2^{(0)} - z_1\right)^2}.$$

Subtraction yields

$$\begin{aligned} \Delta s = s - s^{(0)} &= \sqrt{\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2 + \left(z_2 - z_1\right)^2} - \\ &\quad - \sqrt{\left(x_2^{(0)} - x_1\right)^2 + \left(y_2^{(0)} - y_1\right)^2 + \left(z_2^{(0)} - z_1\right)^2} \approx \\ &\quad \approx \frac{\partial s}{\partial x_2} \overbrace{\left(x_2 - x_2^{(0)}\right)}^{\Delta x_2} + \frac{\partial s}{\partial y_2} \overbrace{\left(y_2 - y_2^{(0)}\right)}^{\Delta y_2} + \frac{\partial s}{\partial z_2} \overbrace{\left(z_2 - z_2^{(0)}\right)}^{\Delta z_2}, \end{aligned}$$

a truncated Taylor expansion.

Here, the abstract “vector” ℓ of the observations has only one element, $\ell = \begin{bmatrix} \Delta s \end{bmatrix} = \begin{bmatrix} s - s^{(0)} \end{bmatrix}$, and design matrix A (equation 14.10) consists of the partial derivatives of this one observable with respect to all unknowns x_2 , y_2 , z_2 :

$$A = \begin{bmatrix} \frac{\partial s}{\partial x_2} & \frac{\partial s}{\partial y_2} & \frac{\partial s}{\partial z_2} \end{bmatrix}.$$

The elements are derived by differentiating equation 14.19:

$$\frac{\partial s}{\partial x_2} = \frac{x_2 - x_1}{s}, \quad \frac{\partial s}{\partial y_2} = \frac{y_2 - y_1}{s}, \quad \frac{\partial s}{\partial z_2} = \frac{z_2 - z_1}{s}.$$

The numerical result is obtained by substituting the approximate values:

$$A \approx \begin{bmatrix} \frac{x_2^{(0)} - x_1}{s^{(0)}} & \frac{y_2^{(0)} - y_1}{s^{(0)}} & \frac{z_2^{(0)} - z_1}{s^{(0)}} \end{bmatrix}.$$

Using the approximate values of the co-ordinates is allowed if they are “good enough”, so close to reality that the linearisation is valid.

The elements of this matrix A are the components of a unit vector! By using the approximate values for the horizontal direction angle (azimuth) α and the zenith angle ζ , we may write

$$A \approx \begin{bmatrix} \cos \alpha^{(0)} \sin \zeta^{(0)} & \sin \alpha^{(0)} \sin \zeta^{(0)} & \cos \zeta^{(0)} \end{bmatrix},$$

or symbolically

$$A = \bar{\mathbf{e}}_{12}^\top,$$



in which $\bar{\mathbf{e}}_{12}$ is the *direction vector* between points 1 and 2 (as an abstract vector of components), for which holds

$$\|\bar{\mathbf{e}}_{12}\| = 1.$$

The more general case in which both points are unknown is discussed next. The vector of unknowns is formed

$$\mathbf{x} \stackrel{\text{def}}{=} \left[x_1 - x_1^{(0)} \quad y_1 - y_1^{(0)} \quad z_1 - z_1^{(0)} \mid x_2 - x_2^{(0)} \quad y_2 - y_2^{(0)} \quad z_2 - z_2^{(0)} \right]^T,$$

and the “vector” of observations is again $\ell = \left[\Delta s \right] = \left[s - s^{(0)} \right]$. Design matrix A , equation 14.10, is

$$A = \left[\frac{\partial s}{\partial x_1} \quad \frac{\partial s}{\partial y_1} \quad \frac{\partial s}{\partial z_1} \mid \frac{\partial s}{\partial x_2} \quad \frac{\partial s}{\partial y_2} \quad \frac{\partial s}{\partial z_2} \right].$$

Based on the previous, the end result is

$$A \approx \left[-\frac{x_{12}^{(0)}}{s^{(0)}} \quad -\frac{y_{12}^{(0)}}{s^{(0)}} \quad -\frac{z_{12}^{(0)}}{s^{(0)}} \mid \frac{x_{12}^{(0)}}{s^{(0)}} \quad \frac{y_{12}^{(0)}}{s^{(0)}} \quad \frac{z_{12}^{(0)}}{s^{(0)}} \right] = \left[-\bar{\mathbf{e}}_{12}^T \quad \bar{\mathbf{e}}_{12}^T \right],$$

in which we used the abbreviations

$$x_{12}^{(0)} \stackrel{\text{def}}{=} x_2^{(0)} - x_1^{(0)}, \quad y_{12}^{(0)} \stackrel{\text{def}}{=} y_2^{(0)} - y_1^{(0)}, \quad z_{12}^{(0)} \stackrel{\text{def}}{=} z_2^{(0)} - z_1^{(0)}.$$

The linearised observation equation is now

$$\left[s - s^{(0)} \right] = \left[-\bar{\mathbf{e}}_{12}^T \quad \bar{\mathbf{e}}_{12}^T \right] \begin{bmatrix} \Delta \bar{\mathbf{r}}_1 \\ \Delta \bar{\mathbf{r}}_2 \end{bmatrix}$$

in which the linearised co-ordinates

$$\Delta \bar{\mathbf{r}}_i \stackrel{\text{def}}{=} \begin{bmatrix} \Delta x_i \\ \Delta y_i \\ \Delta z_i \end{bmatrix} = \begin{bmatrix} x_i - x_i^{(0)} \\ y_i - y_i^{(0)} \\ z_i - z_i^{(0)} \end{bmatrix}, \quad i = 1, 2$$

are the unknowns, six in total.

Write this in stochastic form, with residual v :

$$\overbrace{\left[\Delta s \right]}^{\ell} + \overbrace{\left[v \right]}^v = \overbrace{\left[-\bar{\mathbf{e}}_{12}^T \quad \bar{\mathbf{e}}_{12}^T \right]}^A \overbrace{\begin{bmatrix} \Delta \hat{\mathbf{r}}_1 \\ \Delta \hat{\mathbf{r}}_2 \end{bmatrix}}^{\hat{\mathbf{x}}},$$

in which

$$\bar{\mathbf{e}}_{12}^T = \left[\frac{x_2^{(0)} - x_1^{(0)}}{s^{(0)}} \quad \frac{y_2^{(0)} - y_1^{(0)}}{s^{(0)}} \quad \frac{z_2^{(0)} - z_1^{(0)}}{s^{(0)}} \right] =$$



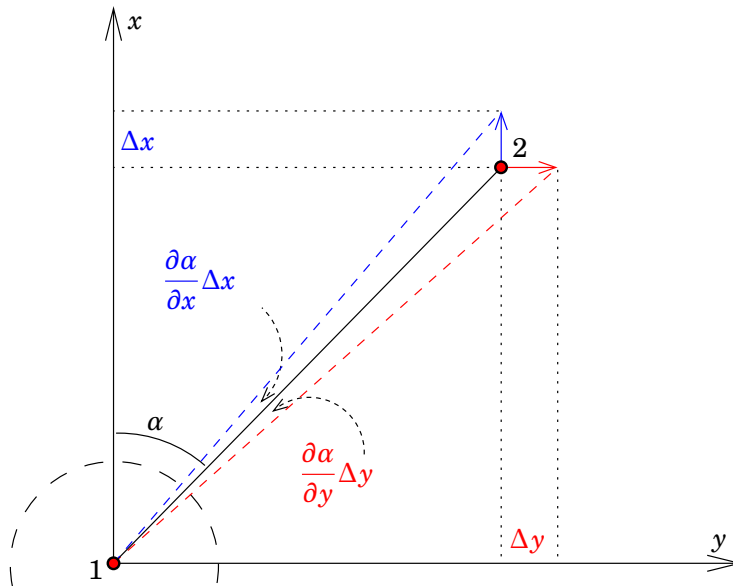


FIGURE 14.9. The geometry of azimuth measurement (seen from above) and elements of the design matrix.

$$= \begin{bmatrix} \frac{x_2 - x_1}{s} & \frac{y_2 - y_1}{s} & \frac{z_2 - z_1}{s} \end{bmatrix}^{(0)} = \begin{bmatrix} \frac{x_{12}}{s} & \frac{y_{12}}{s} & \frac{z_{12}}{s} \end{bmatrix}^{(0)},$$

$$\Delta \hat{\mathbf{r}}_i \stackrel{\text{def}}{=} \begin{bmatrix} \Delta \hat{x}_i \\ \Delta \hat{y}_i \\ \Delta \hat{z}_i \end{bmatrix} = \begin{bmatrix} \hat{x}_i - x_i^{(0)} \\ \hat{y}_i - y_i^{(0)} \\ \hat{z}_i - z_i^{(0)} \end{bmatrix}, \quad i = 1, 2,$$

$$\Delta \underline{s} \stackrel{\text{def}}{=} \underline{s} - s^{(0)}.$$

Here, the superscript (0) always identifies an approximate value. For approximate values, the *functional model applies exactly*:

$$s^{(0)} = f(\mathbf{x}^{(0)}) = \sqrt{\left(x_2^{(0)} - x_1^{(0)}\right)^2 + \left(y_2^{(0)} - y_1^{(0)}\right)^2 + \left(z_2^{(0)} - z_1^{(0)}\right)^2}.$$

14.9.2 Azimuth measurement

If the azimuth or horizontal direction α is measured between unknown points 1 and 2:

$$\alpha = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) + k \cdot 200^g,$$

then the vector of unknowns can be formed as

$$\mathbf{x} = \left[x_1 - x_1^{(0)} \quad y_1 - y_1^{(0)} \mid x_2 - x_2^{(0)} \quad y_2 - y_2^{(0)} \right]^T.$$

The design matrix is obtained again using the chain rule, as follows, abbreviating $x_{12} = x_2 - x_1$, $y_{12} = y_2 - y_1$. See figure 14.9 for the geometric



logic:

$$\begin{aligned} \frac{\partial \arctan(y_{12}/x_{12})}{\partial x_1} &= \frac{\partial \arctan(y_{12}/x_{12})}{\partial (y_{12}/x_{12})} \cdot \frac{\partial (y_{12}/x_{12})}{\partial x_{12}} \cdot \frac{\partial x_{12}}{\partial x_1} = \\ &= \frac{1}{1 + (y_{12}/x_{12})^2} \cdot \left(-\frac{y_{12}}{x_{12}^2}\right) \cdot (-1) = \frac{y_{12}}{x_{12}^2 + y_{12}^2} = \frac{y_{12}}{\rho^2} = \frac{\sin \alpha}{\rho}. \end{aligned}$$

The whole matrix is now

$$\begin{aligned} A &\approx \left[\begin{array}{cc|cc} +\frac{\sin \alpha}{\rho} & -\frac{\cos \alpha}{\rho} & -\frac{\sin \alpha}{\rho} & +\frac{\cos \alpha}{\rho} \end{array} \right]^{(0)} = \\ &= \left[\begin{array}{cc|cc} +\frac{y_2 - y_1}{\rho^2} & -\frac{x_2 - x_1}{\rho^2} & -\frac{y_2 - y_1}{\rho^2} & +\frac{x_2 - x_1}{\rho^2} \end{array} \right]^{(0)}, \end{aligned}$$

in which $\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$, the distance between instrument and signal, projected onto the horizontal plane.

From this, the linearised observation equation is obtained:

$$\underbrace{\Delta \alpha}_{\ell} + \underbrace{\underline{v}}_{\underline{v}} = \overbrace{\left[\begin{array}{cc|cc} +\frac{y_2 - y_1}{\rho^2} & -\frac{x_2 - x_1}{\rho^2} & -\frac{y_2 - y_1}{\rho^2} & +\frac{x_2 - x_1}{\rho^2} \end{array} \right]^{(0)}}^A \begin{array}{c} \widehat{x} \\ \Delta \widehat{x}_1 \\ \Delta \widehat{y}_1 \\ \Delta \widehat{x}_2 \\ \Delta \widehat{y}_2 \end{array},$$

in which the deltas are formed in the usual way by subtracting approximate values from observed or estimated values:

$$\begin{aligned} \Delta \alpha &= \alpha - \alpha^{(0)}, & \Delta \widehat{x}_1 &= \widehat{x}_1 - x_1^{(0)}, & \Delta \widehat{y}_1 &= \widehat{y}_1 - y_1^{(0)}, \\ & & \Delta \widehat{x}_2 &= \widehat{x}_2 - x_2^{(0)}, & \Delta \widehat{y}_2 &= \widehat{y}_2 - y_2^{(0)}. \end{aligned}$$



14.9.3 Horizontal-angle measurement

In practice, azimuth measurement is only possible using a gyrotheodolite, an expensive specialty instrument popular in underground measurements. Ordinary theodolites are not able to measure absolute horizontal directions but only direction *differences*, also known as horizontal angles.

One way of deriving an observation equation for this observation type is by adding an *orientation unknown* to the vector of unknowns, as already done in subsection 6.4.1: in that case, instrument station 1 still has, in addition to the three co-ordinates x_1 , y_1 , and z_1 , a fourth unknown Ω_1 , and the observation equation is

$$\theta = \arctan \frac{y_2 - y_1}{x_2 - x_1} - \Omega_1 + k \cdot 200^g.$$



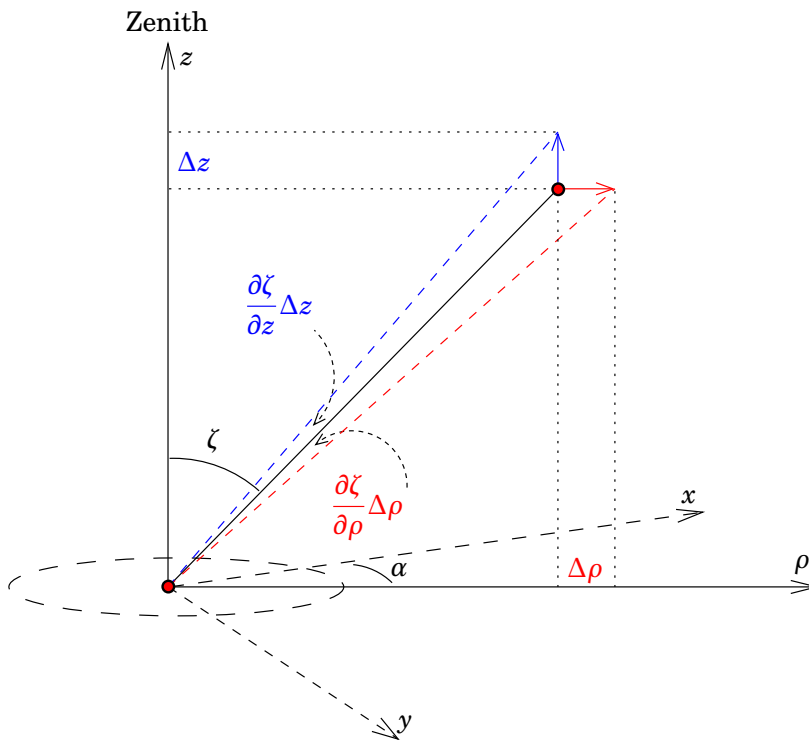


FIGURE 14.10. The geometry of zenith-angle measurement.

Linearisation is done as in the preceding subsection, with the result

$$\underbrace{\frac{\Delta \ell}{\Delta \theta}} + \underbrace{\frac{v}{v}} = \overbrace{\left[\begin{array}{cc|cc} +\frac{y_2 - y_1}{\rho^2} & -\frac{x_2 - x_1}{\rho^2} & -1 & -\frac{y_2 - y_1}{\rho^2} & +\frac{x_2 - x_1}{\rho^2} \end{array} \right]}^{A} \begin{matrix} \widehat{x} \\ \Delta \widehat{x}_1 \\ \Delta \widehat{y}_1 \\ \Delta \widehat{\Omega}_1 \\ \Delta \widehat{x}_2 \\ \Delta \widehat{y}_2 \end{matrix},$$

in which $\Delta \widehat{\Omega}_1 = \widehat{\Omega}_1 - \Omega_1^{(0)}$. This unknown represents the unknown azimuth of the zero mark on the horizontal circle of the instrument while it stands on point 1. vaakakehä

14.9.4 Zenith-angle measurement

The zenith angle between points 1 and 2 is measured,

$$\zeta = \arccos \frac{z_2 - z_1}{s} = 100^g - \arctan \frac{z_2 - z_1}{\rho},$$

in which again $\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$.

Write

$$A = \begin{bmatrix} -A_{12} & A_{12} \end{bmatrix}$$



in which (chain rule)

$$\begin{aligned} A_{12} &= \left[\frac{\partial \zeta}{\partial \rho} \cdot \frac{\partial \rho}{\partial x_2} \quad \frac{\partial \zeta}{\partial \rho} \cdot \frac{\partial \rho}{\partial y_2} \quad \frac{\partial \zeta}{\partial z_2} \right]^{(0)} = \\ &= \left[\frac{1}{s} \cos \zeta \cdot \cos \alpha \quad \frac{1}{s} \cos \zeta \cdot \sin \alpha \quad -\frac{1}{s} \sin \zeta \right]^{(0)} = \\ &= \left[\frac{z_2 - z_1}{s^2} \cdot \frac{x_2 - x_1}{\rho} \quad \frac{z_2 - z_1}{s^2} \cdot \frac{y_2 - y_1}{\rho} \quad -\frac{\rho}{s^2} \right]^{(0)}. \quad (14.20) \end{aligned}$$

Here, the partial derivatives of ζ have been computed in cylindrical coordinates (ρ, α, z) , with the definition $z \stackrel{\text{def}}{=} z_2 - z_1$. In these co-ordinates, $\zeta = \zeta(\rho, z)$, so ζ does not depend on the azimuth, and we obtain (figure 14.10):

$$\frac{\partial \zeta}{\partial \rho} = \frac{1}{s} \cos \zeta = \frac{1}{s} \frac{z_2 - z_1}{s}, \quad \frac{\partial \zeta}{\partial z_2} = -\frac{1}{s} \sin \zeta = -\frac{1}{s} \frac{\rho}{s}.$$

Now, the linearised observation equation is

$$\underbrace{\Delta \zeta}_{\ell} + \underbrace{v}_{\nu} = \underbrace{\begin{bmatrix} -A_{12} & A_{12} \end{bmatrix}}_A \underbrace{\begin{bmatrix} \Delta \hat{\mathbf{r}}_1 \\ \Delta \hat{\mathbf{r}}_2 \end{bmatrix}}_{\hat{\mathbf{x}}},$$

in which A_{12} as above, equation 14.20, and

$$\Delta \hat{\mathbf{r}}_i = \begin{bmatrix} \Delta \hat{x}_i \\ \Delta \hat{y}_i \\ \Delta \hat{z}_i \end{bmatrix} = \begin{bmatrix} \hat{x}_i - x_i^{(0)} \\ \hat{y}_i - y_i^{(0)} \\ \hat{z}_i - z_i^{(0)} \end{bmatrix}, \quad i = 1, 2.$$



14.10 Tacheometer measurement

We can bring the results of the previous four subsections together in describing tacheometer measurement, the simultaneous measurement of a horizontal direction, a vertical angle and a distance:

$$\begin{bmatrix} s \\ \theta \\ \zeta \end{bmatrix} + \begin{bmatrix} v_s \\ v_\theta \\ v_\zeta \end{bmatrix} = \begin{bmatrix} \begin{matrix} [3 \times 3] \\ -A_{12} \end{matrix} & \begin{matrix} 0 \\ -1 \\ 0 \end{matrix} & \begin{matrix} [3 \times 3] \\ A_{12} \end{matrix} \end{bmatrix} \begin{bmatrix} \Delta \hat{x}_1 \\ \Delta \hat{y}_1 \\ \Delta \hat{z}_1 \\ \Delta \hat{\Omega}_1 \\ \Delta \hat{x}_2 \\ \Delta \hat{y}_2 \\ \Delta \hat{z}_2 \end{bmatrix},$$

with

$$A_{12} = \begin{bmatrix} \frac{x_2 - x_1}{s} & \frac{y_2 - y_1}{s} & \frac{z_2 - z_1}{s} \\ -\frac{y_2 - y_1}{\rho^2} & +\frac{x_2 - x_1}{\rho^2} & 0 \\ \frac{z_2 - z_1}{s^2} \frac{x_2 - x_1}{\rho} & \frac{z_2 - z_1}{s^2} \frac{y_2 - y_1}{\rho} & -\frac{\rho}{s^2} \end{bmatrix}^{(0)}$$



with all the above definitions and results.

14.11 Helmert transformation in the plane

14.11.1 Theory

In the plane, if more than two points are given in both co-ordinate frames, it is possible to derive the unknown parameters of a Helmert transformation between the two frames. We start from equation 3.4 for a *single point* (note the notation change):

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + K \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Writing $K \stackrel{\text{def}}{=} 1 + m$, with m the *scale distortion*, assumed to be small, this becomes, for also small rotation angles θ :

kiertokulma

$$\begin{aligned} \begin{bmatrix} x' \\ y' \end{bmatrix} &\approx \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + (1+m) \begin{bmatrix} 1 & \theta \\ -\theta & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \approx \\ &\approx \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 & \theta \\ -\theta & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \end{aligned}$$

from which

$$\begin{bmatrix} x' \\ y' \end{bmatrix} - \begin{bmatrix} x \\ y \end{bmatrix} \approx \begin{bmatrix} x_0 + mx + \theta y \\ y_0 + my - \theta x \end{bmatrix},$$

which can be rearranged into

$$\underbrace{\begin{bmatrix} x' - x \\ y' - y \end{bmatrix}}_{\ell} = \underbrace{\begin{bmatrix} 1 & 0 & x^{(0)} & y^{(0)} \\ 0 & 1 & y^{(0)} & -x^{(0)} \end{bmatrix}}_A \underbrace{\begin{bmatrix} x_0 \\ y_0 \\ m \\ \theta \end{bmatrix}}_x.$$

This is recognised as a set of two observation equations, with the observation vector, the vector of unknowns, and the design matrix being, respectively, ℓ , x , and A , and in which $(x^{(0)}, y^{(0)})$ are *approximate values*¹⁶ of the co-ordinates (x, y) of the point.

16

For multiple points given in the two frames:

$$\begin{bmatrix} x'_i - x_i \\ y'_i - y_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_i^{(0)} & y_i^{(0)} \\ 0 & 1 & y_i^{(0)} & -x_i^{(0)} \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ m \\ \theta \end{bmatrix},$$

¹⁶Use of approximate values is allowed here, as both m and θ are assumed to be small.

with i the point number. We see that, if the number of available points n exceeds 2, there will be *redundancy*: more observation equations — $2n$ — than there are unknowns — 4. Written stochastically, we now have, with residuals,

$$\underbrace{\begin{bmatrix} x'_i - x_i \\ y'_i - y_i \end{bmatrix}}_{\ell} + \underbrace{\begin{bmatrix} v_{2i-1} \\ v_{2i} \end{bmatrix}}_{\nu} = \underbrace{\begin{bmatrix} 1 & 0 & x_i^{(0)} & y_i^{(0)} \\ 0 & 1 & y_i^{(0)} & -x_i^{(0)} \end{bmatrix}}_A \underbrace{\begin{bmatrix} \hat{x}_0 \\ \hat{y}_0 \\ \hat{m} \\ \hat{\theta} \end{bmatrix}}_{\hat{x}}, \quad i = 1, \dots, n. \quad (14.21)$$



14.11.2 Propagation of uncertainty

With more than two points common between the (x', y') and (x, y) frames, it will be possible not only to derive the Helmert transformation parameters, but also how the point uncertainties propagate into those of the transformation parameters. See for example [Kahmen and Faig \(1988\)](#), pages 253–255.

We start from the Helmert transformation equation rewritten as observation equations, 14.21. It is assumed that the co-ordinate values $(x_i^{(0)}, y_i^{(0)})$ available for the points are good enough, and that the scale distortion m and the rotation angle θ of the axes are both small. This is typically the case as many, also local, co-ordinate reference frames are approximately of the correct scale and oriented correctly to the north.

The above observation equations have $2n$ rows, where n is the number of points: $i = 1, \dots, n$. The design matrix A looks like

$$A = \begin{bmatrix} 1 & 0 & x_i^{(0)} & y_i^{(0)} \\ 0 & 1 & y_i^{(0)} & -x_i^{(0)} \end{bmatrix}.$$

If we assume that the co-ordinates of the given points are all equally precise, we may derive the *normal matrix*, or weight matrix of the unknowns, P_{xx} as follows (the symbol \sum means summation over all given points, all sub- and superscripts dropped):

$$P_{xx} = A^T A = \begin{bmatrix} n & 0 & \sum x & \sum y \\ 0 & n & \sum y & -\sum x \\ \sum x & \sum y & \sum (x^2 + y^2) & 0 \\ \sum y & -\sum x & 0 & \sum (x^2 + y^2) \end{bmatrix}.$$

The significance of this is that the variance matrix of the unknowns



$\hat{\mathbf{x}} = \begin{bmatrix} \hat{x}_0 & \hat{y}_0 & \hat{m} & \hat{\theta} \end{bmatrix}^\top$ is precisely

$$\text{Var}\{\hat{\mathbf{x}}\} \stackrel{\text{def}}{=} \begin{bmatrix} \sigma_{x_0}^2 & \sigma_{x_0 y_0} & \sigma_{x_0 m} & \sigma_{x_0 \theta} \\ \sigma_{y_0 x_0} & \sigma_{y_0}^2 & \sigma_{y_0 m} & \sigma_{y_0 \theta} \\ \sigma_{m x_0} & \sigma_{m y_0} & \sigma_m^2 & \sigma_{m \theta} \\ \sigma_{\theta x_0} & \sigma_{\theta y_0} & \sigma_{\theta m} & \sigma_{\theta}^2 \end{bmatrix} = \sigma^2 \mathbf{Q}_{xx} = \sigma^2 \mathbf{P}_{xx}^{-1},$$

in which $\mathbf{Q}_{xx} = \mathbf{P}_{xx}^{-1}$ is the weight-coefficient matrix of the unknowns and σ is the mean error of unit weight, in this case, the assumed precision of one point co-ordinate. Here, the diagonal elements $\sigma_{x_0}^2 = \text{Var}\{\hat{x}_0\}$ and so on are variances, $\sigma_{x_0 y_0} = \text{Cov}\{\hat{x}_0, \hat{y}_0\}$ and so on, covariances.

If now $\sum x = \sum y = 0$, in other words, (x, y) are *centre-of-mass co-ordinates*, we obtain

$$\mathbf{P}_{xx} = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & \sum(x^2 + y^2) & 0 \\ 0 & 0 & 0 & \sum(x^2 + y^2) \end{bmatrix}$$

and, with $\mathbf{Q}_{xx} = \mathbf{P}_{xx}^{-1}$,

$$\begin{aligned} \text{Var}\{\hat{\mathbf{x}}\} &= \text{Var}\left\{ \begin{bmatrix} \hat{x}_0 \\ \hat{y}_0 \\ \hat{m} \\ \hat{\theta} \end{bmatrix} \right\} = \sigma^2 \mathbf{Q}_{xx} = \\ &= \sigma^2 \begin{bmatrix} 1/n & 0 & 0 & 0 \\ 0 & 1/n & 0 & 0 \\ 0 & 0 & 1/\sum(x^2 + y^2) & 0 \\ 0 & 0 & 0 & 1/\sum(x^2 + y^2) \end{bmatrix}. \end{aligned}$$


Now, the variances of the transformation parameters are

$$\begin{aligned} \text{Var}\{\hat{x}_0\} &= \sigma_{x_0}^2 = \sigma^2 (\mathbf{Q}_{xx})_{11} = \frac{\sigma^2}{n}, \\ \text{Var}\{\hat{y}_0\} &= \sigma_{y_0}^2 = \sigma^2 (\mathbf{Q}_{xx})_{22} = \frac{\sigma^2}{n}, \\ \text{Var}\{\hat{m}\} &= \sigma_m^2 = \sigma^2 (\mathbf{Q}_{xx})_{33} = \sigma^2 / \sum(x^2 + y^2), \\ \text{Var}\{\hat{\theta}\} &= \sigma_{\theta}^2 = \sigma^2 (\mathbf{Q}_{xx})_{44} = \sigma^2 / \sum(x^2 + y^2). \end{aligned}$$

Moreover, the parameters do not statistically correlate with each other: the covariances between them vanish.

Now, the actual observation equations to be solved are 14.21, in which the quantities $\underline{v}_{2i-1}, \underline{v}_{2i}$ are the important *residuals*, containing valuable quality-control information.



 TABLE 14.2. Point set given in two different co-ordinate frames. The points are in the municipality of Porvoo, the co-ordinate frames are **ETRS-GK27** and **KKJ** zone 3.

Point	x'	y'	x	y
10	6697976.388	27427023.033	6698108.117	3427192.039
36	6700867.976	27428566.410	6700999.695	3428735.464
714	6701212.704	27424871.393	6701344.461	3425040.439
717	6696502.735	27428579.172	6696634.432	3428748.182
17_vara	6697821.437	27424568.639	6697953.182	3424737.628
35_vara	6709919.416	27434433.390	6710051.148	3434602.545
2061	6694497.478	27432539.402	6694629.164	3432708.409
2062	6704170.468	27432163.151	6704302.189	3432332.245
2063	6703595.075	27426736.815	6703726.833	3426905.879



Self-test questions

1. What is redundancy?
2. Why are observations weighted in an adjustment?
3. Why are the residuals of an adjustment of interest?
4. What is the mean error of unit weight?
5. What is the difference between variance matrix and weight-coefficient matrix?
6. What is a better-known name for the matrix of Jacobi resulting from the linearisation of observation equations?
7. What is the propagation law of variances for a stochastic vector quantity?
8. How many independent invariants has the variance matrix of a point location in the plane? What are they (for example)?
9. What are the parameters of a two-dimensional or plane Helmert transformation? What does the variance matrix of the least-squares solution of the plane Helmert transformation parameters look like in centre-of-mass co-ordinates?



Exercise 14–1: Helmert transformation parameter estimation

A set of points is given in two different co-ordinate frames, see table 14.2 (Porvoo). With this data, do the following operations:


1. Calculate (solve) the four Helmert transformation parameters¹⁷ x_0 ,



- y_0 , m , and θ . An octave script is provided for the purpose, tableau 14.3.
2. Calculate the parameters x'_0 , y'_0 , m' , and θ' of the *inverse transformation* — so, swap columns x' and y' with columns x and y .
 3. Verify that the scales $K = m + 1$ and $K' = m' + 1$ are each other's inverses: $K' = 1/K$ and thus $m' \approx -m$, and that the rotation parameters θ are each other's opposites: $\theta' = -\theta$.
 4. Look at the *residuals*. How precise were the input co-ordinates of the points?
 5. How many *observations* are there (hint: the number of observations equals the number of residuals)? How many *unknowns*? What is the number of *degrees of freedom* b (the difference between the number of observations and the number of unknowns, the *redundancy*)?
 6. Estimate $\hat{\sigma} \stackrel{\text{def}}{=} \sqrt{\frac{1}{b} \sum_{i=1}^{2n} v_i^2}$, in which b is the number of degrees of freedom, v_i the residuals, and $2n$ the number of observations. The estimated quantity σ is called the *mean error of unit weight*, the typical precision of a point co-ordinate.¹⁸
 7. Repeat the calculation after changing one x value by adding one metre. Note the effect on the residuals. Would you be able to *identify* this co-ordinate as the one containing the gross error? Also recompute $\hat{\sigma}$. What do you see?

¹⁷We leave off the stochastic underlines here, as these parameter values are just single realisations of the corresponding stochastic variables.

¹⁸More correctly, this only holds true if one of the two co-ordinate sets given, (x', y') and (x, y) , is assumed to be exact. Otherwise it is the typical precision of a co-ordinate *difference*.

 TABLEAU 14.3. Calculation script for Helmert transformation. Note that in this code, no reduction to center-of-mass co-ordinates takes place. The co-ordinate convention is the geodetic one: x points north, y east. Unlike in the text, however, the rotation angle θ is assumed positive in the anti-clockwise direction.

```
format long

% Data input:
npts = 9;
pts = [10, 36, 714, 717, 17, 35, 2061, 2062, 2063];
x2 = [6697976.388, 6700867.976, 6701212.704, 6696502.735, 6697821.437, ...
      6709919.416, 6694497.478, 6704170.468, 6703595.075];
y2 = [27427023.030, 27428566.410, 27424871.393, 27428579.172, 27424568.639, ...
      27434433.390, 27432539.402, 27432163.151, 27426736.815];
x1 = [6698108.117, 6700999.695, 6701344.461, 6696634.432, 6697953.182, ...
      6710051.148, 6694629.164, 6704302.189, 6703726.833];
y1 = [3427192.039, 3428735.464, 3425040.439, 3428748.182, 3424737.628, ...
      3434602.545, 3432708.409, 3432332.245, 3426905.879];

% Observable vector:
yoffset = (27-3)*1E6;
ell = [x2'-x1';y2'-y1'-yoffset];

% Construct design matrix:
A = zeros(2*npts,4);
for i = 1:npts
    A(i,:) = [1, 0, x1(i), -y1(i)];
    A(i+npts,:) = [0, 1, y1(i), x1(i)];
end

% Solve for unknowns:
x = inv(A'*A)*A'*ell;
x0 = x(1);
y0 = x(2) + yoffset;
scale = 1.0 + x(3);
rot = 57.296*x(4);
printf('\nUnknowns:\n\n');
printf('x0: %20.10f\ny0: %20.10f\nscale :%20.10f\nrot: %20.10f\n\n',...
      x0, y0, scale, rot);

% Residuals:
v = A*x - ell;
printf('Residuals:\n\n');
for i = 1:npts
    printf('%5i %10.3f %10.3f\n', pts(i), v(i), v(i+npts));
end
```




Statistical methods in geodesy

15

[...] There has not been a single date in the history of the theory of gravitation when a modern significance test would not have rejected all laws and left us with no law. Nevertheless the law did lead to improvement for centuries, [...]

Harold Jeffreys, 1939 (Jeffreys, 1998, page 391)

IN THIS CHAPTER we present two interrelated subjects:

- Statistical testing, in the context of the validation of and outlier detection in geodetic network measurements. poikkeava havaintoarvo
- The reliability of geodetic networks.

The framework of hypothesis testing with null and alternative hypotheses is adopted. We also show how this framework may be used, for example for geodetic deformation analysis.

We would be amiss in not pointing out that hypothesis testing is not the appropriate framework for settling all scientific disputes. There are other, often more appropriate, techniques, like the Akaike information criterion (Burnham and Anderson, 2013) and Bayesian approaches. The above quote from Harold Jeffreys¹ is apposite. 1



15.1 The method of least-squares

Explaining the method of least-squares is simplest if one assumes that all observables are stochastic quantities that are *normally distributed* (figure 2.5), both individually and together: they form a *multivariate* pienimmän neliösumman menetelmä

¹Sir Harold Jeffreys FRS (1891–1989) was a British mathematician, statistician, geophysicist, and astronomer, an influential advocate of Bayesian statistics.

normal distribution. If normally distributed observables are statistically independent of each other — for example, if they were produced by independent measurement processes — this is automatically the case.

The method of least-squares as a way to minimise the random errors in estimated quantities such as point co-ordinates has been presented above.

parametrinen tasoitus ehtotasoius monikulmiojono In chapter 14, the *parametric adjustment method* was presented, in which observations are expressed as functions of the unknowns. The alternative, adjustment by conditions, is suitable, for example, for computing traverses.

Here, the parametric method will be discussed in more detail.

Let the *observations*, as a vector $\underline{\ell}$, be linear functions of the *unknowns* \underline{x} :

$$\begin{array}{c} \underline{\ell} \\ \left[\begin{array}{c} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_n \end{array} \right] \\ [n] \end{array} = \begin{array}{c} A \\ \left[\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{array} \right] \\ [n \times m] \end{array} \begin{array}{c} \underline{x} \\ \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array} \right] \\ [m] \end{array} + \begin{array}{c} \underline{n} \\ \left[\begin{array}{c} n_1 \\ n_2 \\ \vdots \\ n_n \end{array} \right] \\ [n] \end{array} .$$

Here, n is the number of observations, m the number of unknowns. The observations, elements of the vector $\underline{\ell}$, are stochastic quantities. Assume that they are normally distributed around the “true” value of the observed quantity. Then, the elements of the vector of *observation errors* \underline{n} are also normally distributed.²

pienimmän neliösumman ratkaisu In this, rather general, case we may compute the *least-squares solution* in the following way:

$$\hat{\underline{x}} = \begin{array}{c} (A^T Q_{\ell\ell}^{-1} A)^{-1} \\ [m \times m]^{-1} \end{array} \begin{array}{c} (A^T Q_{\ell\ell}^{-1}) \\ [m \times n] \end{array} \begin{array}{c} \underline{\ell} \\ [n] \end{array}, \quad (15.1)$$

in which $Q_{\ell\ell}$ is the *weight-coefficient matrix* of the observations, size $[n \times n]$:

$$Q_{\ell\ell} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} .$$

²Often, they are also assumed to be statistically independent from each other, meaning that their random variations happen independently of each other. However, neither the elements of solution vector $\hat{\underline{x}}$ nor those of the vector of residuals \underline{v} will be statistically independent of each other.



This matrix is related to the variance matrix of the observations as follows:

$$\text{Var}\{\underline{\ell}\} \stackrel{\text{def}}{=} \Sigma_{\ell\ell} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} = \sigma^2 \mathbf{Q}_{\ell\ell},$$

in which

$$\sigma_i^2 = \text{Var}\{\underline{\ell}_i\} = E\left\{\left(\underline{\ell}_i - E\{\underline{\ell}_i\}\right)^2\right\} = \sigma^2 q_{ii},$$

$$\sigma_{ij} = \text{Cov}\{\underline{\ell}_i, \underline{\ell}_j\} = E\left\{\left(\underline{\ell}_i - E\{\underline{\ell}_i\}\right)\left(\underline{\ell}_j - E\{\underline{\ell}_j\}\right)\right\} = \sigma^2 q_{ij}.$$

Here, σ is the mean error of unit weight.

The variance matrix of the solution is obtained through propagation of variances. Let

$$\widehat{\mathbf{x}} = \mathbf{L}\underline{\ell},$$

in which

$$\mathbf{L} \stackrel{\text{def}}{=} (\mathbf{A}^\top \mathbf{Q}_{\ell\ell}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{Q}_{\ell\ell}^{-1}.$$

Then, based on equation 15.1:

$$\begin{aligned} \mathbf{Q}_{\mathbf{x}\mathbf{x}} &= \mathbf{L} \mathbf{Q}_{\ell\ell} \mathbf{L}^\top = (\mathbf{A}^\top \mathbf{Q}_{\ell\ell}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{Q}_{\ell\ell}^{-1} \cdot \mathbf{Q}_{\ell\ell} \cdot \mathbf{Q}_{\ell\ell}^{-1} \mathbf{A} (\mathbf{A}^\top \mathbf{Q}_{\ell\ell}^{-1} \mathbf{A})^{-1} = \\ &= (\mathbf{A}^\top \mathbf{Q}_{\ell\ell}^{-1} \mathbf{A})^{-1} \end{aligned}$$

by suitable elimination. So, the variance matrix of the solution vector $\widehat{\mathbf{x}}$, $\Sigma_{\mathbf{x}\mathbf{x}} = \sigma^2 \mathbf{Q}_{\mathbf{x}\mathbf{x}}$, is obtained in any case as a side product of computing the solution, equation 15.1!

painoyksikön
keskivirhe
variانس sien
kasautumislaki

15.2 The residuals from the adjustment

The least-squares estimators of the observations $\widehat{\underline{\ell}}$ and unknowns $\widehat{\mathbf{x}}$ are connected to each other through the functional model

$$\widehat{\underline{\ell}} = \mathbf{A}\widehat{\mathbf{x}},$$

and from the original observations

$$\underline{\ell} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

one computes the *residuals*:³

$$\underline{\mathbf{v}} \stackrel{\text{def}}{=} \widehat{\underline{\ell}} - \underline{\ell} = \mathbf{A}\widehat{\mathbf{x}} - \underline{\ell} = \mathbf{A}(\widehat{\mathbf{x}} - \mathbf{x}) - \mathbf{n}.$$

jäännös virhe
3



Residuals are central in the quality control of geodetic network solutions.

karkea virhe

- The size of the residuals tells something about the contradictions present in the network solution, possible gross errors, or even model errors.
- The size of the residual of a certain observation can indicate whether there may be a gross error hiding in this observation.
- The network must be *reliable*: there has to be redundancy, an over-determination by the observational material. For example, all kinds of closing errors offer possibilities for checking.

With little or no redundancy, the residuals may well be small or even zero, but this means nothing!

An often-used form of the *observation equations* is

$$\underline{\ell} + \underline{v} = A\hat{x}.$$

pienimmän
neliösumman
tasoitus

The residuals \underline{v} of a least-squares adjustment have four nice properties, here given without proof:

1. The quadratic form

$$\underline{\mathcal{E}} \stackrel{\text{def}}{=} \underline{v}^T P_{\ell\ell} \underline{v} = \underline{v}^T Q_{\ell\ell}^{-1} \underline{v} = \sigma^2 (\underline{v}^T \Sigma_{\ell\ell}^{-1} \underline{v}),$$

the weighted sum of the squares of the residuals, is *minimised* — this is what the methods of least-squares got its name from. In fact, the square root of this quantity is the *norm* of the vector of residuals \underline{v} , or its length, in the $Q_{\ell\ell}$ metric, which is thus minimised:⁴ $\underline{\mathcal{E}} = \|\underline{v}\|_Q^2$.

³The vector \underline{v} of residuals is not the same as the vector of observation errors, or “noise”, \underline{n} ! The residual is the difference between the original observation and the adjusted observation: in other words, a *correction*. However, not even an adjusted observation — or unknown — is the “truth”. The truth is not precisely knowable; it is only approximable at best, and the values of the elements of the vector \underline{n} , unlike the values of the elements of the vector \underline{v} , cannot be computed.

⁴One could eliminate the weight matrix altogether by applying a co-ordinate transformation in the vector space of observations: do a Cholesky decomposition $P_{\ell\ell} = \Gamma\Gamma^T$, resulting in $\underline{\mathcal{E}} = \underline{v}^T P_{\ell\ell} \underline{v} = \underline{v}^T \Gamma\Gamma^T \underline{v} = \tilde{\underline{v}}^T \tilde{\underline{v}}$, with $\tilde{\underline{v}} \stackrel{\text{def}}{=} \Gamma^T \underline{v}$. This is automatically achieved by redefining the observables as $\tilde{\underline{\ell}} \stackrel{\text{def}}{=} \Gamma^T \underline{\ell}$. This is the straightforward way of reducing the general least-squares problem 15.1 to the simpler unweighted one 14.3. The new observables $\tilde{\underline{\ell}}$ do not correlate with each other and have identical mean errors.



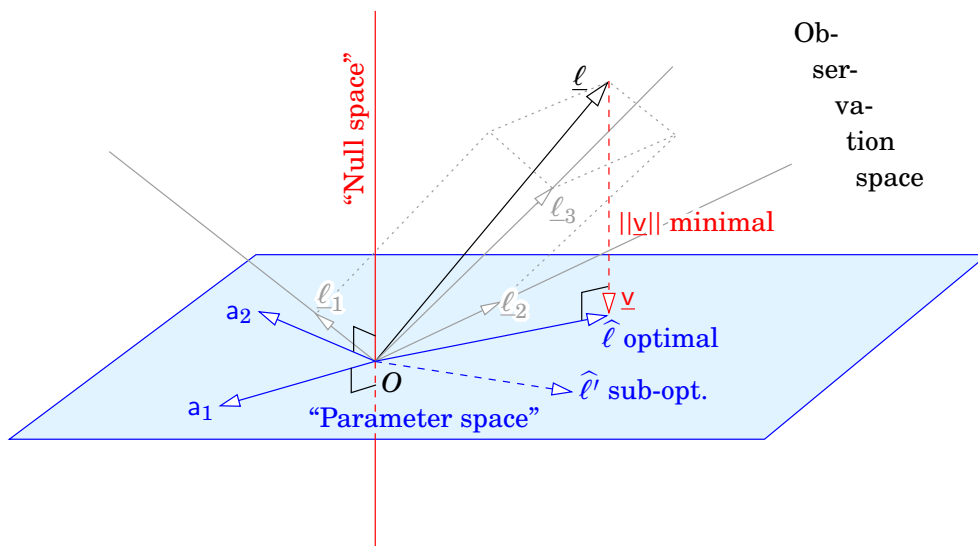


FIGURE 15.1. Least-squares adjustment as an orthogonal projection. For visualisation, it is assumed that the observation space has three dimensions, the parameter space two, and thus the number of degrees of freedom, the dimensionality of the null space, is one.



2. The variance $\Sigma_{\lambda\lambda}$ of an arbitrary linear combination $\hat{\lambda} = \Lambda\hat{x}$ of the unknowns \hat{x} (and its mean error $\sqrt{\Sigma_{\lambda\lambda}}$) is minimised.
3. The adjusted observables $\hat{\ell}$ and the residuals \underline{v} are mutually orthogonal in the $Q_{\ell\ell}$ metric: if the scalar product is defined as $\langle a \cdot b \rangle_Q \stackrel{\text{def}}{=} a^T Q_{\ell\ell}^{-1} b$:

tasoitettu havaintosuure

$$\langle \hat{\ell} \cdot \underline{v} \rangle_Q = \langle A\hat{x} \cdot \underline{v} \rangle_Q = (A\hat{x})^T Q_{\ell\ell}^{-1} \underline{v} = \hat{x}^T A^T Q_{\ell\ell}^{-1} \underline{v} = 0,$$

because

$$a_i^T Q_{\ell\ell}^{-1} \underline{v} = 0, \quad i = 1, \dots, m,$$

in other words, the vector of residuals is orthogonal to all columns a_i of the design matrix A .

rakennematriisi

Figure 15.1 gives a geometrical interpretation: the unknowns are those coefficients in the linear combination of the columns of the A matrix that minimise the norm of the vector of residuals.

4. The covariance matrix between the vector of unknowns \hat{x} and the vector of residuals \underline{v} vanishes: $\Sigma_{xv} = \sigma^2 Q_{xv} = 0$. So, they do not correlate with each other.

jäännös-
virheiden
vektori

Because

$$\underline{\ell} = A\hat{x} - \underline{v},$$



it follows, based on the law of propagation of variances and the above-mentioned property 4, that

$$Q_{\ell\ell} = A Q_{xx} A^T + Q_{vv} \implies Q_{vv} = Q_{\ell\ell} - A Q_{xx} A^T, \quad (15.2)$$

a useful equation for computing the *weight-coefficient matrix* Q_{vv} — and the variance matrix $\Sigma_{vv} = \sigma^2 Q_{vv}$ — of the residuals.



15.3 Testing, hypotheses for testing

The observational material may contain *gross errors*. In a real-life adjustment calculus we must be able to say, based on our knowledge of the statistical distribution of the observations,

- Something about the possible occurrence of gross errors.
- How large gross errors would have to be in order to be noticed and removed.
- The propagation of gross errors of this magnitude into the unknowns of interest.

Finding gross errors belongs to the field of *statistical testing*.

Gross errors that are found can be handled in two ways:

- They are removed from the observation set, and the measurements in question are repeated. After the fact, this is laborious and costly, wherefore at least part of statistical testing is done already in the field.
- They are simply left out. This assumes that the measurement was planned redundantly from the start: so many measurements have been made that one can afford to leave a (small) fraction of them out.

Statistical testing always requires the formulation of *hypotheses*. One of the hypotheses is always the

Null hypothesis All the measurements in the network are correct, there are no gross errors in them. This hypothesis is designated by the symbol H_0 .

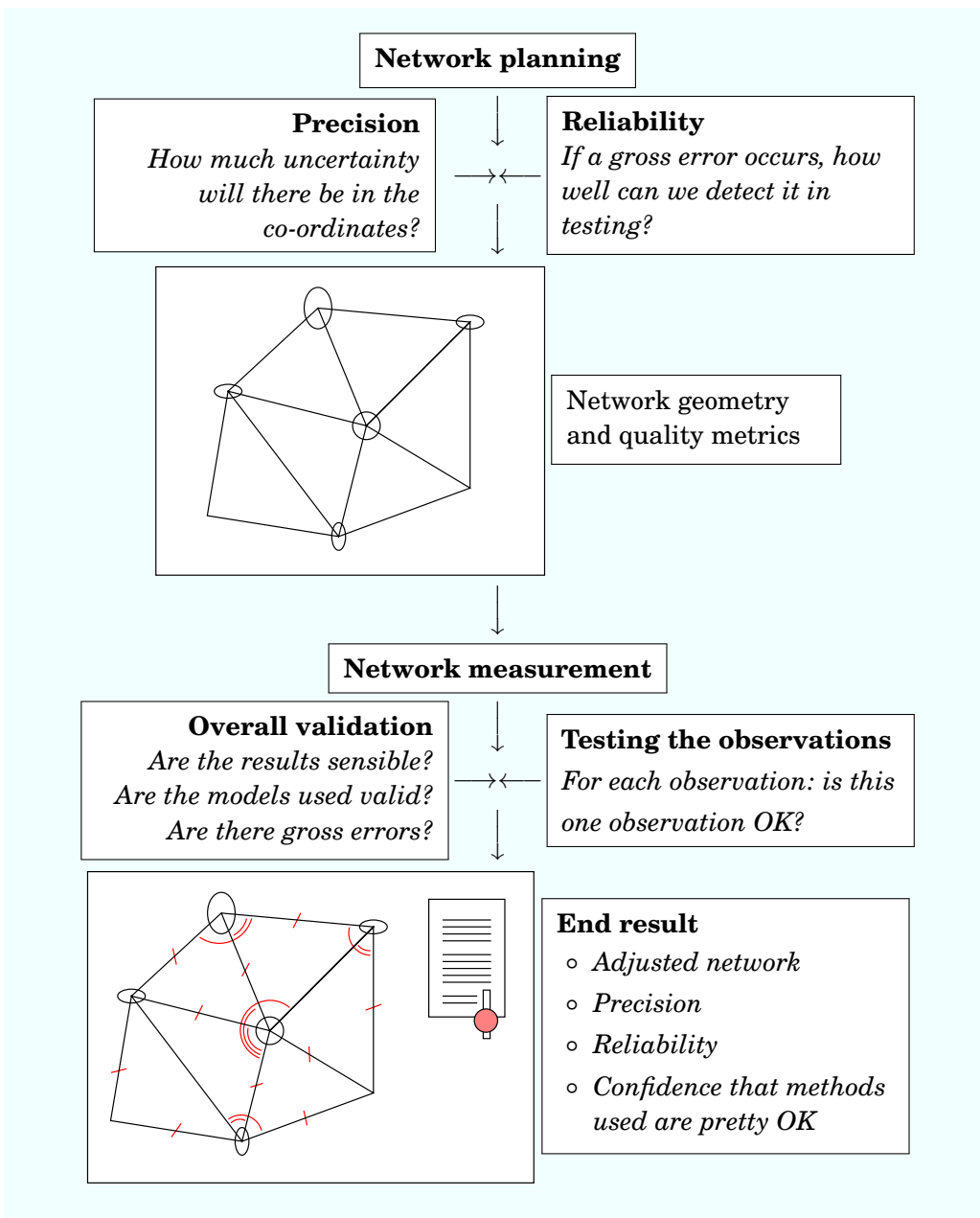
In addition, there must always be at least one

Alternative hypothesis The network contains some gross error, or some combination of gross errors, or a specific gross error. This hypothesis is designated by the symbol H_a .





TABLEAU 15.1. The planning and measurement process.



Generally we wish to know, or make a judgement on, two matters:

- Are there *generally* any gross errors left in this observation set?
- Is *this specific observation* in error?

Tableau 15.1 shows the role of testing in the whole planning and measurement process.

These questions will be discussed separately in the following sections.





15.4 Overall validation



15.4.1 The χ^2 distribution

Firstly we choose the following alternative hypothesis:

H_a : somewhere in the measurement material (we do not yet know where) there is a gross error.

This kind of hypothesis can be tested using the χ^2 test. The method, and the tables belonging to it, can be found in statistics textbooks and on the Internet. The quantity to be tested is the length of the vector of residuals in the $\Sigma_{\ell\ell}$ or $Q_{\ell\ell}$ metric, its *norm*, squared:

$$\frac{1}{\sigma^2} \mathcal{E} = \|\underline{v}\|_{\Sigma}^2 \stackrel{\text{def}}{=} \underline{v}^T \Sigma_{\ell\ell}^{-1} \underline{v} = \frac{1}{\sigma^2} (\underline{v}^T Q_{\ell\ell}^{-1} \underline{v}) \stackrel{\text{def}}{=} \frac{1}{\sigma^2} \|\underline{v}\|_Q^2. \quad (15.3)$$

painoyksikön
varianssi

This quantity — note the scaling with the variance of unit weight — is distributed according to the χ_{n-m}^2 distribution: the χ^2 distribution with $n - m$ degrees of freedom, figure 15.2. The number of *degrees of freedom* is the difference between the number of observations and the number of unknowns, also known as the *redundancy* $b = n - m$.

odotusarvo

Conceptually, a stochastic quantity with the χ_b^2 distribution, for b degrees of freedom, is obtained as the sum of the squares of b independent, standard-normally distributed — that is, having an expectancy zero and a mean error one — stochastic quantities \underline{n}_i , $i = 1, \dots, b$. As the expectancy of one such square equals the variance of the standard normal distribution, being one, it follows that the expectancy

$$E\{\chi_b^2\} = \sum_{i=1}^b E\{\underline{n}_i^2\} = \sum_{i=1}^b \text{Var}\{\underline{n}_i\} = \sum_{i=1}^b 1 = b.$$

Yet another perspective is that the quantity

$$\widehat{\sigma^2} \stackrel{\text{def}}{=} \frac{\mathcal{E}}{n-m} = \sigma^2 \frac{\underline{v}^T \Sigma_{\ell\ell}^{-1} \underline{v}}{n-m}$$

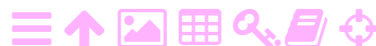
harhaton
estimaattori

has the expectancy $E\{\widehat{\sigma^2}\} = \sigma^2$ and is thus an *unbiased estimator* of σ^2 . It is called the *a posteriori* variance of unit weight. See also subsection 6.4.3. The ratio $\widehat{\sigma^2} / \sigma^2$ is expected to be close to unity if the null hypothesis is valid.



15.4.2 The overall test

By testing the above quantity 15.3, one may infer whether the measure-



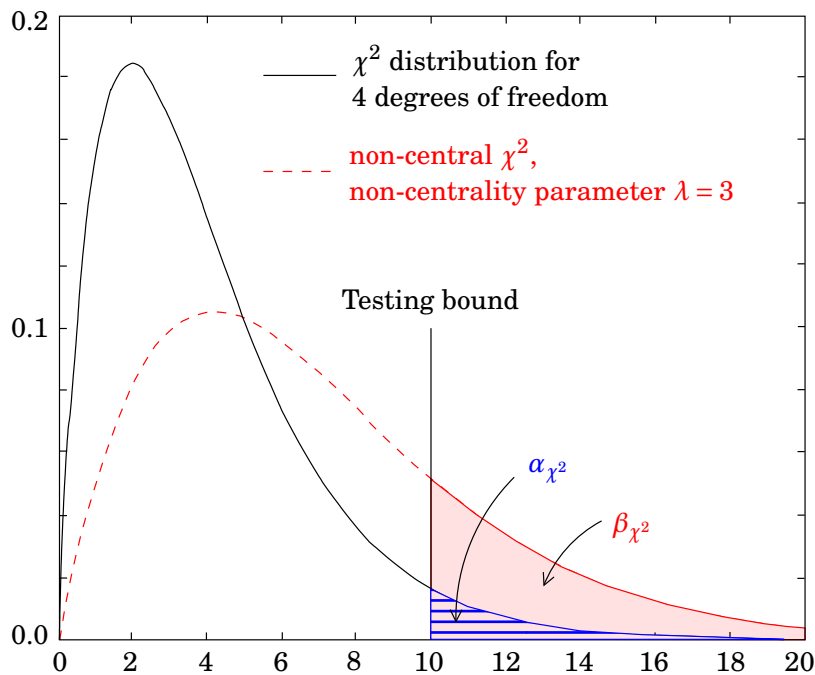


FIGURE 15.2. The χ^2 distribution with four degrees of freedom.

ment material contains *some* gross error or not,⁵ without yet stating in which observation it might be found.

MATLAB contains ready routines for applying the χ^2 method.

The quantity⁶ $\underline{\mathcal{E}}/\sigma^2$ is distributed according to χ_{n-m}^2 only in the case that the material contains no gross errors, that is, the *null hypothesis* H_0 applies. Then, the expectancy of the testing variate is

$$E\left\{\frac{1}{\sigma^2}\underline{\mathcal{E}} \mid H_0\right\} = E\{\chi_{n-m}^2\} = n - m.$$

Assume now instead however, that the observations contain one or more gross errors, taken together $\nabla\ell$: the *alternative hypothesis* H_a applies. The effect of this error vector on the *residuals* is ∇v . In this case, the distribution of the quantity $\underline{\mathcal{E}}/\sigma^2$ is the *non-central* χ^2 , in figure 15.2 the red curve.

⁵The χ^2 test cannot distinguish between actual gross errors and the possible unsuitability of the functional model $\ell = A \times$ applied to the material. If it happens that the χ^2 test rejects the null hypothesis, but all observations appear to be okay, it might be that there is a problem with the functional model: some systematic effect may have been overlooked.

⁶In Baarda's terminology: "shifting variate". Willem Baarda (1917–2005) was a pioneer of modern adjustment theory and statistical geodesy.





15.4.3 Reliability

The interesting question is now how large the effect of $\nabla\ell$ on $\nabla\mathbf{v}$ is going to be. We may hope that it will be large, because then, the network is ⁷ *reliable*.⁷ Generally the whole of $\nabla\ell$ does not propagate into $\nabla\mathbf{v}$; the *tasoitus* adjustment conveys part into the vector of unknowns, as a gross-error ⁸ effect,⁸ of magnitude $\nabla\mathbf{x}$. See also section 15.9.

In this situation, the vector of residuals is $\underline{\mathbf{v}} + \nabla\mathbf{v}$. Written out into terms, the testing variate becomes

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \|\underline{\mathbf{v}} + \nabla\mathbf{v}\|_{\Sigma}^2 = \underline{\mathbf{v}}^T \Sigma_{\ell\ell}^{-1} \underline{\mathbf{v}} + \underline{\mathbf{v}}^T \Sigma_{\ell\ell}^{-1} \nabla\mathbf{v} + \nabla\mathbf{v}^T \Sigma_{\ell\ell}^{-1} \underline{\mathbf{v}} + \nabla\mathbf{v}^T \Sigma_{\ell\ell}^{-1} \nabla\mathbf{v}.$$

The expectancy of the testing variate becomes

$$\begin{aligned} E\left\{\frac{1}{\sigma^2}\underline{\mathcal{E}} \mid H_a\right\} &= \\ &= E\left\{\underline{\mathbf{v}}^T \Sigma_{\ell\ell}^{-1} \underline{\mathbf{v}}\right\} + E\left\{\underline{\mathbf{v}}^T\right\} \Sigma_{\ell\ell}^{-1} \nabla\mathbf{v} + \nabla\mathbf{v}^T \Sigma_{\ell\ell}^{-1} E\left\{\underline{\mathbf{v}}\right\} + \nabla\mathbf{v}^T \Sigma_{\ell\ell}^{-1} \nabla\mathbf{v} = \\ &= E\left\{\chi_{n-m}^2\right\} + 0 + 0 + \nabla\mathbf{v}^T \Sigma_{\ell\ell}^{-1} \nabla\mathbf{v} = (n-m) + \lambda, \end{aligned}$$

in which $\lambda \stackrel{\text{def}}{=} \nabla\mathbf{v}^T \Sigma_{\ell\ell}^{-1} \nabla\mathbf{v}$ is called the *non-centrality parameter* of the χ^2 distribution. It describes how far the effect of the assumed gross error on the residuals, $\nabla\mathbf{v}$, extends outside the uncertainty area of the observations as described by the matrix $\Sigma_{\ell\ell}$.

A quadratic quantity is always positive. Therefore, the χ^2 test is *one-sided*, unlike the later presented test for the normal distribution. So, because $\underline{\mathcal{E}}$ is a quadratic quantity, every gross error — and even systematic errors, that is, errors in the functional model used — will tend to increase it. *Each and every error tends to make χ^2 larger*, and makes noticing the error more likely. This makes the χ^2 test such a useful overall test.

In fact, the χ^2 test validates a lot more than just the observations. It assures that

- The observation set does not contain any (large) gross errors.
- The functional model used (the observation equations) is valid with sufficient accuracy.
- The assumed mean errors of the observations (and the possible assumption of non-correlatedness) are realistic.

⁷This is called *interior reliability*.

⁸This is called *exterior reliability*. A small effect means a good exterior reliability.





15.5 Locating gross errors

If we have inferred that the material presumably contains one or more gross errors, we next want to find out which observations are under suspicion. Let us assume for simplicity that a gross error occurs only in one observation, although there might well be errors in several observations simultaneously.

The simplest way to search for gross errors, or rather, to undertake *outlier detection*, is to *look at the residuals*.

Let the vector of residuals be

$$\underline{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_i & \cdots & v_{n-1} & v_n \end{bmatrix}^T.$$

The element v_i is the residual of observation number i , that is the observation ℓ_i . The element's variance is

$$\sigma_{v_i}^2 = \sigma^2 [Q_{vv}]_{ii},$$

and its mean error σ_{v_i} is the square root of this.

Assume that the residuals v_i are *normally distributed*. Then, we may *test* every observation $i = 1, \dots, n$:

$$\begin{aligned} |v_i| > 1.96 \sigma_{v_i} &\implies \ell_i \text{ is probably in error} \\ |v_i| \leq 1.96 \sigma_{v_i} &\implies \ell_i \text{ is presumably correct.} \end{aligned}$$

This two-sided test based on the normal distribution uses a *significance level* of 95%: *Even if there is no gross error*, there nevertheless is a probability of 100% – 95% = 5% that, based on the test, observation ℓ_i will be rejected.

merkitsevyy-
taso


Table 15.2 gives the rejection bounds for different significance levels in the two-sided test based on the standard normal distribution.

The method described here works correctly only if the observations do not correlate with each other, so that the matrix $Q_{\ell\ell}$ is a diagonal matrix. If it is not, the literature offers an adapted⁹ testing method called *data snooping* (Baarda, 1968).

⁹The trick is simply that, instead of the residuals \underline{v} , *weighted residuals* $\underline{w} \stackrel{\text{def}}{=} Q_{\ell\ell}^{-1} \underline{v}$ and their variances are used.

The logic is, that if we search for a gross error in observation number i , we look for the linear combination of residuals in which the error shows clearest. We calculate the *orthogonal projection* of \underline{v} (in the $Q_{\ell\ell}$ metric) on the direction of the assumed gross error



 TABLE 15.2. Rejection bounds h for significance levels α_2 in a two-sided test based on the standard normal distribution, having mean error $\sigma = 1$ and expectancy $\mu = 0$. See figure 15.4.

$1 - \alpha_2, \%$	$\alpha_2, \%$	h
5	95	1.96
2.5	97.5	2.24
1	99	2.57
0.1	99.9	3.29


 **15.6 Calculation example: linear regression**

Let us return to the linear regression example already used in subsection 14.5.3, see table 15.3.

Recall that the least-squares solution found was

$$\begin{aligned} \hat{a} &= 1.76 \pm 1.25\sigma, \\ \hat{b} &= 0.495 \pm 0.349\sigma. \end{aligned}$$

We compute the function values $\hat{a} + \hat{b}x_i$ of the fitted line, as well as its residuals $\underline{v}_i = (\hat{a} + \hat{b}x_i) - \underline{y}_i$. The condition $\sum_{i=1}^n \underline{v}_i = 0$ is a good sanity check.

 TABLE 15.3. Example of linear regression.

$i \rightarrow$	1	2	3	4	5	$\sum_{i=1}^5$
x_i	1.51	2.44	3.34	4.41	5.05	16.75
\underline{y}_i	2.32	3.12	3.57	3.93	4.15	17.09
$\hat{a} + \hat{b}x_i$	2.51	2.97	3.41	3.94	4.26	
\underline{v}_i	+0.19	-0.15	-0.16	+0.01	+0.11	0.00
\underline{v}_i^2	0.0361	0.0225	0.0256	0.0001	0.0121	0.0964

$$\underline{e}_i \stackrel{\text{def}}{=} \begin{bmatrix} 0 & 0 & \dots & 1 & \dots & 0 & 0 \end{bmatrix}^T \text{ (where the element "1" is in place } i\text{):}$$

$$\underline{w}_i \stackrel{\text{def}}{=} \langle \underline{e}_i \cdot \underline{v} \rangle_Q = \underline{e}_i^T \underline{Q}_{\ell\ell}^{-1} \underline{v}.$$


Together, the components \underline{w}_i form the vector \underline{w} and they are optimally suited for discerning gross errors, or “outliers”.

In the test, also the mean error σ_{w_i} of every individual \underline{w}_i is needed, to be computed from the diagonal elements of the matrix (equation 15.2)

$$\underline{Q}_{\underline{w}\underline{w}} \stackrel{\text{def}}{=} \underline{Q}_{\ell\ell}^{-1} \underline{Q}_{\underline{v}\underline{v}} \underline{Q}_{\ell\ell}^{-1} = \underline{Q}_{\ell\ell}^{-1} (\underline{Q}_{\ell\ell} - \underline{A} \underline{Q}_{\underline{x}\underline{x}} \underline{A}^T) \underline{Q}_{\ell\ell}^{-1},$$

just as when testing using \underline{v} .



 TABLE 15.4. Values of the cumulative χ_3^2 distribution. α_{χ^2} is the significance level of the χ^2 test.

x	$\int_0^x \chi_3^2(x') dx'$ (α_{χ^2})	$\int_x^\infty \chi_3^2(x') dx'$ ($1 - \alpha_{\chi^2}$)
4.642	0.80	0.20
6.251	0.90	0.10
7.815	0.95	0.05
9.837	0.98	0.02
11.345	0.99	0.01
12.838	0.995	0.005
14.796	0.998	0.002
16.266	0.999	0.001

If the observations y_i have a variance matrix $\Sigma_{\ell\ell} = \sigma^2 I$, then the “shifting variate” to be tested is

siirtosuure

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \underline{v}^\top \Sigma_{\ell\ell}^{-1} \underline{v},$$

in which \underline{v} is the vector formed by the residuals v_i . We obtain

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \frac{1}{\sigma^2} \sum_{i=1}^n v_i^2 \implies \underline{\mathcal{E}} = \sum_{i=1}^n v_i^2.$$

If it is given *a priori* that $\sigma = \pm 0.15$, it follows that

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \frac{0.0964}{0.0225} = 4.28.$$

The quantity $\underline{\mathcal{E}}/\sigma^2$ is distributed according to χ_3^2 : there are $n = 5$ observations and $m = 2$ unknowns (a and b), so the *number of degrees of freedom* (redundancy) is $n - m = 3$. According to table 15.4, the probability that under the null hypothesis $\chi_3^2 > 4.642$ is 20%, so the value 4.28 is fully acceptable, at least on a significance level of 80%.


Next, the individual residuals are tested. Compute first the weight-coefficient matrix of the vector of residuals using equation 15.2:

$$Q_{vv} = Q_{\ell\ell} - A Q_{xx} A^\top,$$

in which $Q_{\ell\ell} = I$,

$$Q_{xx} = \begin{bmatrix} 1.5640 & -0.4072 \\ -0.4072 & 0.1215 \end{bmatrix}$$



 TABLE 15.5. Example of linear regression. Computing the residuals, their variance-covariance matrix, and normalised residuals.

	$i \rightarrow$	1	2	3	4	5
v_i		+0.19	-0.15	-0.16	+0.01	+0.11
		+0.3887	-0.4032	-0.2019	+0.0375	+0.1807
		-0.4032	+0.6998	-0.2006	-0.0821	-0.0112
Q_{vv}		-0.2019	-0.2006	+0.8007	-0.1978	-0.1969
		+0.0375	-0.0821	-0.1978	+0.6646	-0.4178
		+0.1807	-0.0112	-0.1969	-0.4178	+0.4502
σ_{v_i}		0.0935	0.1255	0.1342	0.1223	0.1006
$ v_i /\sigma_{v_i}$		2.03	1.20	1.19	0.08	1.09

was already computed in subsection 14.5.3, and

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} 1 & 1.51 \\ 1 & 2.44 \\ 1 & 3.34 \\ 1 & 4.41 \\ 1 & 5.05 \end{bmatrix}.$$

After careful calculation (MATLAB!), we obtain the matrix Q_{vv} , shown in table 15.5. Of this matrix, it is mostly the *diagonal elements* that are interesting:

$$\sigma_{v_1} = \sigma \sqrt{[Q_{vv}]_{11}} = 0.15 \cdot \sqrt{0.3887} = 0.15 \cdot 0.623 = 0.0935,$$

$$\sigma_{v_2} = \sigma \sqrt{[Q_{vv}]_{22}} = 0.1255,$$

¹⁰ and so on. (Remember that $\sigma = 0.15$.) See table 15.5.¹⁰

As can be seen, all observations are acceptable, with the exception of \underline{y}_1 , which, on the 95% significance level, is barely rejected (rejection bound 1.96). However, already on a significance level of 97.5%, it too is accepted.


15.7 Adding a gross error

karkea virhe Next, we add to the observed value \underline{y}_3 a simulated gross error +1.0.

¹⁰Note how the mean errors of the residuals are *systematically smaller* than the mean errors of the observations $\sigma = \pm 0.15$, especially close to the edges! With a large number of points, this phenomenon vanishes and we may write $Q_{vv} \approx Q_{\ell\ell}$.

This is often done in any case. Then, gross errors in the edge points will not be noticed sufficiently well.



 TABLE 15.6. Example of linear regression. A simulated gross error in point 3: original data, linear regression, residuals, testing.

	$i \rightarrow$	1	2	3	4	5	$\sum_{i=1}^5$
x_i		1.51	2.44	3.34	4.41	5.05	16.75
y_i		2.32	3.12	4.57	3.93	4.15	18.09
x_i^2		2.28	5.95	11.16	19.45	25.50	64.34
$x_i y_i$		3.50	7.61	15.26	17.33	20.96	64.66
$\hat{a} + \hat{b}x_i$		2.71	3.17	3.62	4.14	4.46	
v_i		+0.39	+0.05	-0.95	+0.21	+0.31	0.01
v_i^2		0.1521	0.0025	0.9025	0.0441	0.0961	1.1973
σ_{v_i}		0.0935	0.1255	0.1342	0.1223	0.1006	
$ v_i /\sigma_{v_i}$		4.17	0.40	7.08	1.72	3.08	
Rejection?		*		**		*	

Now, as the least-squares solution we obtain the result of table 15.6:

$$\hat{b} = \frac{5 \cdot 64.66 - 16.75 \cdot 18.09}{5 \cdot 64.34 - 16.75^2} = \frac{20.2925}{41.1375} = 0.493,$$

$$\hat{a} = \frac{1}{5} \left(18.09 - 16.75 \cdot \hat{b} \right) = 1.97.$$

In table 15.6 the σ_{v_i} values have not changed.

Compute the “shifting variate”

$$\frac{1}{\sigma^2} \mathcal{E} = \frac{1}{\sigma^2} \sum_{i=1}^n v_i^2 = \frac{1.1973}{0.0225} = 53.21!$$

There is something very wrong here...

Now look at table 15.6. The largest testing value by far, 7.08, is seen for the erroneous observation 3. But observations 1 and 5 are also rejected at even a 99% significance level! For this reason, one should proceed carefully. *Based on the test, one should reject only one observation at a time*, after which the whole least-squares computation should be repeated.

pienimmän
neliösumman
laskenta

15.8 Significance level of the test

15.8.1 Choice of rejection bound

When we test a certain alternative hypothesis against the null hypothesis using an assumedly normally distributed testing quantity or variate, one must choose a suitable *rejection bound*. If the variate to be tested exceeds



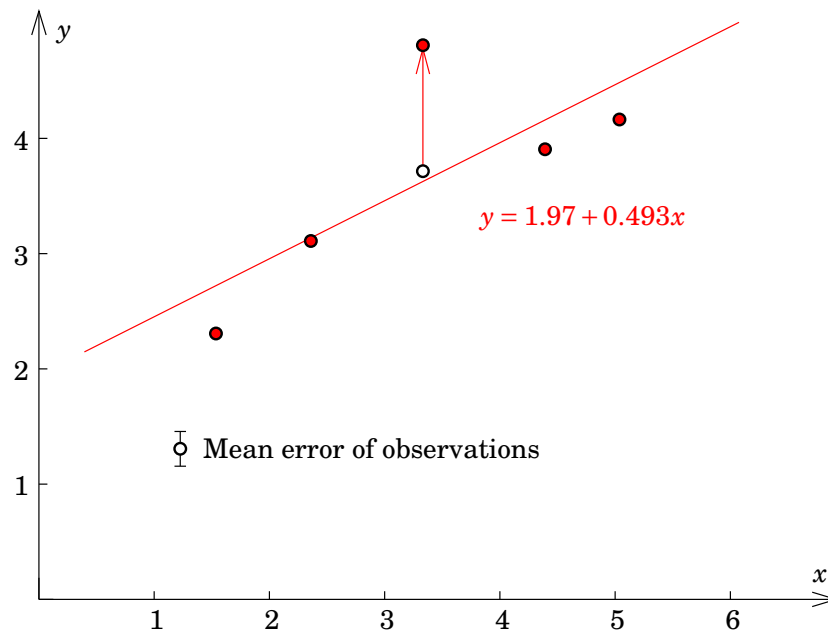


FIGURE 15.3. Example of linear regression, observation 3 contains a simulated gross error.

this bound, H_0 is rejected and H_a accepted. Choosing the rejection bound is an important strategic decision.

See figure 15.4. In the figure, the rejection bound chosen is $h = 2.5\sigma$: if the testing variate exceeds 2.5 times its own mean error σ , the null hypothesis H_0 is rejected and the alternative hypothesis H_a is accepted.

Now, the strategy may lead to two types of error:

ensimmäisen
lajin virhe
11

- The null is rejected although it is valid. This is called an *error of the first kind*.¹¹ The probability of this error happening is the size of the vertically hatched (blue) area. In the case of normal distribution, it amounts to $1 - \alpha = 1.24\%$ (two-sided), if the rejection bound is

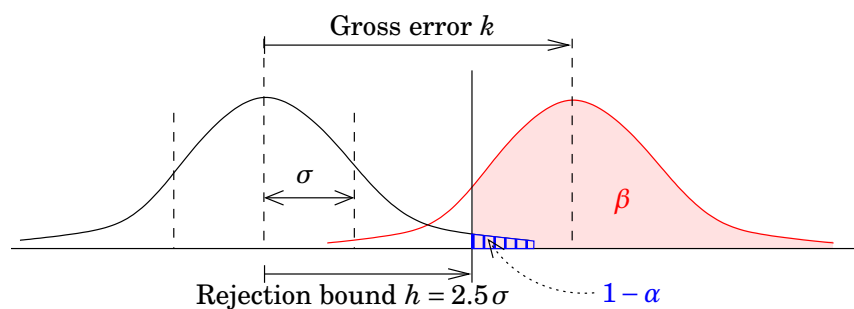



FIGURE 15.4. Statistical testing based on normal distribution.

¹¹Also rejection error.



 TABLE 15.7. Rejection bound h and probability of rejection in a test in the case of normal distribution. α_1 is the significance level of the one-sided, α_2 of the two-sided test.

h/σ	$1 - \alpha_1, \%$	$1 - \alpha_2, \%$
2.0	2.28	4.56
2.5	0.62	1.24
3.0	0.13	0.27
3.5	0.02	0.05

$h = 2.5\sigma$, see table 15.7. The quantity α is called the *significance level* of the test.


- The null is accepted although there is a gross error: H_0 is false, the alternative hypothesis H_a is true. This is an *error of the second kind*.¹² The probability of it happening depends on the size k of the gross error, more precisely, on the size of the normalised difference $(k - h)/\sigma$.

toisen lajin
virhe
12

Its complement, the probability of rejection, is called the *power* β ¹³ of the test. In figure 15.4 it is the size of the pink area.

erotuskyky
13

Choosing the *testing strategy*, choosing h , is thus always a *compromise*. It depends on the relative costs of errors of the first and second kind — including non-monetary, such as reputational, “costs”. $h = 3\sigma$ is often used — the “three-sigma rule”.

 TABLE 15.8. Assumed size k of gross error and corresponding power β of the test. Normal distribution and rejection bound $h = 2.5\sigma$ assumed.

k/σ	$(k - h)/\sigma$	$\beta, \%$
3.0	0.5	69.1
3.5	1.0	84.1
4.0	1.5	93.3
4.5	2.0	97.7
5.0	2.5	99.4
5.5	3.0	99.9
6.0	3.5	99.98

¹²Also *acceptance error*.

¹³So, the probability of an error of the second kind, if there is indeed a gross error in the observation, is $1 - \beta$, or $100\% - \beta$.



15.8.2 Harmonisation of overall and per-observation tests

In testing the body of observations, there is a *link* between the overall validation test and the per-observation tests! The link is through the significance levels. See figure 15.5: if the significance level of the χ^2 test is α_{χ^2} and that of the test for a single observation is α , the connection is

$$\alpha_{\chi^2} = \alpha^{n-m},$$

- ¹⁴ with $n - m$ the number of degrees of freedom.¹⁴ In other words, the joint probability that all observations individually pass their tests must be the same as the probability of passing the overall validation. Only on that condition may it be expected that, if the common χ^2 test finds something “rotten”, the tests for the individual observations will also point to the “guilty” observation.

Example If $\alpha_{\chi^2} = 95\%$ with ten degrees of freedom, it follows that

$$\alpha = \sqrt[n-m]{\alpha_{\chi^2}} = \sqrt[10]{0.95} \approx 0.99489 = 99.489\%,$$

_____ some ten times closer to 100%.

- After removing or correcting the “guilty” observation, the testing procedure is repeated, until the χ^2 test is passed.¹⁵



15.9 Reliability



15.9.1 Principle

The *reliability* of a measurement network is the property that gross errors are found easily, and are found even if they are relatively small.

Reliability corresponds to the network being “strong”. It is however not the same kind of strength as when the network is precise.

See figure 15.6. From points A , B , and C are measured the directions to a fourth point. Error ellipses for three different cases are drawn:

- I when the point is far from the points A and B
- II when the point is in a location where the directions to points A and B are perpendicular to each other, and

¹⁴This procedure is similar to the well-known *Bonferroni correction*, [Wikipedia, Bonferroni correction](#).

¹⁵If the test still fails, perhaps the other models used should be checked, such as the assumed precisions of the observations and so on.



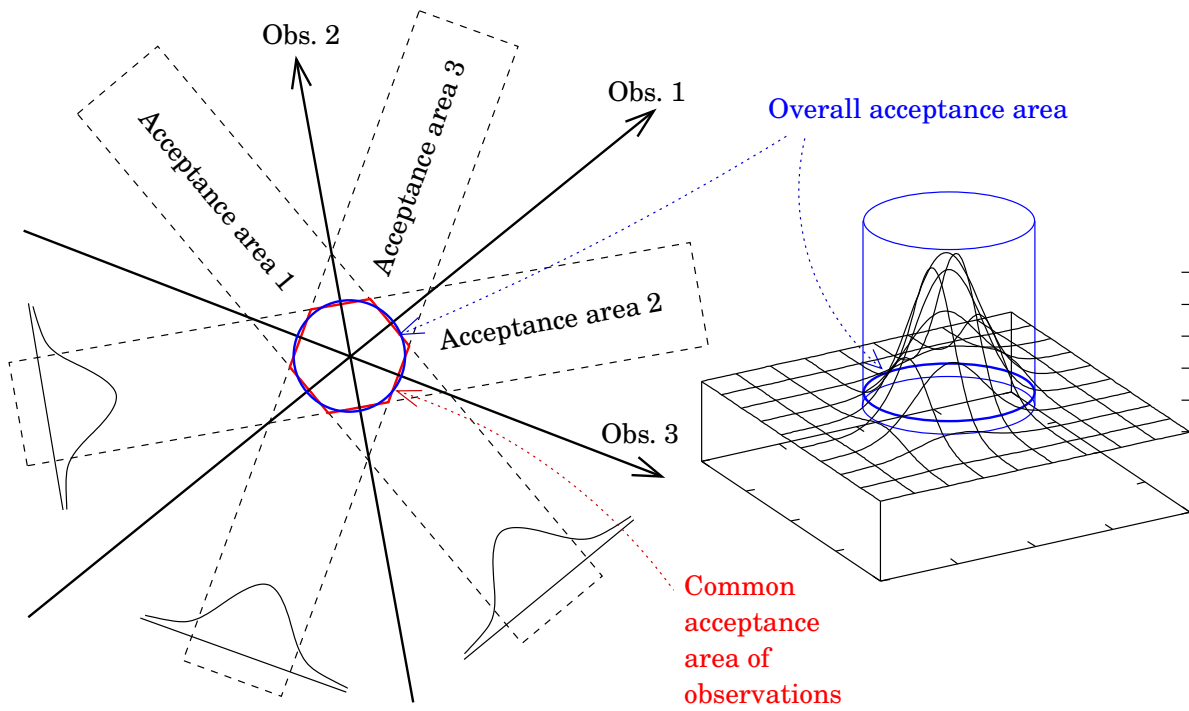


FIGURE 15.5. Harmonisation of the significance levels of the overall validation and per-observation tests.

III when the point is between points *A* and *B*.

As can be seen, the most precise result is obtained in case III. The error ellipse is the smallest.

However, *reliability is poor* (non-existent) in case III. If the measurement made from point *C* contains a gross error (dashed line), we still

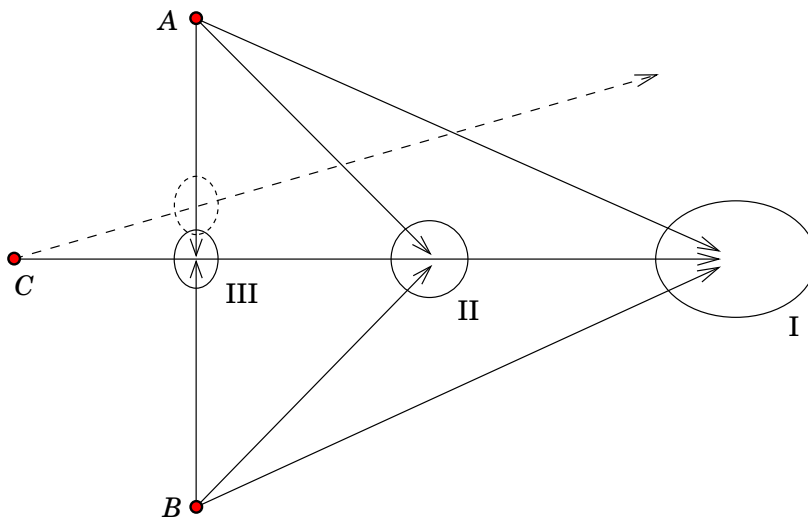


FIGURE 15.6. An example of reliability.



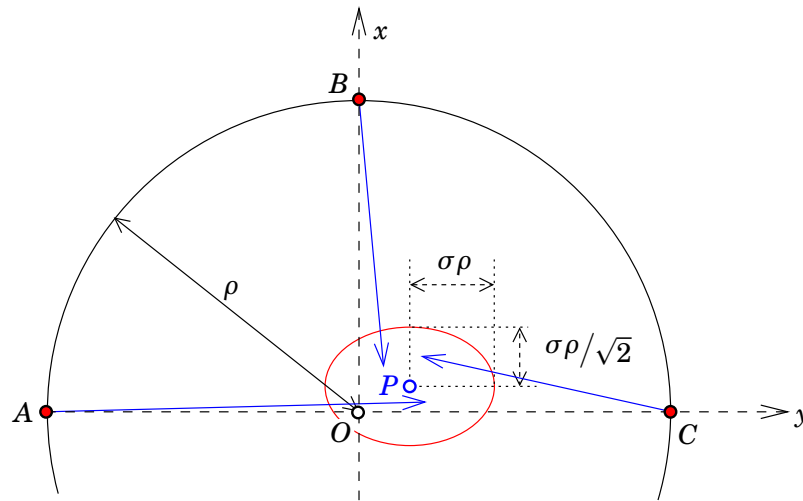


FIGURE 15.7. Another example of reliability.

obtain in case III a seemingly good — precise — but *erroneous* result. See the dashed error ellipse.

In cases I and II, a gross error in the observation from *C* will produce a contradictory result. It is not possible to find a location for the target point that is compatible with the direction measurements from all three points. This is a *good thing*, because it enables the detection of the gross error. The network is then called *reliable*.

In network planning, attention must be paid to reliability, of course in addition to precision. The network must be designed with appropriate redundancy: it must contain measurements that check each other. Common sense helps a great deal here. There are mathematical and software tools for evaluating the reliability of a network. One must always ask, “what if this or that observation were in error... would I notice?”

15.9.2 Another example

In this example, figure 15.7, the observation points *A*, *B*, and *C* are located on the edge of a circle, and point *P*, the direction to which is being measured, is located near the origin (centre point of the circle) *O*.

The observation equations are obtained by looking at the geometry:

$$\begin{bmatrix} \theta_{AP} - \theta_{AO} \\ \theta_{BP} - \theta_{BO} \\ \theta_{CP} - \theta_{CO} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \frac{1}{\rho} \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{x}_P \\ \hat{y}_P \end{bmatrix}$$



Symbolically

$$\underline{\ell} + \underline{v} = A\hat{x},$$

in which

$$\underline{\ell} = \begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ \underline{\ell}_3 \end{bmatrix} = \begin{bmatrix} \underline{\theta}_{AP} - \theta_{AO} \\ \underline{\theta}_{BP} - \theta_{BO} \\ \underline{\theta}_{CP} - \theta_{CO} \end{bmatrix}, \quad A = \frac{1}{\rho} \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

The least-squares solution is

$$\hat{x} = (A^T A)^{-1} A^T \underline{\ell} = \rho \cdot \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \underline{\ell}_3 - \underline{\ell}_1 \\ -\underline{\ell}_2 \end{bmatrix} = \rho \cdot \begin{bmatrix} \frac{1}{2}(\underline{\ell}_3 - \underline{\ell}_1) \\ -\underline{\ell}_2 \end{bmatrix}.$$

From this are obtained the residuals

$$\underline{v} = A\hat{x} - \underline{\ell} = \begin{bmatrix} -\frac{1}{2}(\underline{\ell}_3 - \underline{\ell}_1) \\ \underline{\ell}_2 \\ \frac{1}{2}(\underline{\ell}_3 - \underline{\ell}_1) \end{bmatrix} - \begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ \underline{\ell}_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\underline{\ell}_1 - \frac{1}{2}\underline{\ell}_3 \\ 0 \\ -\frac{1}{2}\underline{\ell}_1 - \frac{1}{2}\underline{\ell}_3 \end{bmatrix}. \quad (15.4)$$

Note 1 As can be seen, the observation $\underline{\ell}_2$ has vanished from the residuals! If $\underline{\ell}_2 = \underline{\theta}_{BP} - \theta_{BO}$ contains a gross error, we are never going to notice it as an overly large residual.

Note 2 From the residuals it cannot be seen whether a gross error comes from observation $\underline{\ell}_1$ or observation $\underline{\ell}_3$. In the residuals, their coefficients are identical.

We may also write equation 15.4 as

$$\underline{v} = A\hat{x} - \underline{\ell} = A(A^T A)^{-1} A^T \underline{\ell} - \underline{\ell} = -R\underline{\ell},$$

with the *redundancy matrix*

$$R \stackrel{\text{def}}{=} I - A(A^T A)^{-1} A^T = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

Each diagonal element of the redundancy matrix is a rough-and-ready measure for how well the geometry controls for a gross error in the corresponding observation. $R_{11} = R_{33} = \frac{1}{2}$ tells us that observations $\underline{\ell}_1$ and $\underline{\ell}_3$ are somewhat controlled, but $R_{22} = 0$ tells us that $\underline{\ell}_2$ is not controlled at all. A sensible requirement is that all $R_{ii} \gtrsim 0.5$.

Next, compute the shifting variate

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \underline{v}^T \Sigma_{\ell\ell}^{-1} \underline{v}.$$



Here, $\Sigma_{\ell\ell}$ is the variance matrix of the observations. Assume that the observations do not correlate with each other and that their mean error is σ . Then

$$\Sigma_{\ell\ell} = \sigma^2 \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}.$$

We obtain (H_0 is the null hypothesis):

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} \mid H_0 = \frac{1}{\sigma^2} \sum_{i=1}^3 v_i^2 = \frac{1}{2\sigma^2} (\underline{\ell}_1 + \underline{\ell}_3)^2.$$

Because the mean errors of both $\underline{\ell}_1$ and $\underline{\ell}_3$ are σ and they do not correlate, the mean error of the sum $\underline{\ell}_1 + \underline{\ell}_3$ is $\sigma\sqrt{2}$ and its variance $2\sigma^2$. The number of degrees of freedom is 1 and the variate $\underline{\mathcal{E}}/\sigma^2$ is distributed according to χ_1^2 , as should be the case according to theory.

By comparing the value $\underline{\mathcal{E}}/\sigma^2$ computed from the observations with the values from the χ_1^2 table, one can test whether the observations might contain a gross error. If all observations are free of gross errors, the expectancy of $\underline{\mathcal{E}}/\sigma^2$ is 1.

Nevertheless, as already pointed out above, we cannot observe any gross errors in $\underline{\ell}_2$ at all. We say¹⁶ that the measurement geometry is *reliable* for observations $\underline{\ell}_1$ and $\underline{\ell}_3$, but *unreliable* for observation $\underline{\ell}_2$. If observation $\underline{\ell}_2$ contained a gross error of size ∇ , it would slip *in its entirety* into the co-ordinate \hat{y}_P as an error $-\rho \cdot \nabla$! We also say¹⁷ that the measurement geometry is *unreliable* for unknown \hat{y}_P , but *reliable* for unknown \hat{x}_P .

A sufficiently large gross error ∇ in observations $\underline{\ell}_1$ or $\underline{\ell}_3$ would again be detected as an overly large value for the shifting variate (alternative hypothesis H_a):

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} \mid H_a = \frac{1}{2\sigma^2} (\underline{\ell}_1 + \underline{\ell}_3 + \nabla)^2,$$

the expectancy of which is $1 + \frac{1}{2} (\nabla/\sigma)^2$, for a non-centrality parameter $\lambda = \frac{1}{2} (\nabla/\sigma)^2$, see figure 15.2. If $\nabla \gg \sigma$, this would be detected with considerable confidence.

Note Reliability has nothing to do with *precision*! The precision of the

¹⁶This is called interior reliability.

¹⁷Exterior reliability.



unknowns $\hat{x} = \begin{bmatrix} \hat{x}_P & \hat{y}_P \end{bmatrix}^\top$ is described by their variance matrix

$$\text{Var}\{\hat{x}\} = \sigma^2 (A^\top Q_{\ell\ell}^{-1} A)^{-1} = \sigma^2 \rho^2 \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix},$$

so, the mean error of \hat{x}_P is $\frac{1}{2}\sigma\rho\sqrt{2}$, and that of \hat{y}_P is $\sigma\rho$, and they are uncorrelated with each other. See the error ellipse in figure 15.7.

However, a good mean error gives no solace if the co-ordinate solution \hat{y}_P contains a gross error...



15.9.3 The meaning of redundancy

Even though the reliability of a measurement network is good, we may still ask whether it is easy to identify the observation in which the gross error has occurred. If this is not easy, we end up measuring all suspect observations again, or throwing them out. This is not good.

From the viewpoints of both good reliability and identifiability of gross errors, the *degree of redundancy* of a geodetic measurement network should not be too low. If the number of observations is n and the number of unknowns m , then the number of conditions, or *degrees of freedom*, is $n - m$. The degree of redundancy is then $(n - m)/n$. This is often stated as a percentage. For example, linear regression of a straight line through five points: $n = 5$, $m = 2$, so a degree of redundancy of $\frac{3}{5} = 60\%$. On the other hand, a levelling line of ten points between two known points: $n = 11$, $m = 10$, the degree of redundancy being $\frac{1}{11} = 9\%$ — weak, but unfortunately common. By measuring in both directions we obtain $n = 22$, $m = 10$, so a degree of redundancy of $\frac{12}{22} \approx 55\%$, which is already good. A good rule of thumb is that a degree of redundancy of 50% is desirable.



15.10 Deformation analysis

Deformation analysis is one practical application of statistical testing. The null hypothesis H_0 in these tests is, that no observable deformation has happened. There may be many different alternative hypotheses H_a , from the hypothesis that some unspecified deformation took place, to many concrete hypotheses about the precise nature of the deformation sought.

Deformation analysis is also an application that involves the time dimension: measurements collected in two or more measurement epochs



are compared. The deformations studied may be natural, like deformations in the Earth's crust brought about by tectonic movements or by varying glacial loading; or they may be brought about by human activity, like the subsidence caused by mineral extraction (petroleum, natural gas, irrigation water, ...). The object of study may be the Earth's crust in an area, or a building or other built structure like a reservoir dam. The possibilities are very broad.

Deformation analysis is discussed in the textbooks by Cooper (1987) pages 331–352 and Vaníček and Krakiwsky (1986) pages 611–659.



15.10.1 Height deformation analysis

In one dimension, *height deformation analysis* studies vertical movement, for example using levelling. In the simplest case, the same levelling line or network of n points has been measured twice:

$$\underline{H}_i(t_1), \quad \underline{H}_i(t_2), \quad i = 1, \dots, n,$$

and the variance matrices of the heights, $\Sigma(t_1)$ and $\Sigma(t_2)$, are also available.

Clearly, comparison is possible only if both measurements are first reduced to the same reference or *datum point*. We choose the first network point, point 1, as the datum point:

$$H_1^{(1)}(t_1) = H_1^{(1)}(t_2) \quad (= \text{some agreed value}).$$

After this, the variance matrices for both measurement times or *epochs* are only of size $(n-1) \times (n-1)$, because now point 1 is *known* and no longer has (co-)variances.

$$\Sigma^{(1)}(t_1) = \begin{bmatrix} \sigma_{22}^{(1)}(t_1) & \sigma_{23}^{(1)}(t_1) & \cdots & \sigma_{2n}^{(1)}(t_1) \\ \sigma_{32}^{(1)}(t_1) & \sigma_{33}^{(1)}(t_1) & \cdots & \sigma_{3n}^{(1)}(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n2}^{(1)}(t_1) & \sigma_{n3}^{(1)}(t_1) & \cdots & \sigma_{nn}^{(1)}(t_1) \end{bmatrix},$$

and the same for $\Sigma^{(1)}(t_2)$. Here

$$\left. \begin{aligned} \sigma_{ii}^{(1)}(t_\ell) &= \text{Var}\{H_i^{(1)}(t_\ell)\}, \\ \sigma_{ij}^{(1)}(t_\ell) &= \text{Cov}\{H_i^{(1)}(t_\ell), H_j^{(1)}(t_\ell)\}, \end{aligned} \right\} \quad \ell = 1, 2; \quad i, j = 2, \dots, n.$$



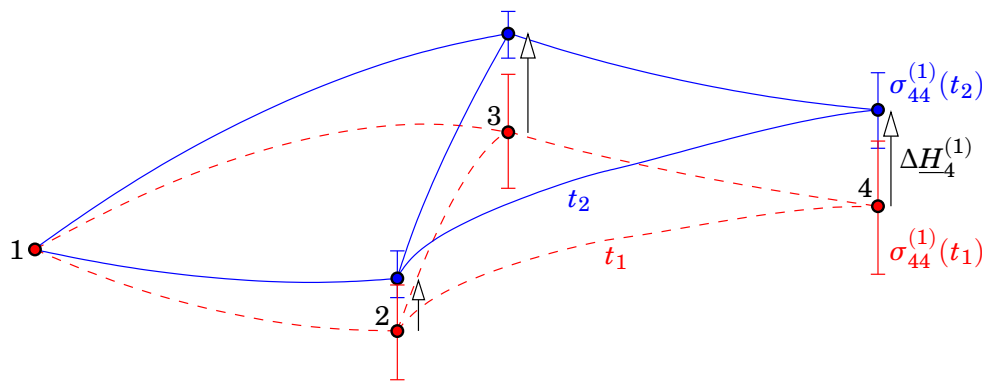


FIGURE 15.8. Height deformation monitoring network for the epochs t_1 (red) and t_2 (blue). Realistic error bars.

Now, calculate the height displacements between the two measurement epochs and their joint variance matrix, assuming that the measurements made at times t_1 and t_2 are statistically independent of each other:

$$\Delta \underline{H}_i^{(1)} \stackrel{\text{def}}{=} \underline{H}_i^{(1)}(t_2) - \underline{H}_i^{(1)}(t_1), \quad i = 2, \dots, n,$$

$$\Sigma_{\Delta H \Delta H}^{(1)} = \Sigma^{(1)}(t_1) + \Sigma^{(1)}(t_2).$$

After this it is intuitively clear — provided that both sets of height measurements are multi-normally distributed — that the following quantity, the *shifting variate*, has the χ_{n-1}^2 distribution:

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \left(\Delta \underline{H}^{(1)} \right)^T \left(\Sigma_{\Delta H \Delta H}^{(1)} \right)^{-1} \Delta \underline{H}^{(1)},$$

in which

$$\Delta \underline{H}^{(1)} \stackrel{\text{def}}{=} \begin{bmatrix} \underline{H}_2^{(1)}(t_2) - \underline{H}_2^{(1)}(t_1) \\ \underline{H}_3^{(1)}(t_2) - \underline{H}_3^{(1)}(t_1) \\ \vdots \\ \underline{H}_n^{(1)}(t_2) - \underline{H}_n^{(1)}(t_1) \end{bmatrix} = \begin{bmatrix} \Delta \underline{H}_2^{(1)} \\ \Delta \underline{H}_3^{(1)} \\ \vdots \\ \Delta \underline{H}_n^{(1)} \end{bmatrix}$$

is the (abstract) vector of height differences.

Statistical testing for deformation is based on this variate $\underline{\mathcal{E}}$.

15.10.2 Horizontal deformation analysis

In two dimensions we proceed in the same way as in the one-dimensional case, except that

1. It is tempting to write the plane co-ordinates as *complex numbers*.
2. There are now *two* datum points, the co-ordinates of which are considered identical between the two epochs.



So, if there are n points, the size of the variance matrix is now $(n-2) \times (n-2)$. The variance matrix is now also complex valued, and *Hermitian*:

liittoluku its transpose is its complex conjugate.

18 The testing variate is again the shifting variate¹⁸

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \left(\underline{\mathbf{d}}^{(AB)} \right)^\dagger \left(\underline{\Sigma}_{\text{dd}}^{(AB)} \right)^{-1} \underline{\mathbf{d}}^{(AB)},$$

in which $\underline{\mathbf{d}}$ is the complex vector of all co-ordinate differences, or *displacement vector*:

$$\begin{aligned} \underline{\mathbf{d}}^{(AB)} &\stackrel{\text{def}}{=} \\ &= \begin{bmatrix} \underline{x}_3^{(AB)}(t_2) - \underline{x}_3^{(AB)}(t_1) + i \left(\underline{y}_3^{(AB)}(t_2) - \underline{y}_3^{(AB)}(t_1) \right) \\ \underline{x}_4^{(AB)}(t_2) - \underline{x}_4^{(AB)}(t_1) + i \left(\underline{y}_4^{(AB)}(t_2) - \underline{y}_4^{(AB)}(t_1) \right) \\ \vdots \\ \underline{x}_n^{(AB)}(t_2) - \underline{x}_n^{(AB)}(t_1) + i \left(\underline{y}_n^{(AB)}(t_2) - \underline{y}_n^{(AB)}(t_1) \right) \end{bmatrix} = \begin{bmatrix} \Delta \underline{\mathbf{z}}_3^{(AB)} \\ \Delta \underline{\mathbf{z}}_4^{(AB)} \\ \vdots \\ \Delta \underline{\mathbf{z}}_n^{(AB)} \end{bmatrix}, \end{aligned}$$

with

$$\begin{aligned} \Delta \underline{\mathbf{z}}_i^{(AB)} &= \underline{\mathbf{z}}_i^{(AB)}(t_2) - \underline{\mathbf{z}}_i^{(AB)}(t_1) = \\ &= \left(\underline{x}_i^{(AB)}(t_2) - \underline{x}_i^{(AB)}(t_1) \right) + i \left(\underline{y}_i^{(AB)}(t_2) - \underline{y}_i^{(AB)}(t_1) \right), \quad i = 3, \dots, n. \end{aligned}$$

AB is the chosen datum or starting point for both epochs t_1 and t_2 . The other points are numbered 3, 4, ..., n . The symbol \dagger signifies both

19 transposition and complex conjugate, the *Hermitian*¹⁹ conjugate:

$$A^\dagger \stackrel{\text{def}}{=} \overline{A^\top} = \overline{A}^\top.$$



15.10.3 Example

20 Let the adjusted co-ordinates $\underline{\mathbf{x}}_i(t_1)$, $i = 3, \dots, 6$ of the deformation network from the first measurement epoch be given²⁰ in table 15.9a,

¹⁸Warning: in Cooper's book (Cooper, 1987, page 335) there is a mistake under equation (9.52): the correct equation is (inverse, not transpose):

$$\Omega = \hat{\mathbf{d}}^\dagger \hat{Q}_d^{-1} \hat{\mathbf{d}}.$$

¹⁹Charles Hermite (1822–1901) FRS FRSE was a French mathematician and number theorist.



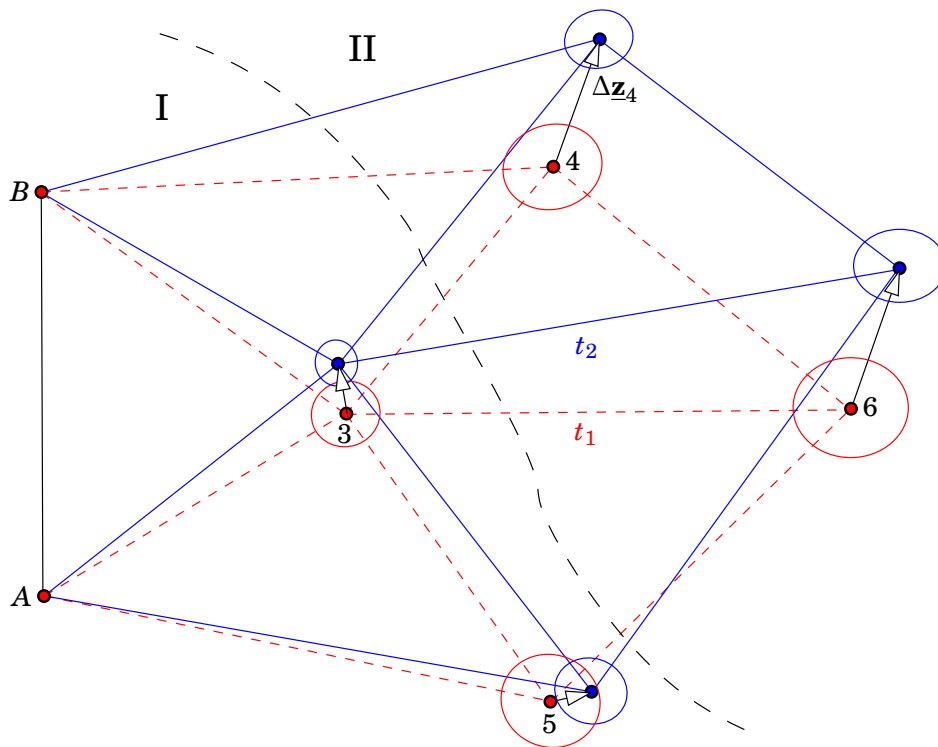


FIGURE 15.9. Horizontal deformation monitoring network for the epochs t_1 (red) and t_2 (blue). Realistic error ellipses. The alternative hypothesis that domain II moves with respect to domain I is also indicated.



and the co-ordinates of the second measurement epoch $\underline{x}_i(t_2)$, $i = 3, \dots, 6$ be given in table 15.9b.

Compute the inter-epoch differences vector \underline{d} , table 15.9c.

Using real numbers, with the definition

$$\underline{d} \stackrel{\text{def}}{=} \underline{x}(t_2) - \underline{x}(t_1) = \begin{bmatrix} \underline{x}_3(t_2) - \underline{x}_3(t_1) \\ \underline{y}_3(t_2) - \underline{y}_3(t_1) \\ \underline{x}_4(t_2) - \underline{x}_4(t_1) \\ \underline{y}_4(t_2) - \underline{y}_4(t_1) \\ \underline{x}_5(t_2) - \underline{x}_5(t_1) \\ \underline{y}_5(t_2) - \underline{y}_5(t_1) \\ \underline{x}_6(t_2) - \underline{x}_6(t_1) \\ \underline{y}_6(t_2) - \underline{y}_6(t_1) \end{bmatrix} = \begin{bmatrix} \Delta \underline{x}_3 \\ \Delta \underline{y}_3 \\ \Delta \underline{x}_4 \\ \Delta \underline{y}_4 \\ \Delta \underline{x}_5 \\ \Delta \underline{y}_5 \\ \Delta \underline{x}_6 \\ \Delta \underline{y}_6 \end{bmatrix},$$

²⁰These are *only* the co-ordinates of the points to be tested. They are assumed to be connected, for both epochs, to the same two unnamed datum points — say, points 1 and 2, or A and B — outside the area, which are assumed to be motionless.





TABLE 15.9. Horizontal deformation analysis, co-ordinates. Unit metre.

(a)			(b)			(c)		
Epoch 1			Epoch 2			Displacements		
i	$\underline{x}_i(t_1)$	$\underline{y}_i(t_1)$	i	$\underline{x}_i(t_2)$	$\underline{y}_i(t_2)$	i	$\Delta\underline{x}_i$	$\Delta\underline{y}_i$
3	1234.123	2134.453	3	1234.159	2134.448	3	+0.036	-0.005
4	1681.045	2507.487	4	1681.123	2507.516	4	+0.078	+0.029
5	755.495	2623.456	5	755.507	2623.487	5	+0.012	+0.031
6	1248.865	3051.775	6	1248.951	3051.807	6	+0.086	+0.032

we find by computation

$$\underline{\mathbf{d}}^T \underline{\mathbf{d}} = \sum_{i=3}^6 \left((\underline{x}_i(t_2) - \underline{x}_i(t_1))^2 + (\underline{y}_i(t_2) - \underline{y}_i(t_1))^2 \right) = 0.017771 \text{ m}^2.$$

Similarly with complex numbers, with the definition

$$\underline{\mathbf{d}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{z}_3(t_2) - \mathbf{z}_3(t_1) \\ \mathbf{z}_4(t_2) - \mathbf{z}_4(t_1) \\ \mathbf{z}_5(t_2) - \mathbf{z}_5(t_1) \\ \mathbf{z}_6(t_2) - \mathbf{z}_6(t_1) \end{bmatrix} = \begin{bmatrix} \Delta \mathbf{z}_3 \\ \Delta \mathbf{z}_4 \\ \Delta \mathbf{z}_5 \\ \Delta \mathbf{z}_6 \end{bmatrix}$$

we obtain in the same way by computation

$$\underline{\mathbf{d}}^\dagger \underline{\mathbf{d}} = \sum_{i=3}^6 (\overline{\mathbf{z}}_i(t_2) - \overline{\mathbf{z}}_i(t_1)) (\mathbf{z}_i(t_2) - \mathbf{z}_i(t_1)) = 0.017771 \text{ m}^2.$$

Here, $\mathbf{z}_i \stackrel{\text{def}}{=} \underline{x}_i + i \underline{y}_i$, and $\overline{\mathbf{z}}_i \stackrel{\text{def}}{=} \underline{x}_i - i \underline{y}_i$ is its complex conjugate.

Let the precisions (mean co-ordinate errors) of the co-ordinates $\underline{x}_i(t_1)$ and $\underline{y}_i(t_1)$ measured at the first epoch be $\sigma_1 = \pm 5 \text{ cm}$, and the precisions of the co-ordinates $\underline{x}_i(t_2)$ and $\underline{y}_i(t_2)$ of the second epoch $\sigma_2 = \pm 1 \text{ cm}$ — for every point, and furthermore the co-ordinates are assumed to be uncorrelated.²¹ The variance matrices of the co-ordinate vectors are thus

$$\Sigma_1 = \sigma_1^2 I, \quad \Sigma_2 = \sigma_2^2 I.$$

We compute the mean error σ_Δ of a single co-ordinate displacement $\Delta \underline{x}_i = \underline{x}_i(t_2) - \underline{x}_i(t_1)$, or, equivalently, $\Delta \underline{y}_i = \underline{y}_i(t_2) - \underline{y}_i(t_1)$. Propagation of

²¹This is obviously unrealistic: in real networks, the point error grows with the distance from the datum points, and the co-ordinate errors within each epoch are strongly correlated.



variances yields

$$\sigma_{\Delta}^2 = \sigma_1^2 + \sigma_2^2 = (25 + 1) \text{ cm}^2 = 26 \text{ cm}^2.$$

Now, the variance matrix of the co-ordinate *displacements* is

$$\Sigma_{\text{dd}} = \Sigma_1 + \Sigma_2 = \sigma_{\Delta}^2 I,$$

with $\sigma_{\Delta} = \sqrt{26} \text{ cm} = 5.1 \text{ cm} = 0.051 \text{ m}$.

Now, we choose for the mean error of unit weight just this value: $\sigma \stackrel{\text{def}}{=} \sigma_{\Delta}$. Then we may also write

$$\Sigma_{\text{dd}} = \sigma_{\Delta}^2 \mathbf{Q}_{\text{dd}} = \sigma^2 \mathbf{Q}_{\text{dd}} = \sigma^2 I,$$

so the weight-coefficient matrix is the unit matrix.

Compute the deformation's *testing variate*, the shifting variate:

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \underline{\mathbf{d}}^T \Sigma_{\text{dd}}^{-1} \underline{\mathbf{d}} = \frac{\underline{\mathbf{d}}^T \mathbf{Q}_{\text{dd}}^{-1} \underline{\mathbf{d}}}{\sigma^2} = \frac{\underline{\mathbf{d}}^T \underline{\mathbf{d}}}{\sigma^2}.$$

Here, $\underline{\mathbf{d}} = \underline{\mathbf{x}}(t_2) - \underline{\mathbf{x}}(t_1)$ is the displacement vector, the abstract vector of co-ordinate differences between the epochs. Because we assume that both co-ordinate sets are given in the same, common datum, the definition points of which nevertheless do *not* belong to the set 3–6, we may assume that all co-ordinates are free. In that case, the number of degrees of freedom is $b = 2n = 8$, where n is the number of points. The variance matrix of the components of the displacement vector, or vector of co-ordinate differences, $\underline{\mathbf{d}}$, is $\sigma^2 I$. We obtain

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \frac{1}{0.0026 \text{ m}^2} (\underline{\mathbf{d}}^T \underline{\mathbf{d}}) = \frac{0.017771 \text{ m}^2}{0.0026 \text{ m}^2} = 6.835.$$

Question The quantity $\underline{\mathcal{E}}/\sigma^2$ is distributed according to the χ_8^2 distribution. If the limit value of this distribution for a significance level of 95% is 15.51 (see [Cooper \(1987\)](#) page 355), has a deformation probably taken place in this case?

Answer No, it has not. $6.835 < 15.51$.

Question If, however, the assumed precisions were $\sigma_1 = \sigma_2 = \pm 1 \text{ cm}$, would then a deformation have probably taken place, at a significance level of 95%?

Answer Yes, it would. $\sigma^2 = (1 + 1) \text{ cm}^2 = 0.0002 \text{ m}^2$ and

$$\frac{1}{\sigma^2} \underline{\mathcal{E}} = \frac{1}{0.0002 \text{ m}^2} (\underline{\mathbf{d}}^T \underline{\mathbf{d}}) = \frac{0.017771 \text{ m}^2}{0.0002 \text{ m}^2} = 88.9 > 15.51.$$





Self-test questions

1. What is the relationship, and difference, between the variance matrix $\Sigma_{\ell\ell}$ and the weight-coefficient matrix $Q_{\ell\ell}$ of the observations?
2. What is the relationship between the *a priori* variance of unit weight σ^2 and the *a posteriori* one $\widehat{\sigma}^2$?
3. What are errors of the first kind and errors of the second kind?
4. What is the power of a statistical test?
5. What is the relationship between the significance level α_{χ^2} of the overall validation test and the significance level α of the outlier test on the individual observations? Why?
6. What is redundancy, and why is it important?
7. What is interior and what is exterior reliability?
8. What are the steps in planning and measuring a geodetic network?
9. What is the Hermitian conjugate of a matrix?
10. Heathrow airport, UK, receives about 35 million incoming international passengers per year. A fancy new system is proposed to be installed that, by analysing the behaviour of people from closed-circuit video, can “flag” them as potential terrorists. The rate of “false positives”, or errors of the first kind, is $1 - \alpha = 1\%$. The rate of errors of the second kind, false negatives or justified but not-called alarms, $1 - \beta$, is believed to be small, less than 50%.

The background to this is that since 1970 there have been some 4000 deaths due to terrorism in the UK.

How would you handle the passengers flagged by the system, and why?

- (a) Have them all killed.
- (b) Arrest and investigate them.
- (c) Send them back to where they came from.
- (d) Have a chat with them and informally look into their backgrounds before doing anything.
- (e) That system is worthless.





Gravity in geodesy

16

Inmiddels was op 6 september 1943 ook het contact met de ‘goede’ Utrechtse hoogleraren tot stand gekomen en werden in nauwe samenwerking met hen hervormingsplannen voor het onderwijs (b.v. betreffende het studium generale) uitgewerkt. Naast prof. dr. J. Jongbloed was vooral prof. dr. ir. F. A. Vening Meinesz, te wiens huize in Amersfoort vele vergaderingen werden gehouden, hierbij een stuwende kracht. In verband met diens studiereizen per onderzeeër werd het contact aangeduid als K18. Vermeldenswaard is, dat tot en met 19 september 1944 33 vergaderingen met de hoogleraren werden belegd.

Reference work *De ondergrondse pers* (“The Underground Press”) 1940–1945 (Winkel and de Vries, 1989, in Dutch)



16.1 Measuring gravity

Gravitation is a fundamental force of the universe. It is an attraction acting between all celestial bodies. According to Newton’s law of gravitation, the force is proportional to the masses of both bodies, and inversely proportional to the square of the distance between the bodies.

Galileo Galilei (1564–1642) was the first to show experimentally, that all bodies fall equally fast: their acceleration in free fall is the same, independently of their mass m .

This may be understood in this way: that as mass m grows, gravitational force F grows, but also the *inertia* of the body m , appearing in the equation $F = ma$, grows, and the acceleration a remains unchanged. One says that the “heavy mass” of the body is the same as its “inertial mass”. hidas massa

However, a broader way to look at this, which makes more theoretical

sense and is more general, is, that there is no difference in principle between gravity and all kinds of *pseudo-forces* — like centrifugal force — caused by non-uniform motion. The Hungarian baron and geophysicist ¹ Loránd Eötvös¹ carried already out many very precise tests to investigate if any difference could be found between “heavy” and “inertial” mass, for example in connection with material composition. The answer was ² consistently “no”.²

From this, Albert Einstein (1879–1955) concluded logically that gravity is a *geometric* property of space-time connected with its curvature, and developed his famous field equations of general relativity, which link together the curvature tensor of space-time and the stress-energy tensor of the matter contained in it. Locally, for example inside a small, closed elevator, it is not possible to know if the gravity sensed inside the elevator is the result of the Earth’s attraction, or of the acceleration produced by a rocket engine in the floor of the elevator!

The measurement and study of the acceleration of gravity is a specialism called gravimetry.

Gravity is the *acceleration of free fall*, and is expressed in the SI unit m/s^2 . On the Earth’s surface, gravity is about 9.8m/s^2 . In gravimetry, however, smaller units are needed, like the milligal (mGal) and the microgal (μGal). In very precise work, even the nanogal (nGal) is encountered.

Unit	In SI units (m/s^2)	As a fraction of gravity (<i>roughly!</i>)
mGal	10^{-5}	10^{-6}
μGal	10^{-8}	10^{-9}
nGal	10^{-11}	10^{-12}

Instruments called *gravimeters* have been built for measuring gravity. An ordinary field gravimeter is in principle just an extremely sensitive spring balance (figure 16.2). Its measurement precision may be 0.01–0.1 mGal. Furthermore, there exist *ballistic* gravimeters, which measure interferometrically the acceleration of a falling body; figure 16.1.

Ballistic gravimeters are *absolute*. Field or spring gravimeters are not *absolute*, and they have a *drift*. This means that the measurement values

käynti

¹Loránd baron Eötvös de Vásárosnamény (1848–1919) was a Hungarian geophysicist and student of gravity.

²Eötvös’ tests have been repeated with even much greater accuracy. The answer did not change.



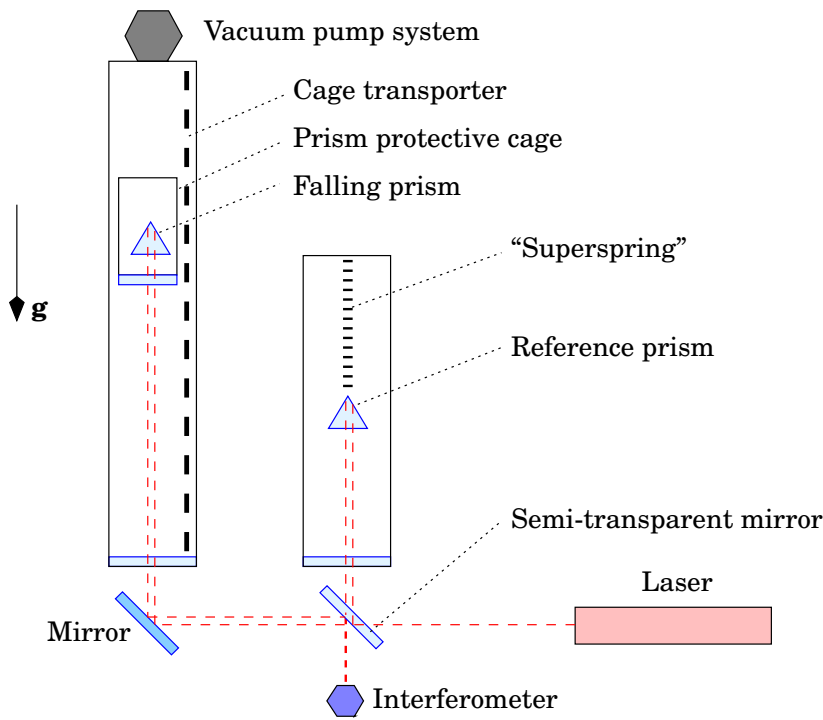


FIGURE 16.1. An absolute or ballistic gravimeter.

produced by the same acceleration of gravity change slowly over time. Therefore, field measurements are planned to always start from a known point and end on a known point. The measurement values are adjusted between the end points in proportion to time according to the drift thus determined.

tasoitus

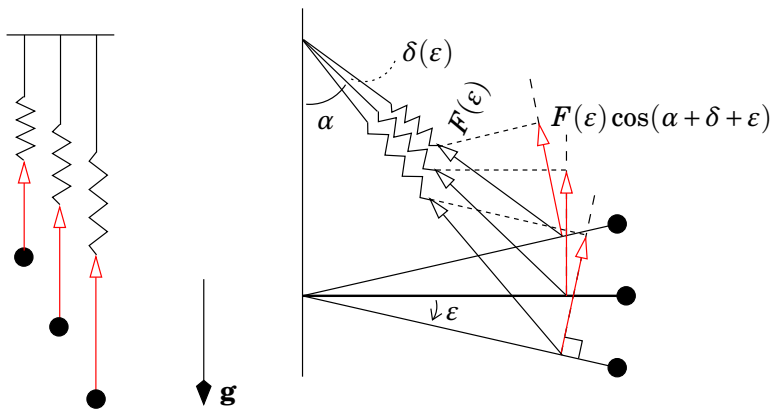


FIGURE 16.2. Principle of operation of a relative or spring gravimeter: increasing sensitivity through a diagonal solution, *astatisation*.



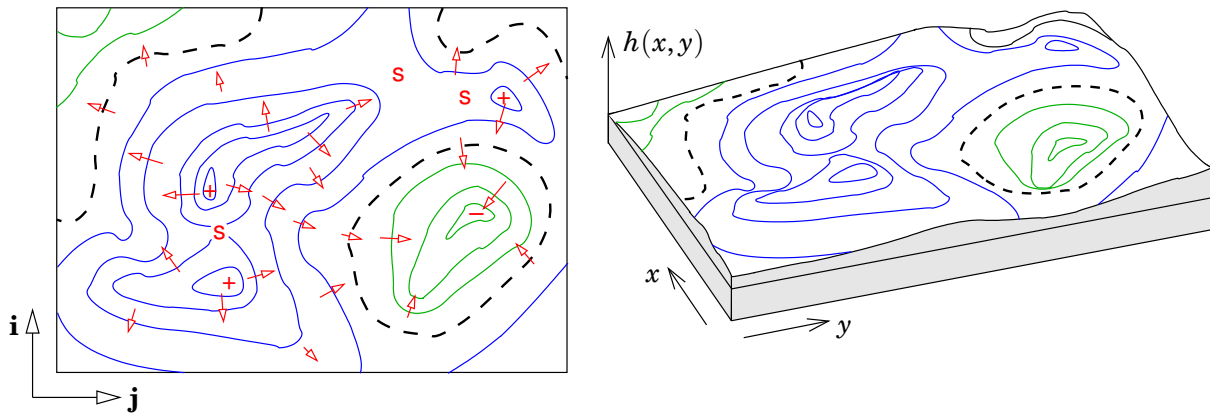


FIGURE 16.3. The terrain height $h(x, y)$ depicted by height contours, and height gradients (arrows). “+”, “-”, local maximum, minimum, “s” saddle point. On the right, the terrain in perspective.



16.2 Gravity and geopotential

16.2.1 The gradient of a scalar field

Figure 16.3 shows how the forms of the terrain are depicted on a map by height contours. In the figure we could have used any scalar function of two variables as an example, instead of the terrain height $h(x, y)$.

³ The figure shows the *gradient* of the height field, the vector field³

$$\mathbf{v}(x, y) = \frac{\partial h(x, y)}{\partial x} \mathbf{i} + \frac{\partial h(x, y)}{\partial y} \mathbf{j} = \partial_x h(x, y) \mathbf{i} + \partial_y h(x, y) \mathbf{j} \stackrel{\text{def}}{=} \text{grad } h = \nabla h$$

as arrows. Here, \mathbf{i} and \mathbf{j} are unit vectors in the x and y co-ordinate directions. This is a vector valued field, the value of which at every point (x, y) consists of two components, the partial derivatives of the height field with respect to the co-ordinates, in this point, in Euler notation $\partial_x h$ and $\partial_y h$.

The gradient vector describes the *slope* of the Earth’s surface: the steeper the slope of the Earth’s surface is, the longer the gradient vector. And the direction of the vector is of course the direction in which the terrain is sloping.⁴

The gradient is always perpendicular to the height contour, which is a set of points having the same height value, an *equi-value curve*. Along it, height is constant.

³The name of the symbol ∇ is *nabla*. This may be an old Greek word for a Phoenician harp, the shape of which it resembles.

⁴Actually the arrows are drawn in the “wrong” direction, the direction into which the Earth’s surface is going down. Thus they describe the vector field $-\mathbf{v}$.



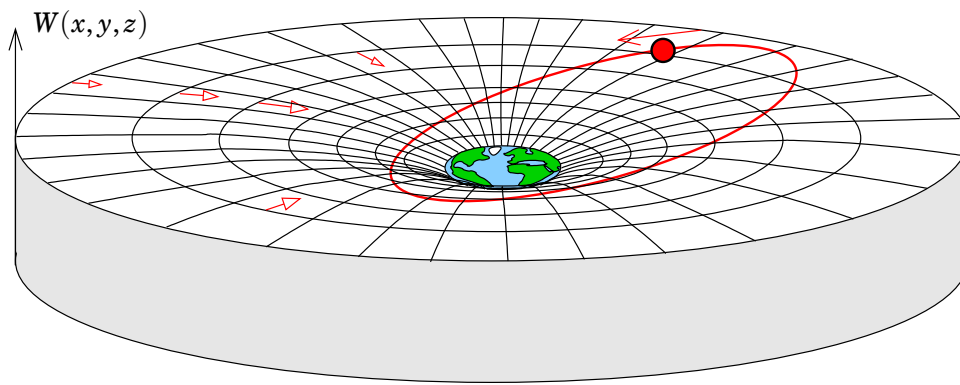


FIGURE 16.4. A geopotential table. Tables like this can be found in science museums. The surface of the table describes the gravity potential of the Earth, albeit only in two dimensions. The arrows again depict the *gradient* of the geopotential, the slope of the table surface.

On the geopotential table, a glass marble can be made to orbit around the “Earth” in an elliptical Kepler orbit, if the figure of the surface agrees sufficiently realistically with Newton’s law of gravitation.

Like the height field $h(x, y)$, we may also visualise the *geopotential* $W(x, y, z)$ in three-dimensional space, with “height contours” or *equipotential surfaces*, and a three-dimensional gradient. On the equipotential surfaces, the value of the geopotential is constant.

Figure 16.4 is a similar visualisation of the geopotential by the curved surface of a table. The distance of the surface from the floor corresponds to the geopotential, representing the energy level of an object on the surface. The circles drawn on the table visualise the equipotential surfaces (in reality three-dimensional) of the Earth’s gravity field, and the curves radiating outwards from the Earth visualise the “lines of force” along which the gradient vector of the potential — in other words, the gravity vector — everywhere points.

voimaviiva

16.2.2 Normal gravity and disturbing potential

Gravity consists of *two parts*:

- the attraction by the Earth’s masses
- the centrifugal force, a pseudoforce caused by the Earth’s rotation.

The contribution of centrifugal force to all of gravity is about -0.3% on the equator, the same order of magnitude as the difference in the gravity between equator and poles.



pyörähdy-
ellipsoidi

The gravity field of the Earth contains all kinds of irregular variations from place to place. Most of the gravity field can however be described as the field of an *ellipsoid of revolution*. This mathematically defined, regular model field, in which the flattening and rotational motion of the Earth have been taken into account, is called the *normal field*.

The lines of force and equipotential surfaces of the normal gravity field are depicted in figure 16.5. The *reference ellipsoid* is one equipotential surface of the normal gravity field, in the same way as the *geoid* (section 16.4) is an equipotential surface of the true gravity field.

The potential of the normal gravity field, the *normal potential*, is written with the symbol $U(x, y, z)$. *Normal gravity* itself is the gradient of this potential.

The gravity vector is the gradient of the geopotential $W(x, y, z)$:

$$\mathbf{g} = \nabla W = \text{grad } W = \frac{\partial W}{\partial x} \mathbf{i} + \frac{\partial W}{\partial y} \mathbf{j} + \frac{\partial W}{\partial z} \mathbf{k} = \partial_x W \mathbf{i} + \partial_y W \mathbf{j} + \partial_z W \mathbf{k},$$

in which \mathbf{i} , \mathbf{j} , and \mathbf{k} are the unit vectors in the x , y , and z directions.

In the same way, the normal gravity vector

$$\boldsymbol{\gamma} = \nabla U = \text{grad } U = \frac{\partial U}{\partial x} \mathbf{i} + \frac{\partial U}{\partial y} \mathbf{j} + \frac{\partial U}{\partial z} \mathbf{k} = \partial_x U \mathbf{i} + \partial_y U \mathbf{j} + \partial_z U \mathbf{k}$$

is also the gradient of the normal gravity potential U .

häiriö-
potentiaali

By subtracting the normal potential from the true gravity potential, the *disturbing potential* is obtained:

$$T \stackrel{\text{def}}{=} W - U.$$

The *strength* of normal gravity is designated by the symbol $\gamma \stackrel{\text{def}}{=} \|\boldsymbol{\gamma}\|$, in the same way as the strength of true gravity $g \stackrel{\text{def}}{=} \|\mathbf{g}\|$. Because the two vectors have nearly identical directions, straight down, we may also write

$$g \approx -\frac{\partial W}{\partial h} = -\partial_h W, \quad \gamma \approx -\frac{\partial U}{\partial h} = -\partial_h U.$$

Normal gravity can be calculated exactly if we know, for a point P , the geodetic latitude φ_P and the height from the reference ellipsoid h_P :

$$\gamma_P = \gamma(\varphi_P, h_P).$$

Normal gravity, like true gravity, diminishes quickly when moving upwards. The rate of diminishing is about 0.3 mGal for every metre. The dependence on latitude is much weaker.



TABLEAU 16.1. Normal potential and normal gravity according to GRS80 (Heikkinen 1981, simplified). In the equations, 9.78... is normal gravity itself, and 0.0000030877... the vertical gradient of normal gravity on the equator on the surface of the reference ellipsoid. Units m, m/s² and m²/s².

$$\begin{aligned}
 U &= 62636860.8500 + \\
 &+ (-9.78032677 - 0.05163075 \sin^2 \varphi - 0.00022761 \sin^4 \varphi - 0.00000123 \sin^6 \varphi) h + \\
 &+ (+0.01543899 - 0.00002195 \sin^2 \varphi - 0.00000010 \sin^4 \varphi) \cdot 10^{-4} \cdot h^2 + \\
 &+ (-0.00002422 + 0.00000007 \sin^2 \varphi) \cdot 10^{-8} \cdot h^3 + 0.00000004 \cdot 10^{-12} \cdot h^4 + \dots, \\
 \frac{\partial U}{\partial h} &= -9.78032677 - 0.05163075 \sin^2 \varphi - 0.00022761 \sin^4 \varphi - 0.00000123 \sin^6 \varphi + \\
 &+ (+0.03087798 - 0.00004390 \sin^2 \varphi - 0.00000020 \sin^4 \varphi) \cdot 10^{-4} \cdot h + \\
 &+ (-0.00007265 + 0.00000021 \sin^2 \varphi) \cdot 10^{-8} \cdot h^2 + 0.00000015 \cdot 10^{-12} \cdot h^3 + \dots.
 \end{aligned}$$

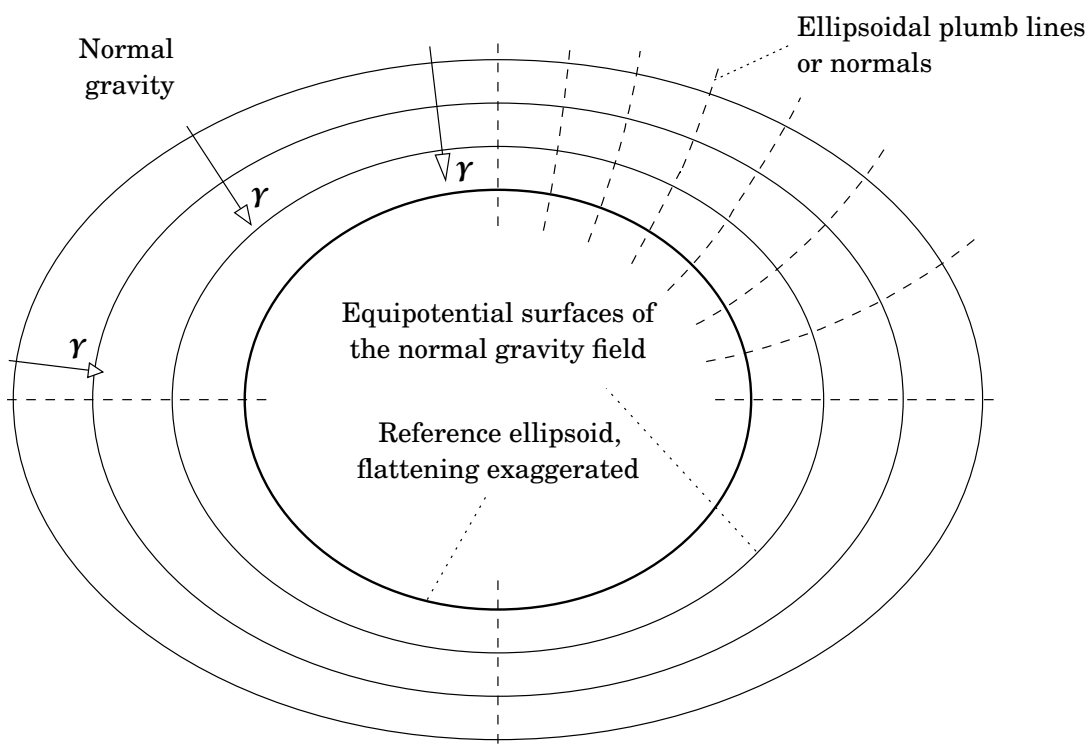


FIGURE 16.5. The normal gravity field of the Earth.



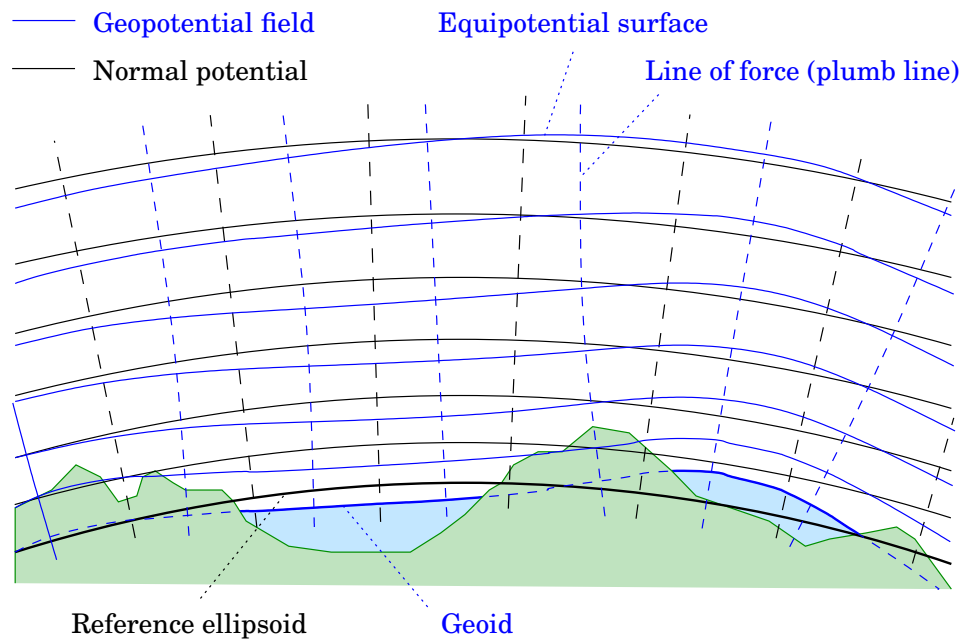


FIGURE 16.6. Equipotential surfaces and lines of force of the geopotential and the normal potential.



16.2.3 Distance between equipotential surfaces

Because the normal gravity field is meant to be an idealised representation of the true gravity field, the equipotential surfaces of both for the same potential value $W = \text{constant}$ and $U = \text{constant}$, for the same constant value, run close to each other. See figure 16.7. The points P and Q lie on the same plumb line: P lies on the surface $W = W_P$ of the W field, whereas Q lies on the corresponding surface $U = U_Q = W_P$ of the U field. So

$$W_P = U_Q$$

and linearisation with respect to height h yields

$$U_P \approx U_Q + \zeta \left. \frac{\partial U}{\partial h} \right|_P = U_Q - \zeta \gamma_P,$$

in which ζ is the distance separating points P and Q .

Subtraction yields the *disturbing potential*

$$T_P \stackrel{\text{def}}{=} W_P - U_P = U_Q - U_P = \zeta \gamma_P \implies \zeta = \frac{T_P}{\gamma_P}. \quad (16.1)$$

⁵ Equation 16.1 is the famous Bruns⁵ equation. The quantity ζ is called the *height anomaly* (of point P). It is the distance between an equipotential

korkeus-anomalia

⁵Ernst Heinrich Bruns (1848–1919) was a gifted mathematician and astronomer, whose greatest achievements were related to the theory of the Earth's gravity field.



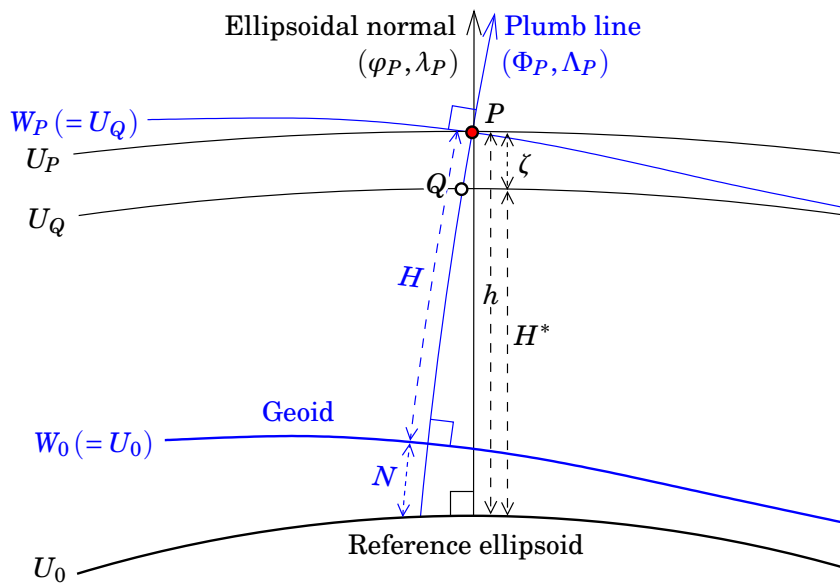


FIGURE 16.7. Equipotential surfaces of true and normal gravity field.

surface of the Earth’s gravity field and the corresponding surface of the normal field, the surface having the same potential. The Bruns equation links this distance directly to the disturbing potential.

When point P is located on the geoid, we have $W_P = W_0$, and Q is located on the reference ellipsoid, so $U_Q = U_0 = W_0$. In this case we use, instead of the notation ζ , the designation N , the *geoid undulation*, the geoid height, the distance of the geoid from the reference ellipsoid. The Bruns equation is in this case

$$N = \frac{T_0}{\gamma_0},$$

in which the values of both T_0 and γ_0 are evaluated on the geoid. In practice, $N \approx \zeta$, except in the mountains. At sea level, $N = \zeta$ exactly.

16.3 Gravity anomalies

In practice, the height h of a point from the reference ellipsoid can be obtained empirically only using satellite positioning.⁶ This is why one writes, using $h_P = H_P + N \approx H_P^* + \zeta$, see figures 16.7 and 16.8:

$$-\left. \frac{\partial T}{\partial h} \right|_P = g_P - \gamma(\varphi_P, h_P),$$

⁶Nowadays, thanks to GNSS, it is easier to obtain the quantity

$$\delta g_P \stackrel{\text{def}}{=} g_P - \gamma(\varphi_P, h_P),$$

which is called the *gravity disturbance*.



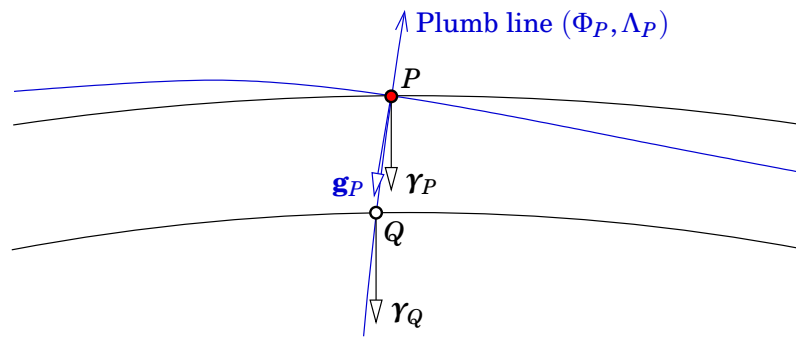


FIGURE 16.8. True and normal gravity vectors.

in which (Taylor series expansion)

$$\gamma(\varphi_P, h_P) \approx \gamma(\varphi_P, H_P^*) + \left. \frac{\partial \gamma}{\partial h} \right|_P \cdot \zeta,$$

so, with the Bruns equation 16.1, we obtain

$$-\left. \frac{\partial T}{\partial h} \right|_P \approx g_P - \gamma(\varphi_P, H_P^*) - \left. \frac{\partial \gamma}{\partial h} \right|_P \cdot \zeta = g_P - \gamma(\varphi_P, H_P^*) - \left. \frac{\partial \gamma}{\partial h} \right|_P \cdot \frac{T_P}{\gamma_P}.$$

From this

$$g_P - \gamma(\varphi_P, H_P^*) = -\left. \frac{\partial T}{\partial h} \right|_P + \left. \frac{\partial \gamma}{\partial h} \right|_P \frac{1}{\gamma_P} T_P. \quad (16.2)$$

This expression is called the *gravity anomaly*, with the definition

$$\Delta g_P \stackrel{\text{def}}{=} g_P - \gamma(\varphi_P, H_P^*). \quad (16.3)$$

The gravity anomaly Δg_P can be computed if two quantities have been measured:⁷

- the gravity value g_P at point P gravimetrically
- the height $H_P \approx H_P^*$ of the point from the geoid, or “above sea level”.

Most often — and before the satellite era, always — the height of a gravity measurement point is determined by reading from a map, from a **ilmapuntari** photogrammetric stereo model, by using a barometer, or from traverse **linjavaaitus** levelling. In all cases, one obtains just the height H_P above sea level. The precision of measurement of the heights varies from several centimetres to around a metre. This uncertainty propagates straight into the anomaly values Δg , by the vertical gradient of normal gravity $\partial_h \gamma \approx -0.3 \text{ mGal/m}$.

The gravity anomaly Δg is an empirical quantity that can be calculated from measurements made on the Earth’s surface.

ilma-anomalia The most commonly used gravity anomaly is the *free-air anomaly* Δg ,

⁷For the latitude φ_P , an approximate value is good enough.



the definition of which was given as 16.3. Gravity anomaly values vary in the interval ± 100 mGal, more rarely, for example in the mountains, ± 200 mGal. Within Finland, the range of variation is ± 60 mGal. Gravity anomalies and their variations from place to place describe the irregularities of the interior mass distribution of the Earth, and are therefore of geophysical and geological interest.

It may be mentioned as a curiosity, that the Dutch researcher Felix A. Vening Meinesz⁸ found, south of the island of Java, at the Java (today Sunda) deep-sea trench, a large deficiency in gravity. We know today that deep-sea trenches are those places on the Earth's surface where, as a part of plate tectonics, the Earth's oceanic crust "dives" into the Earth's mantle where it will be geologically recycled: *subduction*, see figure 18.7.

8
syvänmeren
hauta
alityöntö



16.4 The gravimetric geoid

From equation 16.2 above together with definition 16.3 is obtained

$$\Delta g = -\frac{\partial T}{\partial h} + \frac{\partial \gamma}{\partial h} \frac{1}{\gamma} T = -\partial_h T + \frac{\partial_h \gamma}{\gamma} T, \quad (16.4)$$

which is called the *fundamental equation of physical geodesy*, Heiskanen and Moritz (1967) equation 2-148. So:

fysikaalisen
geodesian
perusyhtälö

The gravity anomaly Δg is a linear combination of disturbing potential T and its vertical derivative of location $\partial_h T$.

The quantity $\partial_h \gamma \approx -0.3$ mGal/m is the, already mentioned, *vertical gradient of normal gravity*.

The *geoid*, the equipotential surface of the Earth's gravity field that describes the figure of the whole field (the "mathematical figure of the Earth", Gauss), may be determined by gravimetric means, starting from the fundamental equation of physical geodesy 16.4. Let us assume the Earth to be a sphere. Then

$$\gamma = \frac{GM}{R^2}$$

and by differentiation

$$\frac{\partial \gamma}{\partial h} = \frac{\partial \gamma}{\partial R} = -2 \frac{GM}{R^3},$$

⁸Felix Andries Vening Meinesz (1887–1966) was a Dutch geophysicist, geodesist and gravimetrist. He wrote together with V. A. Heiskanen the textbook *The Earth and its Gravity Field* (1958).



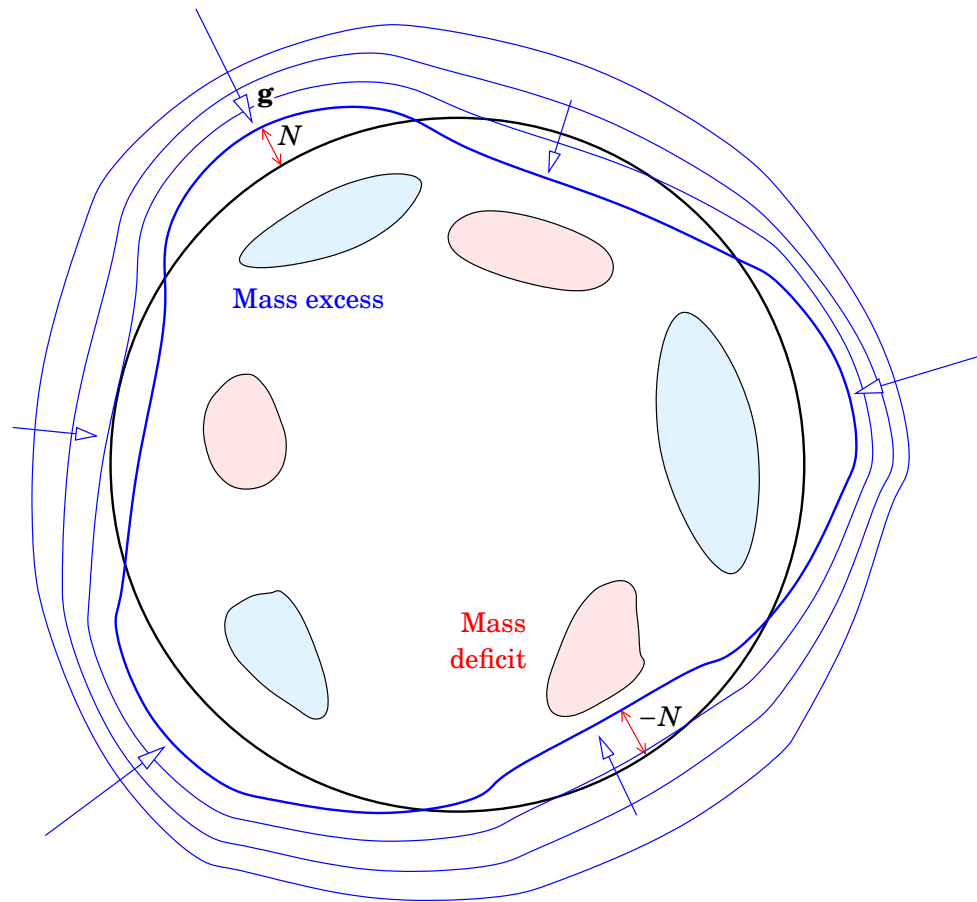


FIGURE 16.9. Relationship between variations in the Earth's gravity and those in geoid height.

from which follows

$$\Delta g = -\frac{\partial T}{\partial h} - \frac{2}{R}T,$$

an equation valid on the surface of a spherical Earth.

It is intuitively clear that there is some kind of link between variations of gravity and variations of the geopotential on the Earth's surface. Both are caused by the uneven distribution of masses inside the Earth. As figure 16.9 shows, the excess masses inside the Earth will cause both an excess in gravity (the level surfaces of the geopotential will be closer together) and a rising of the geoid above the surface of the reference ellipsoid, whereas, on the other hand, mass deficiencies inside the Earth will lead to both a shortfall in gravity and a depression of the geoid below the reference ellipsoid.

The relationship between gravity anomalies Δg and geoid heights N is however not simple. The quantities are connected by an integral equation, the *Stokes equation*. George Gabriel Stokes⁹ derived the following



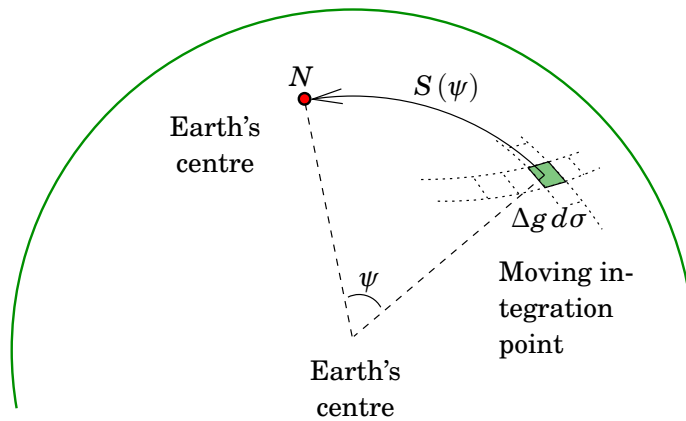


FIGURE 16.10. The geometry of the Stokes integral equation.

classical integral equation¹⁰ in 1849:

10

$$N = \frac{R}{4\pi\gamma} \iint S(\psi) \Delta g d\sigma. \tag{16.5}$$

With the equation, we may compute geoid heights from the global field of gravity anomalies. In the equation, R is the radius of the spherical Earth, γ mean gravity on the Earth's surface, and $S(\psi)$ is the *Stokes function*, also called the Stokes kernel. It depends only on the geocentric angular distance ψ between the evaluation point for the geoid height N and the measurement point for the anomaly Δg seen from the centre of the Earth:

$$S(\psi) = \frac{1}{\sin \psi/2} - 6 \sin \frac{\psi}{2} + 1 - 5 \cos \psi - 3 \cos \psi \ln \left(\sin \frac{\psi}{2} + \sin^2 \frac{\psi}{2} \right).$$

Look more closely at equation 16.5. The quantity $d\sigma$ is the *solid-angle element*, a surface element on a sphere of unit radius, in spherical co-ordinates $d\sigma = \cos \phi d\phi d\lambda$.

[avaruuskulma](#)

Open up the equation in the following way:

$$N(\phi, \lambda) = \frac{R}{4\pi\gamma} \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} S(\psi(\phi, \lambda, \phi', \lambda')) \Delta g(\phi', \lambda') \cos \phi' d\phi' d\lambda'.$$

⁹George Gabriel Stokes (1819–1903) was a gifted English mathematician, physicist and geophysicist.

¹⁰Deriving the equation is difficult and uses the fundamental equation of physical geodesy 16.4 as a *boundary condition* for solving the Laplace field equation in the space exterior to the Earth. See [Heiskanen and Moritz \(1967\)](#) chapter 2.



In this, (ϕ, λ) are the co-ordinates — strictly speaking, the *geocentric* latitude and longitude — of the point at which the geoid height N is being computed. The co-ordinates (ϕ', λ') are again the latitude and longitude of the point at which the gravity anomaly Δg is given. This point traverses the whole Earth's surface with the computation of the double integral. The angle ψ is the angular distance between these two points as seen

¹¹ from the centre of the Earth.¹¹

We see here that computing even a single value N requires Δg values from everywhere on the Earth's surface, in order to evaluate the above integral completely. Divide the Earth's surface into cells, or *blocks*, of size $1^\circ \times 1^\circ$ — in total $360 \times 180 = 64800$ of them — and calculate the value of the integral numerically as a sum

$$N(\phi, \lambda) = \frac{R}{4\pi\gamma} \left(\frac{\pi}{180} \right)^2 \sum_{i=1}^{360} \sum_{j=-89}^{+90} S(\psi(\phi, \lambda, \phi', \lambda')) \Delta g(\phi', \lambda') \cos \phi', \quad (16.6)$$

in which $\lambda' = i^\circ - 0.5$ and $\phi' = j^\circ - 0.5$. Over the whole of the Earth's surface, there are $360 \times 180 = 64800$ different values of N to be calculated, if the desired resolution of the geoid model is also $1^\circ \times 1^\circ$.

Now one also understands why close international collaboration is so essential to studying the Earth's gravity field!

In practice, the greatest influence on the computed geoid height is by *local* gravity anomalies, the values of anomalies in the immediate vicinity of the point of evaluation. The function $S(\psi)$ is for small values of ψ approximately

$$S(\psi) \approx \frac{2}{\psi}.$$

More remote areas also have an effect, but to take them into account, it suffices to use a lower-resolution *global geopotential model* produced by satellite geodesy.

Globally, the geoid deviates from the reference ellipsoid by about ± 100 m. The global mean sea surface in its turn follows the geoid, because it is an equipotential or equilibrium surface. The mean sea surface

¹¹The equation for calculating the angular distance is

$$\cos \psi = \sin \phi \sin \phi' + \cos \phi \cos \phi' \cos(\lambda' - \lambda),$$

or more precisely for small angles ψ , the half-angle formula 3.9:

$$\sin^2 \frac{\psi}{2} = \sin^2 \frac{\phi' - \phi}{2} + \cos \phi \cos \phi' \sin^2 \frac{\lambda' - \lambda}{2}.$$



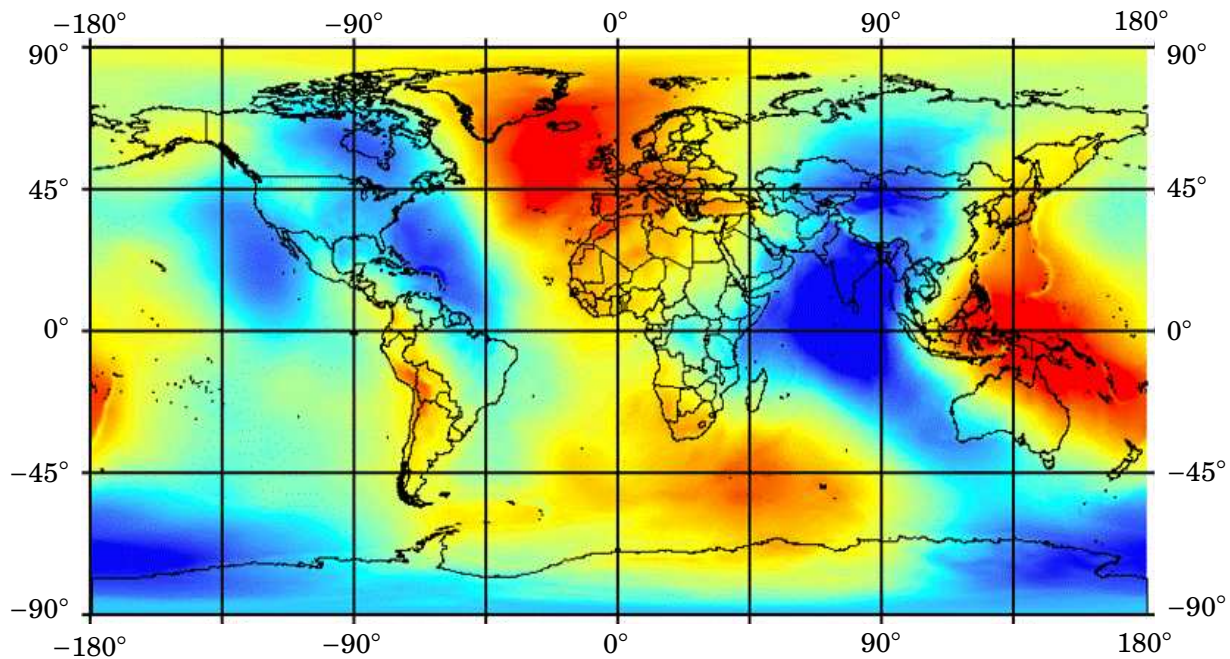


FIGURE 16.11. The global geoid model [EGM2008](#). Geoid heights reckoned from the [GRS80](#) reference ellipsoid range from -107 m (blue) to $+86$ m (red). © 2013 US National Geospatial-Intelligence Agency.

deviates from the geoid by at most ± 2 m. The permanent part of the deviation is called the *sea-surface topography*, see section 1.6. In addition, [meritopografia](#) there are deviations varying in time, like the phenomenon of the tides and the deviations caused by winds and air-pressure variations.

16.5 The gravity field and heights

16.5.1 Geopotential and gradient

In figure 16.3 we see how one can depict the forms of the terrain on a map by means of *height contours*.

The most natural of all measures of height, the *geopotential*, is not a metric height. It is a measure of *energy*: it describes the level of potential energy of a test body, a unit mass, in the gravity field of the Earth. This makes it a geophysically sensible quantity.

The geopotential is connected to gravity in this way, that the gravity vector, \mathbf{g} , is the derivative of place, or *gradient*, of the geopotential W , see figure 16.12:

$$\mathbf{g} = \nabla W = \text{grad } W = \frac{\partial W}{\partial x} \mathbf{i} + \frac{\partial W}{\partial y} \mathbf{j} + \frac{\partial W}{\partial z} \mathbf{k}, \quad (16.7)$$



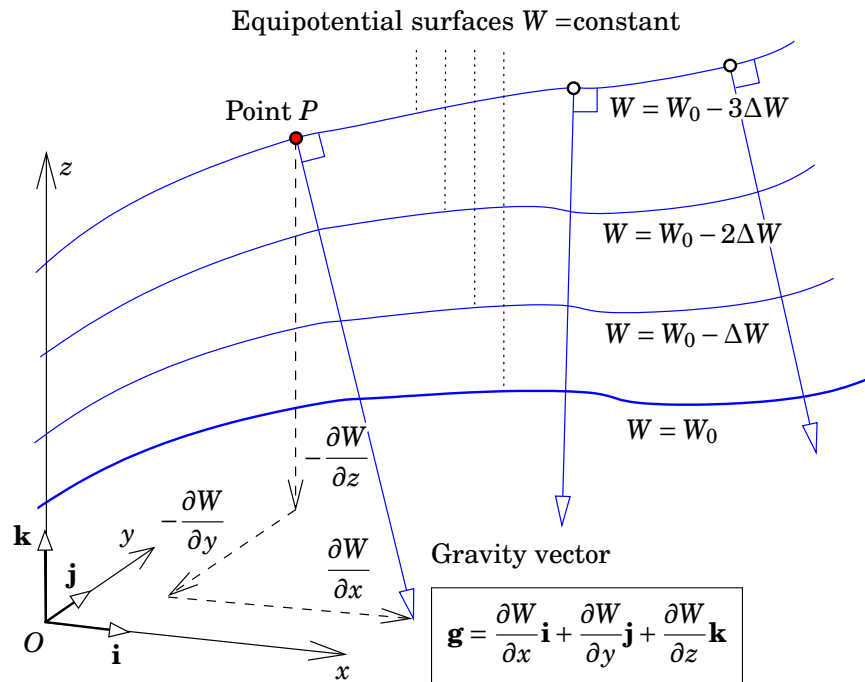


FIGURE 16.12. The gravity vector is the gradient of the geopotential, the derivative with respect to the three co-ordinates of location.

in which $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ are unit vectors in the directions of the axes of the ortonormaali kanta (x, y, z) co-ordinate frame, forming an orthonormal basis.

Because of this, the local gravity vector has the following properties:

- It is always *perpendicular* to equipotential surfaces — also to sea level!
- The closer to each other the equipotential surfaces are, the longer it is.

The gravity field is a *conservative* field. This means that, when one transports a test mass around a closed path, no net work is done. In a conservative force field, the force vector can always be expressed as the gradient of a potential, in the way depicted in figure 16.12.

The difference in potential between points A and B is now the same as the work done by a unit mass moving from point A to point B , and the following integral applies (s is the path length along AB):

$$\begin{aligned}
 W_B - W_A &= \int_A^B dW = \int_A^B \frac{dW}{ds} ds = \int_A^B \left(\frac{\partial W}{\partial x} dx + \frac{\partial W}{\partial y} dy + \frac{\partial W}{\partial z} dz \right) = \\
 &= \int_A^B \langle \text{grad } W \cdot d\mathbf{x} \rangle = \int_A^B \langle \mathbf{g} \cdot d\mathbf{x} \rangle. \quad (16.8)
 \end{aligned}$$

From this it can be seen that the work is the scalar product of the gravity



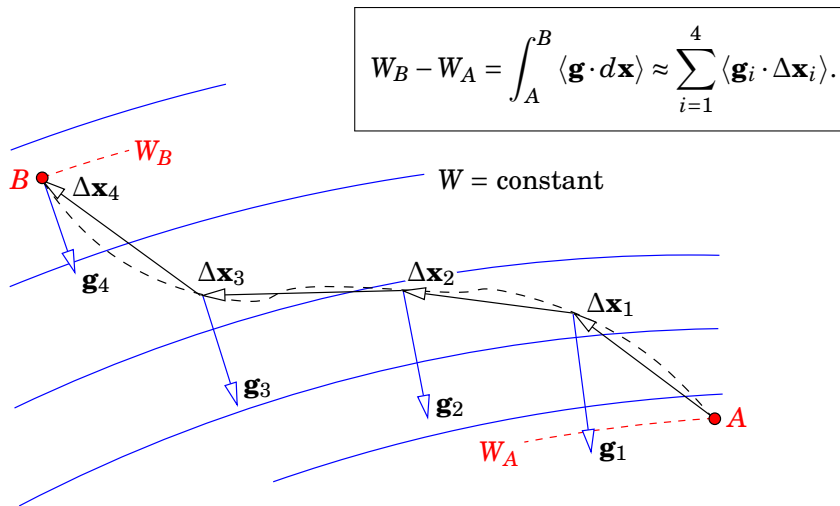


FIGURE 16.13. The path integral of work.

vector \mathbf{g} and the path vector element along the direction of the path $d\mathbf{x} \stackrel{\text{def}}{=} \mathbf{i} dx + \mathbf{j} dy + \mathbf{k} dz$.

In practice, often instead of the geopotential itself W , its difference $C \stackrel{\text{def}}{=} -(W - W_0)$ with the geopotential of mean sea level (or some other suitable reference surface) W_0 is used. This potential difference, which grows in the upward direction, is called the *geopotential number*,¹² and the above integral equation becomes

$$C_B - C_A = - \int_A^B \langle \mathbf{g} \cdot d\mathbf{x} \rangle.$$

In the case of a closed path, we have

$$\oint \langle \mathbf{g} \cdot d\mathbf{x} \rangle = 0.$$

Geopotential numbers are calculated from the measurement results of a *levelling* extending over the country. All metric heights of terrain points, like orthometric height, are calculated from the geopotential numbers of the points.

16.5.2 Geopotential unit, GPU

As we saw in section 16.1, gravity is expressed in the SI unit m/s^2 . We again use the *geopotential unit*, or GPU, as the measurement unit for the geopotential. The SI unit of geopotential is m^2/s^2 : distance \times force/mass

¹²German *geopotentielle Knoten*, French *cotes géopotentielles*.



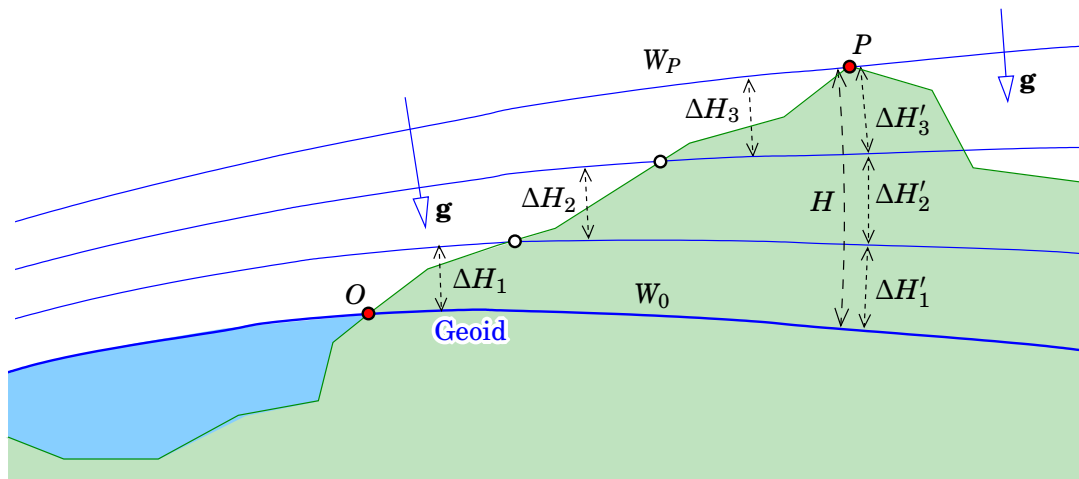


FIGURE 16.14. Heights and equipotential surfaces. The stronger the gravity $g = \|\mathbf{g}\|$ is — always perpendicular to the equipotential surfaces — the closer together are the equipotential surfaces.



= distance \times acceleration, unit $\text{m} \times \text{m}/\text{s}^2$. In the gravity field of the Earth close to the surface, where the acceleration of gravity is $g \approx 9.8 \text{ m}/\text{s}^2$, a height difference of one metre corresponds to a potential difference of about $9.8 \text{ m}^2/\text{s}^2$.

Define

$$1 \text{ GPU} \stackrel{\text{def}}{=} 10 \text{ m}^2/\text{s}^2,$$

then a height difference of one metre corresponds to a potential difference of about 0.98 GPU , and similarly a potential difference of 1 GPU corresponds to a height difference of about 1.02 m .

Thus one can, thanks to the fortuitous circumstance that g is close to $10 \text{ m}/\text{s}^2$, express geopotential differences in a unit that is slightly more intuitive than the corresponding **SI** unit!



16.5.3 Orthometric heights

Orthometric heights H were introduced already. Let us take a closer look.

In figure 16.14 the orthometric height of point P is H . In this simple example, it is the sum of three height differences:

$$H = \Delta H'_1 + \Delta H'_2 + \Delta H'_3, \quad (16.9)$$

in which the $\Delta H'_i$, $i = 1, 2, 3$ are the separations between the equipotential surfaces on the plumb line of the point.

Levelling, however, yields the height differences ΔH_1 , ΔH_2 , and ΔH_3 on the terrain, on the Earth's surface between point and coast. In this



case a levelling has been carried out from coastal point O , the height of which is assumed to be zero. Now,

$$H \neq \Delta H_1 + \Delta H_2 + \Delta H_3!$$

Height differences obtained from levelling may not just be added together to get the height of a point.

This tells us that height, though a metric quantity, is not a very nicely behaved quantity.

For this reason, in scientific work we always use, instead of metric heights, the already-presented *geopotential numbers* $C = -(W - W_0)$.

We may simply write (note that the units also match):

$$\begin{array}{c} \Delta C \\ [\text{m}^2/\text{s}^2] \end{array} = \begin{array}{c} g \\ [\text{m}/\text{s}^2] \end{array} \cdot \begin{array}{c} \Delta H \\ [\text{m}] \end{array}$$

in which ΔC is the *geopotential difference* (between two arbitrary points), the work that needs to be done in order to transport one unit mass over the height difference ΔH between the points.

If in point O it holds that $W = W_0$, it follows that $C_O = 0$. Then, in the example case of figure 16.14, the geopotential number of point P is

$$C = \Delta C_1 + \Delta C_2 + \Delta C_3 = g_1 \Delta H_1 + g_2 \Delta H_2 + g_3 \Delta H_3, \quad (16.10)$$

which is *computable* if, along the levelling line, local gravity g has also been measured.

However, the following also holds:¹³

13

$$C = g'_1 \Delta H'_1 + g'_2 \Delta H'_2 + g'_3 \Delta H'_3,$$

in which the g'_i , $i = 1, 2, 3$ are gravity values *inside the rock*, on the plumb line of point P .

Define the mean gravity along the plumb line by the following equation:¹⁴

14

¹³In the general case the equation is

$$C = \int_0^H g(z) dz,$$

in which z is the arc length measured along the plumb line.



$$\bar{g} \stackrel{\text{def}}{=} \frac{g'_1 \Delta H'_1 + g'_2 \Delta H'_2 + g'_3 \Delta H'_3}{\Delta H'_1 + \Delta H'_2 + \Delta H'_3} = \frac{C}{H}.$$

It follows that

$$C = \bar{g}H \iff H = \frac{C}{\bar{g}},$$

the classical definition of orthometric heights. The equation tells that the amount of work needed to move a unit mass from the geoid to point P , is force \times distance:

$$C = \overset{\text{work}}{C} = \overset{\text{force}}{\bar{g}} \cdot \overset{\text{distance}}{H}.$$

We are left with the problem of determining \bar{g} , the mean of gravity along the plumb line. Measuring values g'_i inside the Earth's crust is usually impossible. . . . Therefore, in practice the determination is based on the value g_P measured on the Earth's surface, by *assuming* that gravity grows going downwards, inside the Earth's crust, according to a certain formula.¹⁵ In this way, an approximate value for the orthometric height is obtained, the accuracy of which, at least for the Finnish territory, is totally adequate.

Orthometric height is but one way of building a metric height system. There are other ways, like normal height and dynamic height. All are heights "above sea level", but the methods of definition and calculation are slightly different. And all three have their own pluses and minuses.

Orthometric heights are not without their problems. The tunnel network of figure 4.2 does not exist, and measuring gravity inside the rock — along the local plumb line — is normally not possible. In practice, orthometric heights are determined with the aid of *levelling*, starting from the coast, along the Earth's surface. If we want to calculate precise orthometric heights from levelling, we unfortunately need detailed data on

- the density of the rock below the height point
- the forms of the terrain around the height point: a *terrain model*.

So, even if orthometric heights are physically elegant, their precise determination may in practice be troublesome. Scientifically one says that orthometric heights are not *hypothesis-free*. The hypotheses required are

¹⁴The general equation is

$$\bar{g} = \frac{C}{H} = \frac{1}{H} \int_0^H g(z) dz.$$

¹⁵For example, Poincaré-Prey reduction, see Heiskanen and Moritz (1967) section 4-2.



precisely the density of the Earth's crust and the terrain's local forms. In practical computation, the effect of the terrain is often omitted, and density values are taken from geological maps. The error thus made is usually small.

16.5.4 Normal heights

Normal heights H^* are, simply stated, orthometric heights computed from geopotential numbers C as if the true gravity field of the Earth were a regular mathematical model field based on an ellipsoid of revolution: a *normal gravity field*. Therefore, no information is needed related to the true, complicated gravity field. Normal heights are computed easily and precisely without any knowledge of local rock density or terrain model.

The equation for normal height is

$$H^* = \frac{C}{\bar{\gamma}},$$

in which $\bar{\gamma}$ is the mean of *normal gravity*, calculated again along the plumb line of the point.¹⁶

16

However, unlike orthometric heights, normal heights have no direct physical interpretation.

On most of the Finnish territory, the differences between orthometric and normal heights are of the order of millimetres. In the mountains globally, they can easily be several decimetres.

In many countries — among others Russia, Sweden, and nowadays also Finland — normal heights are used instead of orthometric heights. Their precise calculation is easier. Whereas orthometric heights are interpreted as heights from the geoid, normal heights are reckoned from a similar surface called the *quasi-geoid*. On the sea, this surface coincides with the geoid and thus also with mean sea level, but under the land, and especially under the mountains, it differs from the geoid. The earlier proposed “tunnel-network metaphor”, letting in sea water under the continents, is not appropriate for the quasi-geoid.

16.5.5 Dynamic heights

Dynamic heights are rarely used. They are calculated simply by dividing the geopotential number C by the normal gravity at latitude 45° at zero height, γ_{45} , which is a constant:

$$H^{\text{dyn}} = \frac{C}{\gamma_{45}}.$$

¹⁶Now, the averaging is done over the range of normal heights $(0, H^*)$.



TABLE 16.2. Properties of various height types.

Height type	Correctness		Hypothesis	
	Metric	Energetic	freeness	Equation
Geopotential number	--	+	+	C
Orthometric	+	-	-	$H = C/\bar{g}$
Normal	-	-	+	$H^* = C/\bar{\gamma}$
Dynamic	--	+	+	$H^{\text{dyn}} = C/\gamma_{45}$



16.5.6 Properties of different height types

It is common among all height types, that a metric height is obtained by *dividing* the geopotential number by some suitable *gravity value*: the SI unit of geopotential numbers is m^2/s^2 , the unit of acceleration of gravity is m/s^2 , and the unit of metric height is indeed $(\text{m}^2/\text{s}^2)/(\text{m}/\text{s}^2) = \text{m}$, as it should be.

Independently of height type, all metric heights are computed from the energy level of the point, the aforementioned *geopotential number* C . The only “heights” that can be measured and computed precisely are geopotential numbers $C = -(W - W_0)$. All other heights are *derived quantities*. Besides precision, some useful properties are always lost in their computation, like when projecting a curved surface onto a flat one.

The user of height values desires from practical heights a number of good properties, familiar from geometric heights within a small area:

oikeellisuus Metric correctness Metric correctness means that, if there are two points P and Q straight above each other, and the distance between them is 1 m, then $H_P - H_Q$ is also precisely 1 m. Only *orthometric heights* have this property. The metric correctness of dynamic heights is especially weak.

The closer to the *true* mean gravity along the plumb line the expression in the denominator of the formula is, the better the metric correctness of the resulting height type will be.

Energetic correctness This means that water always flows “down” in the sense of the height type in question. Of the three types mentioned, only *dynamic heights* are energetically correct — by virtue of their direct proportionality to geopotential numbers C .

Exact calculability, dependence on uncertain hypotheses

Normal heights and dynamic heights may be precisely cal-



culated based on theory. However, the normal field, or reference ellipsoid, chosen for the computations needs to be stated.

Orthometric heights require knowledge of both the true gravity field and the form and density of the topography. Especially the assumed density is an uncertain hypothesis. In practice, however, the uncertainty caused by these factors is fairly small.



16.6 Bouguer anomalies

Earlier we noted that free-air anomalies — equation 16.3 — tell us something about the interior mass distribution of the Earth. However, the whole topography underneath and around a point also affects the free-air anomaly Δg of the point. The forms of the topography above sea level are visible and usually well-known. Therefore it would seem logical to *remove computationally* the effect of the terrain forms from the free-air anomalies, in order to obtain a quantity that tells us only about the mass distribution of the Earth below sea level.

This is how the *Bouguer anomaly* is obtained:

$$\Delta g_B = \Delta g_{FA} - g_{top},$$

in which $\Delta g_{FA} \stackrel{\text{def}}{=} \Delta g$ is the free-air anomaly, Δg_B the Bouguer anomaly and g_{top} the vertical component of the attraction of the topography acting at the point.

Bouguer anomalies may be calculated precisely or approximately. In the first case we use a numerical model of the topography, a *digital terrain model (DTM)*. We also use a density model for the Earth's crust if one exists. In the approximate calculation, we only take into account the effect of the *Bouguer plate*, as a simple closed formula:

$$g_{top} = 2\pi G\rho d,$$

in which

- g_{top} attraction of the plate, only in the vertical direction
- G Newton's universal gravitational constant, see section 1.2
- ρ density of the matter of the plate
- d thickness of the plate, assumed to be equal to the local height H of the terrain.



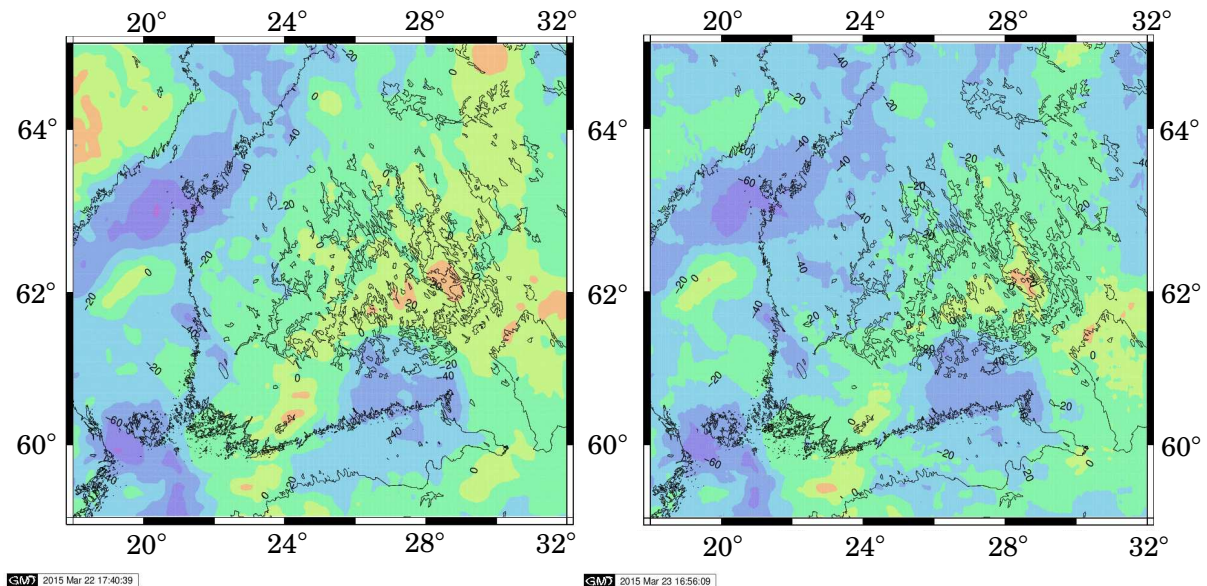


FIGURE 16.15. Free-air and Bouguer anomalies for Southern Finland computed from the EGM2008 geopotential model. Data © Bureau Gravimétrique International (BGI) / International Association of Geodesy.



The plate is assumed to be flat and to extend to infinity in all directions.

If the density is $\rho = 2.67 \text{ g/cm}^3 = 2670 \text{ kg/m}^3$, we obtain

$$g_{\text{top}} = 0.1119d,$$

in which d is in metres and g_{top} in milligals. The milligal unit was explained at the beginning of the chapter.

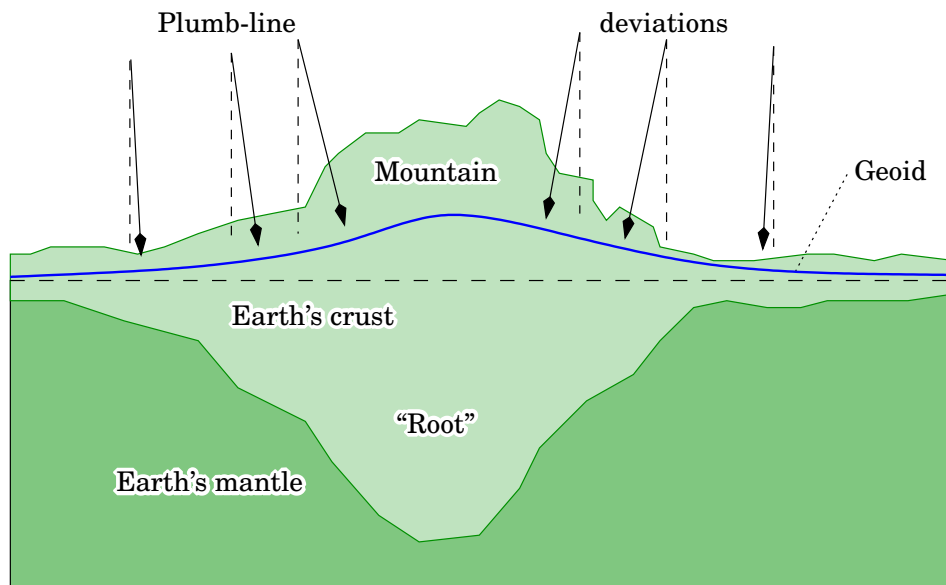



16.7 Astronomical position determination

The local plumb line or direction of the local gravity vector stands perpendicular upon equipotential surfaces. Determining the direction of the plumb line in an absolute sense has been possible by traditional astronomical means.

One speaks of *astronomical position determination* because the first practical application of the method was determining an unknown position, especially at sea. Later, the method was used to study geophysically interesting variations (deviations) of the direction of the plumb line on land. The already mentioned Pierre Bouguer noticed in South America, like George Everest in India, that in the vicinities of mountain ranges the plumb line is deflected towards the mountain range, and interpreted this correctly as being caused by the mountains' own gravitation.





 FIGURE 16.16. The root of a mountain range and its effect on the plumb line.

The attempt to estimate the effect of the mass of the mountains, however, produced a result that was much larger than the actually observed plumb-line deviations. The reason for this is today known to be *isostatic compensation*: under the mountains there is a *root* consisting of lighter rock that keeps the mountain “afloat” on the plastically deforming Earth’s mantle.

luotiviivan
poikkeama

Various instruments are used for astronomical position determination, such as the *meridian circle*, the *astrolabe*, and the *zenith tube*.

The optical axis of an astrolabe points always upwards by a fixed angle. The optical axis of a zenith tube again always points upwards vertically, to the zenith, under an elevation angle $\eta = 90^\circ$. Therefore, by observing the passage through the zenith of stars of which the declination δ is known, the astronomical latitude Φ of the location is obtained by $\Phi = \delta$. The time of transit through the meridian is measured at the same time, from which is obtained the astronomical longitude Λ — because the zenith direction lies in the plane of the meridian. A zenith tube is in a way both a meridian circle and an astrolabe.

korkeuskulma

meridiaanin
läpikulku

In preparation for observations, a *star programme* is drafted, a list of stars that will transit the meridian, a certain elevation circle,¹⁷ or the zenith. In case of the astrolabe, one should take care that the stars are distributed evenly around the whole horizon. In this way, a precise

korkeuskulma-
piiri
17

¹⁷In Arabic *almucantar*.



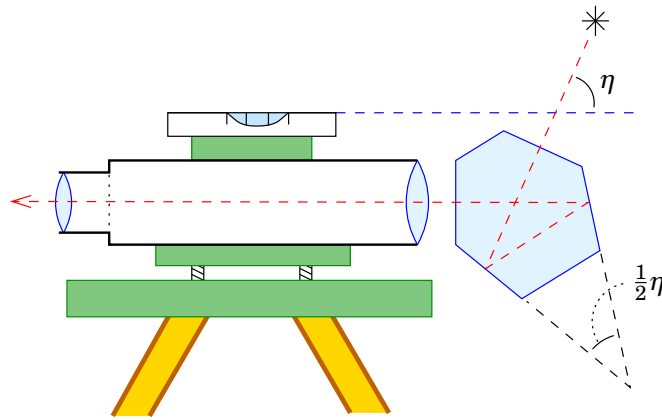


FIGURE 16.17. A levelling instrument converted to astrolabe.

determination of both Φ and Λ is achieved, and the impact of atmospheric refraction minimised.

An astrolabe may be built easily from a levelling instrument, by adding a sixty-degree angle prism in front of the objective; see figure 16.17. The most precise (Danjon) astrolabes again use a mercury mirror together with a sixty-degree prism.



16.8 Measuring the gravity gradient

We have several times mentioned the *potential* of gravity, which is a measure of the *potential energy* (energy content of location) of a test mass inside the gravity field. The acceleration vector of gravity, or free fall, is defined as the *gradient* of this geopotential W , its rate of change with place:

$$\begin{aligned} \mathbf{g} &= \text{grad } W = \nabla W = \\ &= g_x \mathbf{i} + g_y \mathbf{j} + g_z \mathbf{k} = \frac{\partial W}{\partial x} \mathbf{i} + \frac{\partial W}{\partial y} \mathbf{j} + \frac{\partial W}{\partial z} \mathbf{k} = \partial_x W \mathbf{i} + \partial_y W \mathbf{j} + \partial_z W \mathbf{k}, \end{aligned}$$

where the *grad operator* is

$$\text{grad } \cdot = \nabla \cdot = \frac{\partial \cdot}{\partial x} \mathbf{i} + \frac{\partial \cdot}{\partial y} \mathbf{j} + \frac{\partial \cdot}{\partial z} \mathbf{k} = \partial_x(\cdot) \mathbf{i} + \partial_y(\cdot) \mathbf{j} + \partial_z(\cdot) \mathbf{k}.$$

Here $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is an orthogonal triad of unit vectors, or *orthonormal basis*, oriented along the three axes of the (x, y, z) co-ordinate frame.

The acceleration vector of gravity is thus the *gradient of the geopotential*. This vectorial quantity is location-dependent. We know that gravity grows going downwards, at least in free air. Going up, gravity diminishes by some 0.3 mGal for every metre in height.



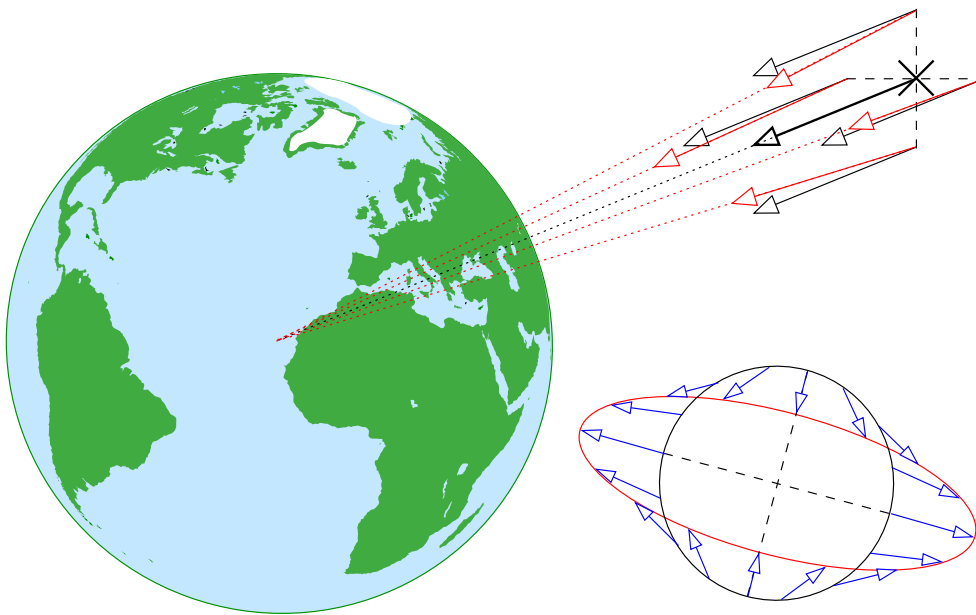


FIGURE 16.18. The gravitational-gradient or tidal force field: gravitation varies with place. The lower diagram gives the *visual ellipsoid* of the gravitational gradient tensor: it depicts the force field inside a freely falling object, for example a satellite.



Gravity as a vector varies in a more complicated way in the vicinity of masses, if one has the use of sufficiently accurate measurement devices. We speak of the *gravity-gradient tensor*, or Eötvös tensor. Written in both the traditional and the more compact Euler notation:

$$\mathcal{M} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial x \partial y} & \frac{\partial^2}{\partial x \partial z} \\ \frac{\partial^2}{\partial y \partial x} & \frac{\partial^2}{\partial y^2} & \frac{\partial^2}{\partial y \partial z} \\ \frac{\partial^2}{\partial z \partial x} & \frac{\partial^2}{\partial z \partial y} & \frac{\partial^2}{\partial z^2} \end{bmatrix} W = \begin{bmatrix} \partial_{xx} & \partial_{xy} & \partial_{xz} \\ \partial_{yx} & \partial_{yy} & \partial_{yz} \\ \partial_{zx} & \partial_{zy} & \partial_{zz} \end{bmatrix} W.$$

In a topocentric co-ordinate frame, where x points north, y east and z up, this matrix has the approximate form

$$\mathcal{M} \approx \begin{bmatrix} -0.15 & & \\ & -0.15 & \\ & & 0.3 \end{bmatrix} \text{mGal/m}, \tag{16.11}$$

in which

$$\partial_{zz} W = \frac{\partial^2}{\partial z^2} W = \frac{\partial}{\partial z} g_z = 0.3 \text{mGal/m}$$

is truly the “free-air” standard gravity gradient, the vertical gradient of normal gravity: if we write according to Newton (note that the vector \mathbf{g}



is pointing down whereas the z co-ordinate grows going up, that is the reason for the minus sign):

$$g_z = -\frac{GM}{(R+z)^2},$$

we obtain by taking the derivative

$$\frac{\partial}{\partial z}g_z = 2\frac{GM}{(R+z)^3}\frac{\partial(R+z)}{\partial z} = -\frac{2g_z}{(R+z)} \approx 3 \cdot 10^{-6} \text{ m/s}^2/\text{m} = 0.3 \text{ mGal/m}.$$

The quantities $\partial_{xx}W$ and $\partial_{yy}W$ again describe the curvatures of the equipotential surfaces in the x and y directions, in the following, geometrically intuitive way:

$$\partial_{xx}W = \frac{\partial^2}{\partial x^2}W = -\frac{g}{r_x}, \quad \partial_{yy}W = \frac{\partial^2}{\partial y^2}W = -\frac{g}{r_y},$$

in which r_x and r_y are the radii of curvature in the x and y directions. Substituting $r_x = r_y = R \approx 6378 \text{ km}$ (try!) gives

$$\partial_{xx}W = \partial_{yy}W = -1.5 \cdot 10^{-6} \text{ m/s}^2/\text{m} = -0.15 \text{ mGal/m}.$$

The already mentioned Hungarian researcher Baron Loránd Eötvös did many ingenious experiments (Eötvös, 1998) in order to measure the components of the gravity-gradient or Eötvös tensor with the *torsion balance* he built. The method continues to be in use in geophysical research, because the gravity gradient as an observable is very sensitive to local variations in the density of the Earth's crust.

In the general case, we can evaluate the Eötvös tensor by performing the partial differentiations above. We do so for a central force field:

$$W = \frac{GM}{r},$$

where $r = \sqrt{X^2 + Y^2 + Z^2}$ is the distance from the geocentre. The coordinates (X, Y, Z) are now geocentric. We obtain

$$\begin{aligned} \mathcal{M} &= \begin{bmatrix} \partial_X & \partial_Y & \partial_Z \end{bmatrix} \begin{bmatrix} \partial_X \\ \partial_Y \\ \partial_Z \end{bmatrix} \frac{GM}{r} = \\ &= \begin{bmatrix} \partial_X & \partial_Y & \partial_Z \end{bmatrix} \left(-\frac{GM}{r^3} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \right) = \\ &= \frac{GM}{r^5} \begin{bmatrix} 3X^2 - r^2 & 3XY & 3XZ \\ 3YX & 3Y^2 - r^2 & 3YZ \\ 3ZX & 3ZY & 3Z^2 - r^2 \end{bmatrix}. \end{aligned}$$



The Eötvös tensor is the matrix of partial derivatives of the gravity acceleration vector

$$\mathbf{g} = -\frac{GM}{r^3} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

with respect to place, as can be seen in the above equation:

$$\mathcal{M} = \begin{bmatrix} \partial_X & \partial_Y & \partial_Z \end{bmatrix} \left(-\frac{GM}{r^3} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \right) = \begin{bmatrix} \partial_X & \partial_Y & \partial_Z \end{bmatrix} \mathbf{g}.$$

In honour of Eötvös we use as the unit of gravity gradient the eötvös, symbol E:

$$1 \text{ E} = 10^{-9} \text{ m/s}^2/\text{m} = 10^{-4} \text{ mGal/m}.$$

The tensor 16.11 given above close to the Earth's surface is now, again in local topocentric (x, y, z) co-ordinates:

$$\mathcal{M} \approx \begin{bmatrix} -1500 & & \\ & -1500 & \\ & & 3000 \end{bmatrix} \text{ E}.$$

In every case

$$\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 W}{\partial z^2} = \partial_{xx} W + \partial_{yy} W + \partial_{zz} W = 0.$$

This condition, the *Laplace field equation*,¹⁸ applies more generally for a gravity potential in vacuum.¹⁹

The gravitational-gradient fields of the Sun and Moon are known on the Earth's surface as the *tidal force field*, which causes the phenomenon of the ocean tides, with a dominant period, due to the Earth's rotation, of twice a day.

vuoroveden
voimakenttä

¹⁸Pierre-Simon Laplace (1749–1827) was a French mathematician, physicist and astronomer, one of the 72 names on the Eiffel Tower, and surely the most gifted of them all ([Eiffel Tower, 72 names](#)).

¹⁹Here we have not considered, besides the atmosphere, the centrifugal force of the Earth's rotation, which causes the Laplace field equation to not be valid in a co-rotating co-ordinate frame, even in a vacuum. The correction term can, however, be precisely calculated. And for the *disturbing potential* $T = W - U$ the Laplace field equation is again exactly true.





Self-test questions

1. What are the differences between an absolute and a relative gravimeter?
2. By how many milligals is gravity less at aircraft height (10 km) than on the ground below? How much is the difference in percentage points of total gravity?
3. The same questions for the International Space Station, mean height 400 km.
4. In videos from the International Space Station, the people there appear to be weightless, so there is no gravity. Yet the station orbits at a height of some 400 km, where the Earth's gravitation is still strong. How is this possible?
5. What is the geopotential, and what is its relationship with gravity?
6. How is a normal gravity field defined?
7. In what situation would you use Bouguer anomalies rather than free-air anomalies?
8. What are deviations of the plumb line, and how are they determined?
9. What causes the deviations of the plumb line close to mountain ranges, and why are they smaller than theoretically expected?
10. What quantity does the unit eötvös (E) describe?



Exercise 16–1: Gravimetric geoid computation

Consider the computational arrangement of section 16.4, specifically calculation equation 16.6:

$$N(\phi, \lambda) = \frac{R}{4\pi\gamma} \left(\frac{\pi}{180} \right)^2 \sum_{i=1}^{360} \sum_{j=-89}^{+90} S(\psi(\phi, \lambda, \phi', \lambda')) \Delta g(\phi', \lambda') \cos \phi', \quad (16.6)$$

in which $\lambda' = i^\circ - 0^\circ.5$ and $\phi' = j^\circ - 0^\circ.5$.

1. How many *elementary operations* — a multiplication, possibly followed by an addition or subtraction — would you have to perform *in total* to calculate a global geoid model at resolution $5^\circ \times 5^\circ$ using gravity anomaly data given at $1^\circ \times 1^\circ$ resolution?

You may assume the Stokes function $S(\psi)$, the cosines of latitude ϕ' and the anomaly values to have been precomputed. A rough



estimate, say within ten percent, suffices for estimating the length in elementary operations of a computer run doing this computation.

2. Interchange the order of the two loops, and move $\cos \phi'$ to outside the inner loop. How does your answer change?



Space geodesy

17

1462 Zamenhof, provisional designation 1938 CA, is a carbonaceous Themistian asteroid from the outer regions of the asteroid belt, approximately 27 kilometers in diameter.

It was discovered on 6 February 1938, by Finnish astronomer Yrjö Väisälä at the Iso-Heikkilä Observatory in Finland. The asteroid was named after L. L. Zamenhof, the creator of Esperanto. It is a recognized Zamenhof–Esperanto object.

[Wikipedia, 1462 Zamenhof](#)



17.1 Earth rotation, orbital motion, sidereal time

The Earth orbits the Sun in 365.25 days. She also rotates around her own axis in a day. Figure 17.1 shows both the physical situation, the orbit of planet Earth around the Sun, and the apparent situation as seen from Earth, the annual path of the Sun along the zodiac or ecliptic.

eläinrata

During a civil day, 24 hours, the Earth turns once around her axis *with respect to the mean Sun*. With respect to the stars, however, the rotation period is a little shorter: when the Earth has, in a year, turned 365.25 times around her axis with respect to the Sun, she has, with respect to the stars, completed 366.25 rotations. Thus, the duration of one rotation is

siviili-
vuorokausi

$$\frac{365.25}{366.25} 24^{\text{h}} = 23^{\text{h}} 56^{\text{m}} 4^{\text{s}}.$$

The name of this period is a *sidereal day*.

tähtivuorokausi

The rotation of the Earth with respect to the stars — or, equivalently, the apparent rotation of the stars with respect to the Earth — is measured by an angle called *sidereal time*. It is calculated from clock time and

tähtiaika

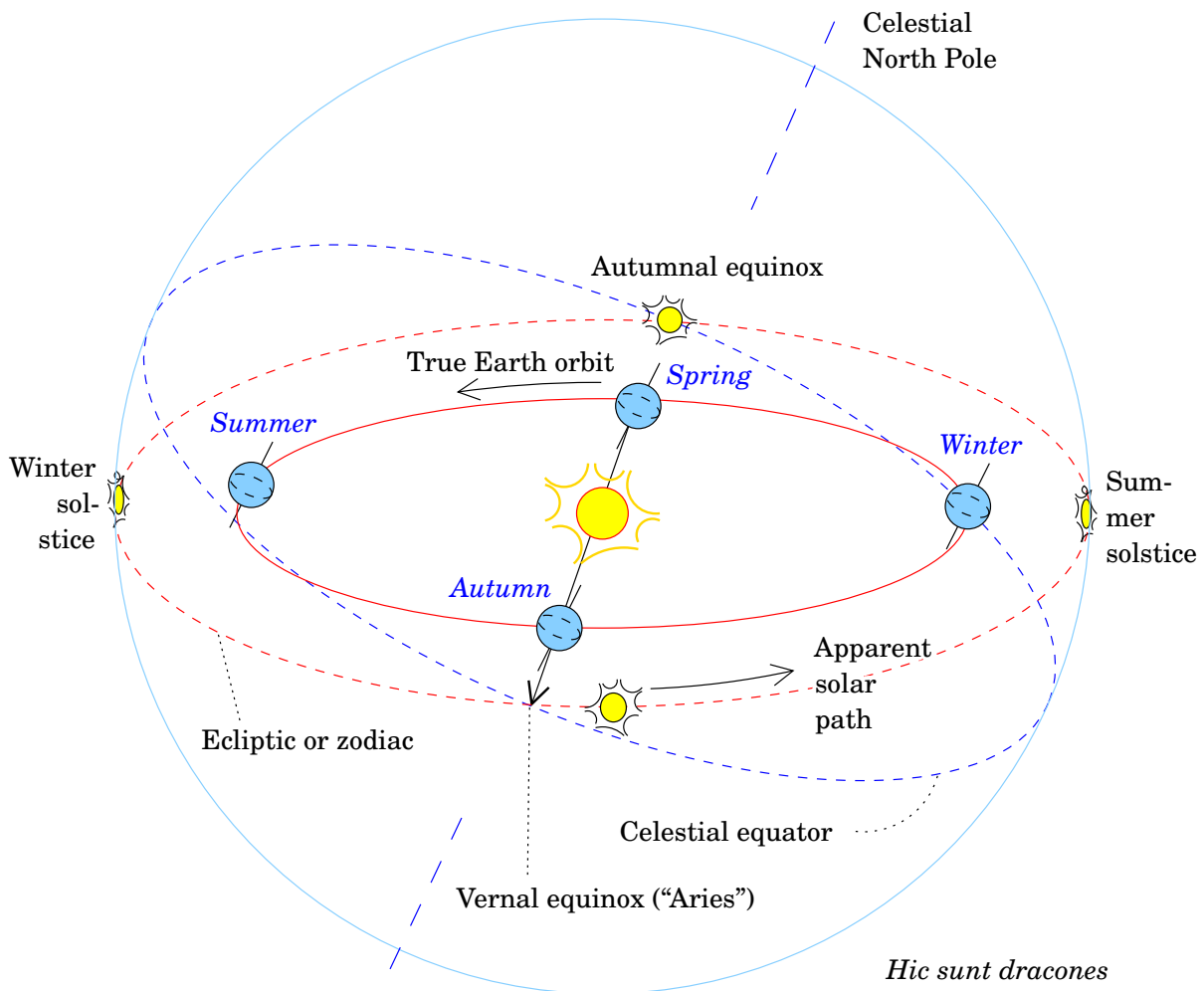


FIGURE 17.1. The orbit of the Earth around the Sun, and the apparent path of the Sun across the celestial sphere. Season names are boreal, referring to the Northern Hemisphere.

tähtikello calendar date using tables drafted for this purpose. In astronomical observatories, *sidereal clocks* are also used that show sidereal time, and run about $1/365.25$ part, or 0.27%, faster than ordinary clocks.

Due to the annual motion of the Earth, the constellations that are visible in the evening hours shift slowly forwards along with the season. Every season has its own distinctive constellations visible in the evening twilight: in winter, Orion, the stars Sirius and Procyon, in summer, the constellations of the Lyre, the Swan and the Eagle.

Seen from Earth, the Sun travels along a yearly path, the zodiac (“ring of beasts”, [Wikipedia, Zodiac, Name](#)) or *ecliptic*.¹ At the beginning of

¹The name “ecliptic” originates from the fact that this is where solar and lunar eclipses happen. Of course, because both types of eclipse require the Sun, Moon and Earth to be

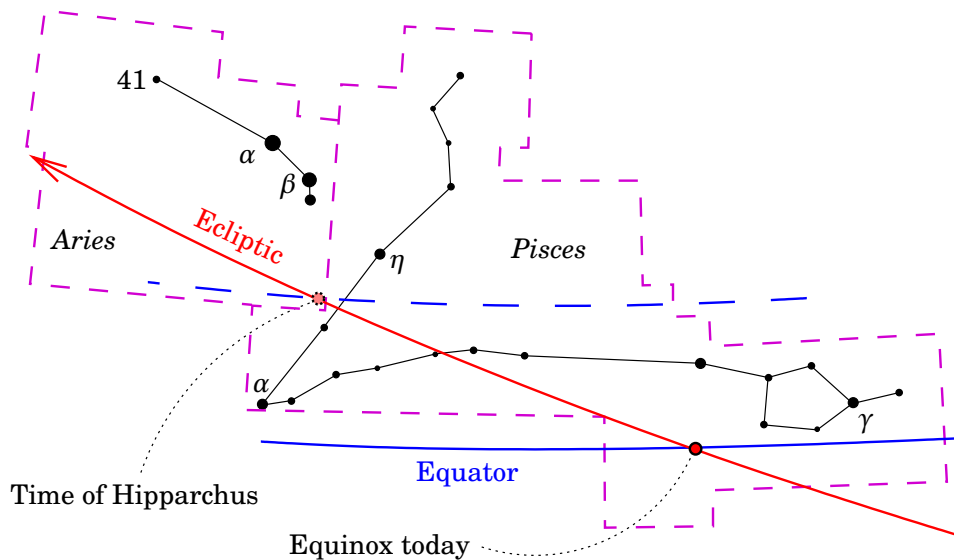


FIGURE 17.2. The vernal equinox and its movement, the *precession* of the equinoxes.



spring, the Sun moves from the Southern Hemisphere to the Northern one, at a point called the *vernal* or *spring equinox*. In the time of the ancient Greeks, this point was in the constellation of the Ram, hence the traditional name “First Point of Aries” — though, due to precession, it is nowadays located in the constellation of Pisces, the Fishes. In the same way, back then, the Sun was at the time of the summer solstice in the constellation of Cancer, the Crab, and at the time of the winter solstice in the constellation of Capricorn (a mythological goat-like creature), and the constellations gave their names to the *tropics*. Due to precession, today’s

kevät-
päiväntasaus

kääntöpiiri

The absolute orientation of the whole globe with respect to the stars is described by Greenwich sidereal time (**GAST**, Greenwich Apparent Sidereal Time). Local sidereal time (**LAST**, Local Apparent Sidereal Time) is obtained using astronomical longitude:

Greenwichin
näennäinen
tähtiäika

$$\text{LAST} = \text{GAST} + \Lambda,$$

in which Λ is the longitude of the site reckoned *east*, and of course converted into time units ($15^\circ \rightarrow 1^{\text{h}}$, $1^\circ \rightarrow 4^{\text{m}}$, etc.).

on the same straight line.



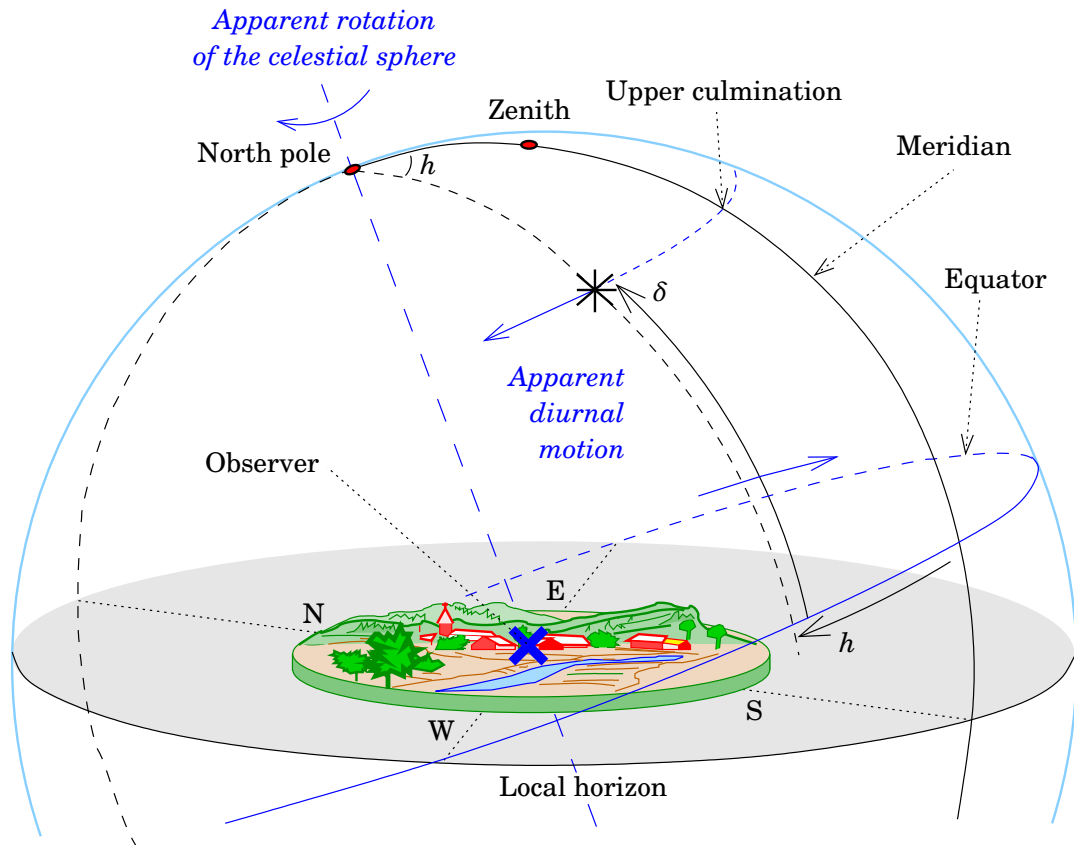


FIGURE 17.3. Hour angle h and declination δ on the celestial sphere.



17.2 Heavenly and Earthly co-ordinates

rektaskensio Celestial co-ordinates are *right ascension* α and *declination* δ . The right ascension is a longitude, measured however eastward from the vernal equinox, the right ascension of which is thus 0.

In the local sky, however, we use the co-ordinates *hour angle* h and declination δ . See figure 17.3. The pair (h, δ) can be directly computed from the azimuth and elevation, if the local astronomical latitude Φ is known.

korkeuskulma

During a sidereal day, a star moves apparently from east to west, and passes through the meridian plane two times: the *upper* and *lower culmination*. The lower culmination remains unseen if the declination of the star is too small compared to the local latitude. Of the southern constellations, even the upper culminations remain unseen by us!

The following relationship exists between sidereal time θ (= LAST), the hour angle h , and the right ascension α of a star:

$$h = \theta - \alpha.$$



If we write for Greenwich sidereal time (**GAST**), θ_0 and for the eastern astronomical longitude of the observation site, Λ , we obtain **Greenwichin tähtiaika**

$$h = \theta_0 + \Lambda - \alpha.$$

If, of the four quantities, three are given, the fourth can be calculated. We speak of

1. *time determination*, if the unknown is θ_0 . Back in time this was a service of astronomy to society, the maintenance of civil time. Today, with clocks so much more precise, it is about monitoring the Earth's rotation.
2. *longitude determination* of the site, if the unknown is Λ . This was critical for navigation at sea (**Sobel, 1998**).
3. determination of the *right ascension* of a star, if the unknown is α . This is how star catalogues are constructed.

In these cases, the measurement is generally done at the moment when $h = 0$: the *meridian transit*. Cases **2** and **3** require the use of a precise clock; case **1** requires knowledge of one's own longitude and the right ascension of the star used. **meridiaanin läpikulku**

The *meridian circle* is often used to precisely time the transit of a star through the meridian. A meridian circle is a telescope having only a horizontal axis, which has been built fixed into the east–west direction, in such a way that the sight axis of the telescope will always be in the local meridian plane. In the eyepiece, one observes how the star moves underneath the crosshair. The precise moment is recorded electrically together with time signals. **tähtäysakseli okulaari hiusviiva**

A graduated circle is attached to the axis, which allows the reading of the elevation or height angle η at the moment of transit. From this, the declination δ of the star can be calculated: between it, the latitude of the site Φ , and the elevation angle η exists a relationship (for the case of upper culmination on the Northern Hemisphere) **jakokehä**

$$\eta + \Delta\eta_{\text{refr}} = (90^\circ - \Phi) + \delta,$$

in which $\Delta\eta_{\text{refr}}$ is the correction for atmospheric refraction. The refraction correction must be made carefully based on local measurements of air pressure, temperature and humidity aimed at determining the index of refraction of the local air. **taitekerroin**

By combining methods **1–3** above one can build catalogues of both the places of stars (α, δ) and of observation stations (Φ, Λ), and at the same



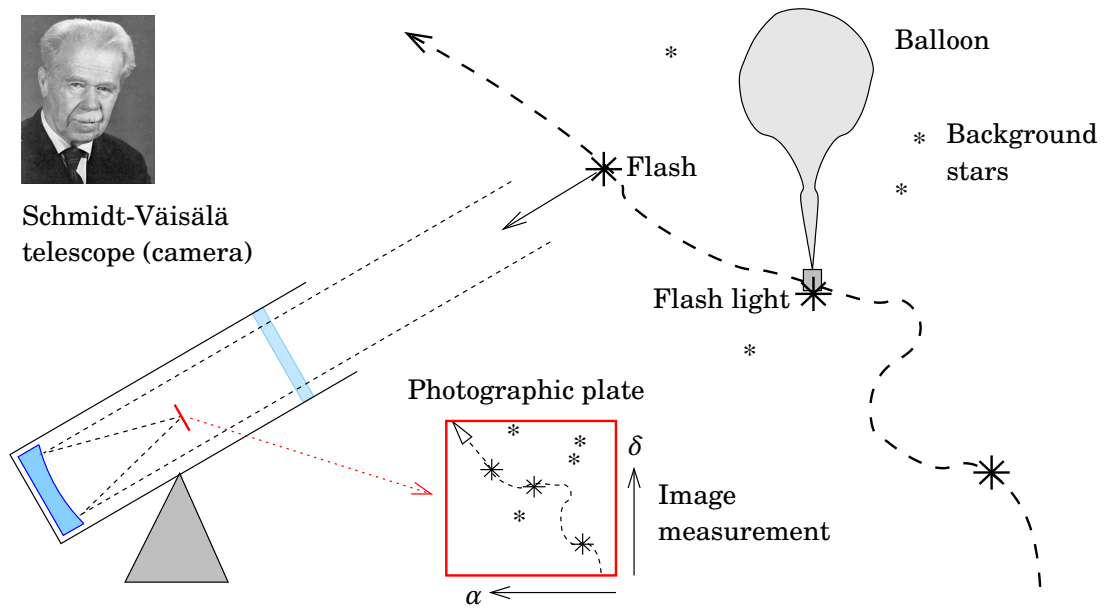


FIGURE 17.4. Yrjö Väisälä's stellar triangulation. Equipment used and principle of photography and direction determination. (a) Equipment used in the measurement.

time follow the progress of sidereal time — in other words, the rotation of the Earth — θ_0 . All this must be done in a consistent way.

[Sobel \(1998\)](#) is a fine book on the role of time in position determination at sea.

17.3 Väisälä's stellar triangulation

ilmatorjunta-
kranaatti

This method was invented by Yrjö Väisälä during the Second World War, while watching anti-aircraft shells exploding over Turku. In his 1946 article “*Maan toinen kuu*” (“*A second moon of the Earth*”, [Väisälä, 1946](#)) he described how targets high up in the sky could be photographed from different places on the Earth's surface, and thus a geodetic network built even connecting points between which there is no direct line of sight.

The method uses meteorological sounding balloons carrying powerful flashtubes. The flash train is photographed against the stellar background, the images are developed and the places of the flashes among the stars are measured and calculated. Thus the momentaneous direction vector between photography site and balloon is obtained, in astronomical co-ordinates ([Kakkuri, 1973](#)). This balloon method has been tried in a production setting in Finland and Hungary ([Czobor and Németh, 1981](#)). The side lengths of the network were several hundreds of kilometres.



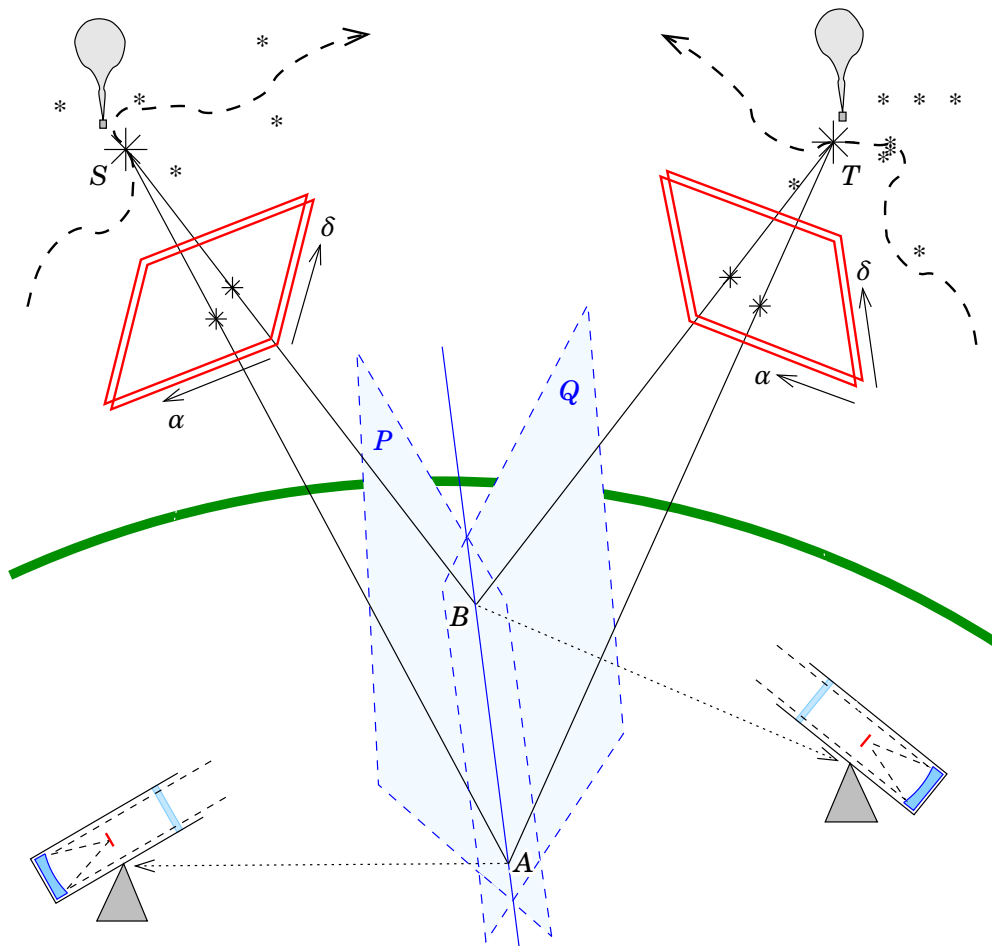


FIGURE 17.5. Yrjö Väisälä's stellar triangulation. Equipment used and principle of photography and direction determination. (b) Measurement geometry.

In short, from two observation sites A and B on the Earth's surface two high signals S and T visible in the sky are photographed against the stellar background. Powerful catadioptric Schmidt²-Väisälä (Wikipedia, Schmidt camera). A glass image plate covered by light-sensitive emulsion is placed into the image plane of every camera.

After development, the image plates³ are measured to extract, in an

²Bernhard Schmidt (1879–1935) was a telescope construction and design genius born in Estonia.

³Back then, glass plates were used instead of films for their geometric stability. The locations of the flashes are measured in plate co-ordinates, as are the locations of the fixed stars captured on the plate. Using the known places of the fixed stars, the places of the flashes are solved in celestial co-ordinates (α, δ) .



accurate measurement device, the locations, in “plate co-ordinates”, of both the signals — the flashes — and chosen background stars. Because the places of fixed stars in terms of celestial co-ordinates, right ascension (α) and declination (δ), are already known, we may derive transformation equations for all plates, which will also yield the places of the *signals* or flashes in the sky (α, δ), seen from both observation sites.

tähy

These “places” on the celestial sphere are in reality directions in three-dimensional space. From the directions are formed planes $P = SAB$ and $Q = TAB$. The intersection line of the planes is AB . Thus the direction vector from A to B has been obtained in three-dimensional space.

isoympyrä

In fact, the plane P is, on the celestial sphere, the same as the great circle through the image points of flash S taken in points A and B , and Q is the great circle through the images of flash T . The intersection point of the great circles is the *direction* in space of connecting line AB : stellar triangulation is direction measurement using heavenly auxiliary points.

tähtikolmio-
mittaus

verkkotasoitus

When a sufficient number of directions between ground stations has been collected, a *network adjustment* can be carried out. Transforming the original directions from the celestial co-ordinate frame to one co-rotating with the Earth, using a model of the Earth’s rotation (“Greenwich sidereal time”) requires that the flashes are accurately time-tagged.

Väisälä’s idea was, as may already be inferred from the name of his article, to use, instead of stratospheric balloons, artificial satellites orbiting the Earth. This method has also been tried: the active geodetic satellite [ANNA 1B](#) (“Army, Navy, [NASA](#), Air Force”) from 1962 was equipped with powerful flashtubes. On the other hand, the passive satellite [PAGEOS](#) (Passive Geodetic Earth Orbiting Satellite) from 1966 was a balloon, over 30 metres in diameter, made of very thin, aluminised Mylar[®] film, in a 4000 km high orbit. The satellite was clearly visible to the naked eye, until it disintegrated in the 1970s, undoubtedly due to the corrosive effect of the Sun’s ultraviolet rays.⁴

4

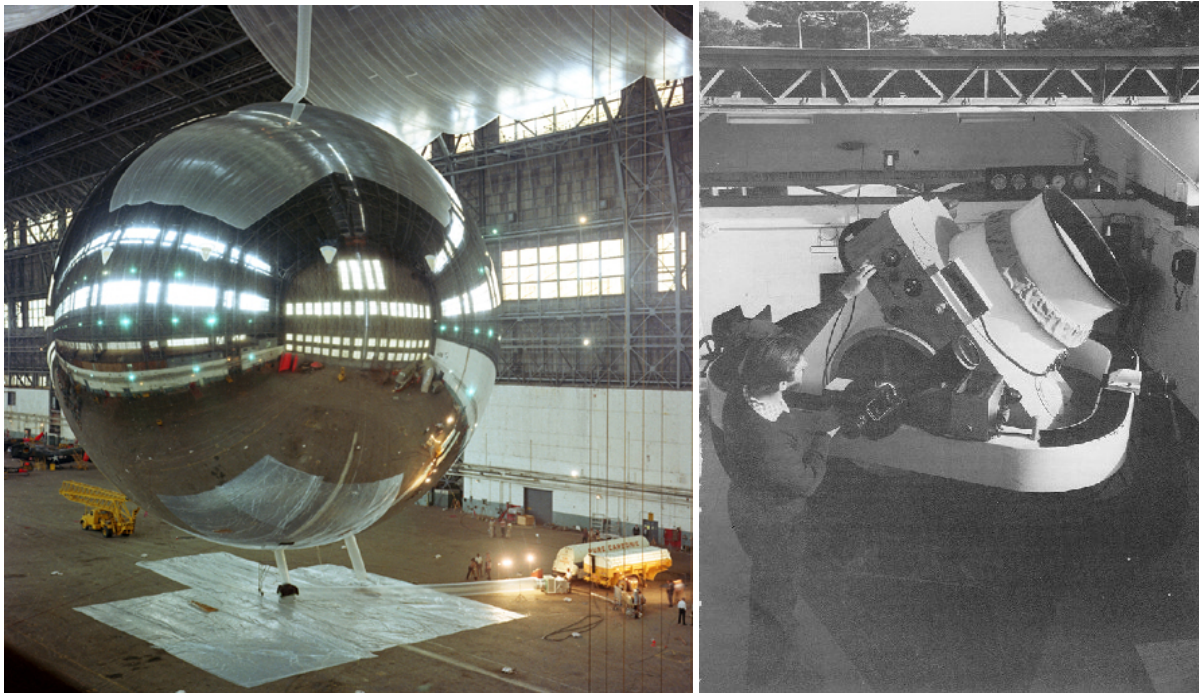
5

Using a global network of fifteen massive Baker–Nunn⁵ cameras, it

⁴DuPont write in the document *Mylar[®] polyester film — Safe Handling*: “Mylar[®] is not recommended for applications requiring prolonged exposure to direct sunlight due to degradation when exposed to ultraviolet rays [...]” ([DuPont Teijin Films](#)).

⁵The Baker–Nunn camera was a variant of the Schmidt camera, which was optimised for observing satellites. It used 55 mm broad Cinemascope motion-picture film, and it had a three-axis mount capable of fast tracking. The aperture was 50 cm, the weight of the whole instrument, 3.5 tonnes. The inventors were Harvard astronomer-optician





(a)

PAGEOS satellite, test inflation (NASA)

(b)

Baker–Nunn camera (NASA)



FIGURE 17.6. Satellite geodesy from the photographic archives of NASA.

became thus possible to build the first intercontinental geodetic triangulation networks, which however, with the advent of so much more precise methods (GNSS, VLBI, satellite laser-ranging), have honourably faded into scientific history.

kolmiomittaus-
verkkolaser-
etäisyysmittaus

17.4 Variations in the Earth's rotation

The rotation of the Earth varies in both speed and direction.

The momentaneous rotation axis of the Earth, or equivalently, the pole, moves with respect to the solid Earth's crust, a movement called *polar motion*. It includes two circular motions:

- An annual motion, period 365 days. This is a forced motion, mainly

James Baker (1914–2005) and mechanical engineer Joseph Nunn (1905–1968). The satellite tracking network consisted of fifteen stations. [Boller and Chivens, Baker Nunn](#), image archive, built twelve cameras for the Smithsonian Astrophysical Observatory and three for the US Air Force.

At least three Baker–Nunn cameras are or were spending their golden years as astronomical sky-survey cameras ([Carter et al., 1992](#); [Fors et al., 2013](#); [Mondal et al., 2009](#); see also [BakerNunn.org](#)).



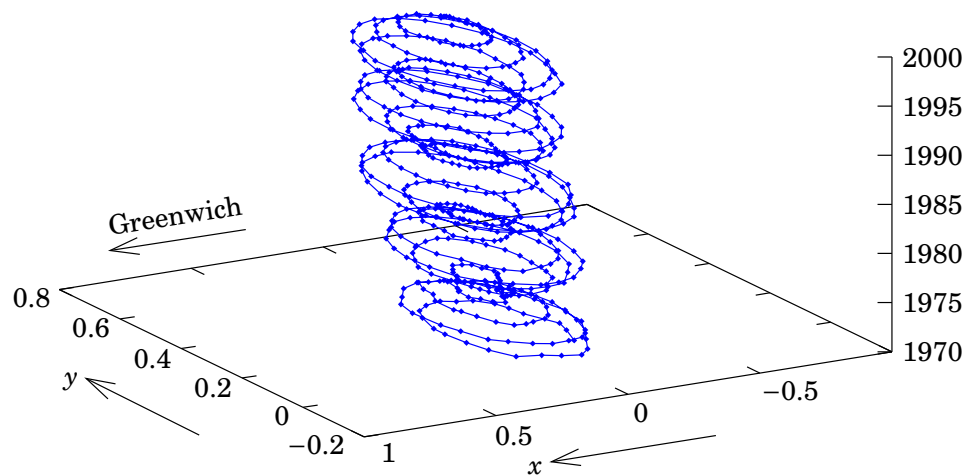


FIGURE 17.7. Polar motion for the period 1970–2000, unit second of arc. Source: IAG International Earth Rotation and Reference Systems Service IERS.



6 Chandlerin huojunta

caused by the atmosphere.

- The “Chandler⁶ wobble”, period about 435 days.
- In addition to these, there has been observed a slow drift of the pole, which is related to changes in mass distribution of the solid Earth, like the post-glacial land uplift.

The amplitude or radius of both circular motions is about $0''.1$ – $0''.2$, on the Earth’s surface about 3–6 m.

The Chandler wobble is theoretically understood: it is *free nutation*, already predicted for a flattened, rigid Earth by Leonhard Euler. The motion is damped, due to the Earth not being rigid but (elastically and plastically) deformable. The fact that the wobble continues requires an explanation. The explanation is that variations in the pressure exerted by oceans and atmosphere are its driving force (Gross, 2000).

vuorokauden pituus ilmakehän pyörähdys- momentti

The *rate* of the Earth’s rotation (length of day, LoD) also varies. This phenomenon is monitored in a similar way to polar motion. It is closely associated with variations in the *angular momentum* of the Earth’s atmosphere, a quantity that numerical weather models (NWP, numerical weather prediction) can calculate very well.

Monitoring of polar motion and length of day, historically and today:

- Measuring the variations in latitude by astronomical means. Use of this method started in 1899 (the International Latitude Service,

⁶Seth Carlo Chandler, Jr. (1846–1913) was an American astronomer.



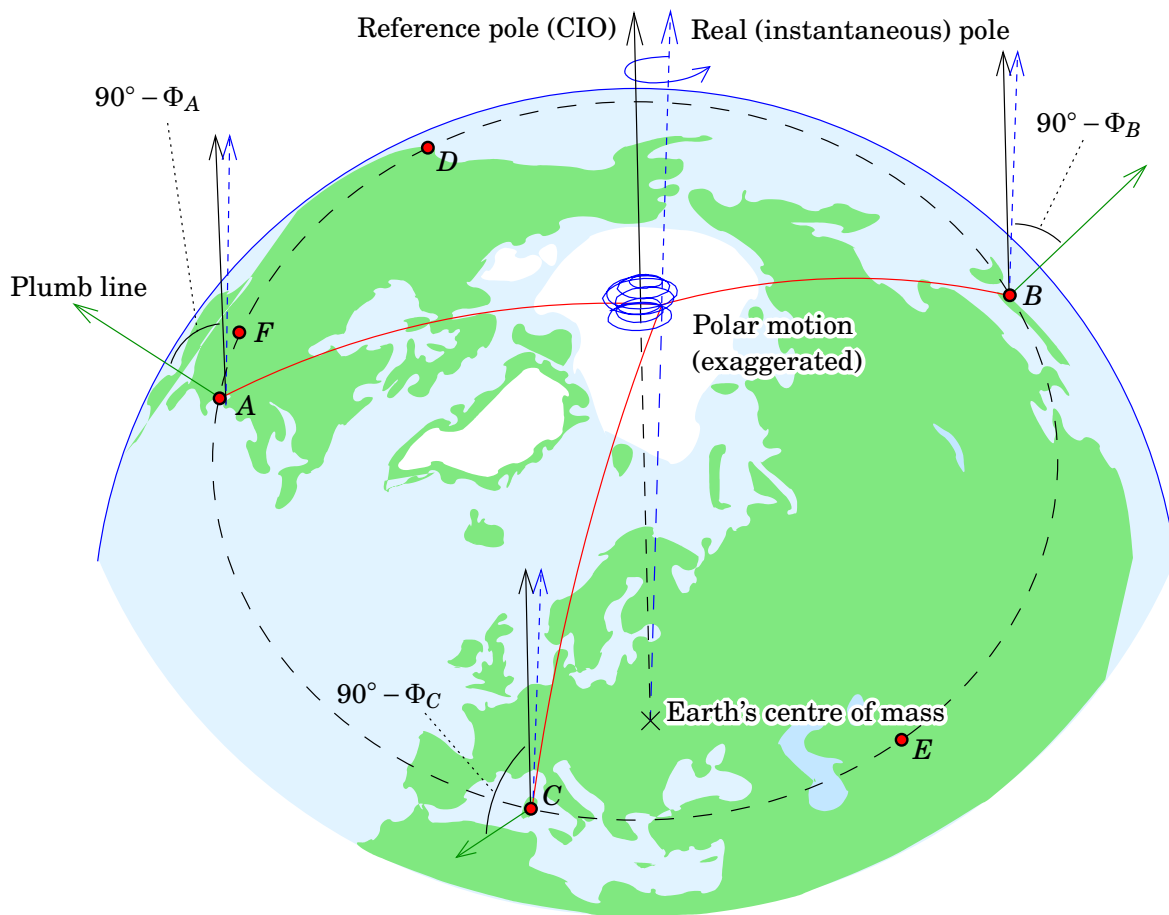


FIGURE 17.8. Polar motion causes variations in the latitudes of observation stations, with the help of which the phenomenon may be monitored.

from 1987 the International Polar Motion Service), using six “International Latitude Observatories” (Misuzawa, Japan; Charjui, Turkestan, later Kitab, Uzbekistan; Carloforte, Italy; Gaithersburg, Ukiah, Cincinnati, USA) which are located on different continents at the same latitude, $39^{\circ}08'$ around the globe. As the instrument used at the observatories was a *zenith tube*, this allowed them to all use the same selection of stars, enabling a uniform data analysis. Monitoring of the rate of rotation of the Earth has also been traditionally done by astronomical means. Civil time was originally defined by means of the Earth's rotation: the Earth herself was used as a clock. The instrument used was the meridian circle (for example in Greenwich, but also at the Helsinki observatory), with which the transit of a star through the meridian was observed. Back then, timekeeping was of vital importance to navigation at sea.

- Using positioning satellites, first the Transit system, nowadays



pitkäkanta-
interferometria

GNSS. The **GNSS** method is today more accurate than the astronomical method.

- With **VLBI**, very long baseline interferometry, which provides the vectors between observation stations in an inertial or celestial system as a function of time. From this, one may compute the momentaneous direction of the Earth's rotation axis in the same system, and even the momentaneous orientation of the whole Earth. Accuracy is even better than for the **GNSS** method, a fraction of a millisecond of arc.

The *origin* of calculation of the polar motion, **CIO**, *Conventional International Origin*, was initially the mean place of the pole over the years 1900–1905. The currently used origin of calculation is close to this.

Polar motion and length-of-day variations, together with nutation and precession, are called *Earth orientation parameters*, **EOP**. Their monitoring and publication is a task of the **IERS** (International Earth Rotation and Reference Systems Service). In precise geodetic work, the **EOP** must be taken into account! The correction information needed can be found on the Internet.

It was not until the 1930s that it was noticed that the rotation of the Earth is slightly irregular, and a more regular time scale was looked for. The first attempt was **ET**, *ephemeris time*, based on the orbital motion of the planets, especially the Moon. When sufficiently accurate atomic clocks appeared, they were taken into use, creating *atomic time* (**TAI**), also suited for demanding scientific use in which a time scale is needed that is strictly uniform.

Today's civil time is **UTC**, Universal Time Co-ordinated. One of its design objectives is to follow everywhere on Earth the cycle of daylight dominating daily life. Therefore it must follow with sufficient precision⁷ the variations in the rotation of the Earth. **UTC**, which is based on **TAI**, follows the variation in the Earth's rotation with an error of at most 0.9s. To this end, twice a year, at the end of December and at the end of June, **UTC** executes, if needed, a *leap second* or seconds (**Wikipedia**, **Leap second**). The difference **UTC** – **TAI**, an integer number of seconds, is tabulated in the almanacs.

karkaussekunti

⁷In recent years there has been a public discussion in the time field on whether the leap seconds are worth the trouble they are causing, when the time of the time-zone system tied to **UTC** agrees with local solar time anyway to no better than half an hour or so.



GPS time differs from both UTC and TAI. Like TAI, it is uniform and does not ever execute leap seconds. In 1980, when GPS timekeeping started, GPS time was identical with UTC. For this reason

$$\text{GPS} = \text{TAI} - 19\text{s}$$

and the difference GPS – UTC is a *varying* number of seconds, included in the GPS navigation message.



17.5 Precession and nutation of the Earth

The attraction of the Sun and Moon, together with the flattening of the Earth, causes a slow motion of the Earth's rotation axis in space. This conical motion is called *precession*. The word “precession” originally denoted the earlier and earlier occurrence of the *equinoxes*, the times of the year when day and night are equally long. This phenomenon, observed first by Hipparchus, brings on, over the centuries, the vernal and autumnal equinox earlier and earlier:⁸ it shifts the place of the Sun in the sky amidst the stars at the moment of equinox backwards along the zodiac.⁹ See figure 17.2.

päiväntasaus

8

9

In fact, the rotation axis of the Earth turns in some 25800 years around an axis that stands perpendicular on the plane of the Earth's orbit. This plane, the apparent plane of the Sun's orbit as seen from Earth, is also called the *zodiac* or *ecliptic*, see above.

Precession is *not* the same as polar motion. Polar motion is the motion of the Earth's rotation axis with respect to the *solid Earth*. Its magnitude is under a second of arc, on the Earth's surface a few metres.

Besides precession, the Earth's rotation axis also goes through a small periodic motion called *nutation*. Its main period is 18 years and it is caused by periodic changes in the orbit of the Moon which take effect through the Moon's attraction. The phenomenon can be precisely computed and is found in almanacs, just like the precession.

The precession makes the rotation axis of the Earth — and thus the place of the celestial pole among the stars — slowly describe a large

⁸Though *not* earlier on the civil calendar, based as that is on the rhythm of the seasons!

⁹Since then, the word has been taken into use to denote the turning motion of the axes of other spinning objects, like spinning tops, gyroscopes, and atomic nuclei in a magnetic field. The latter phenomenon, Larmor precession, is fundamental to magnetic resonance imaging (MRI).



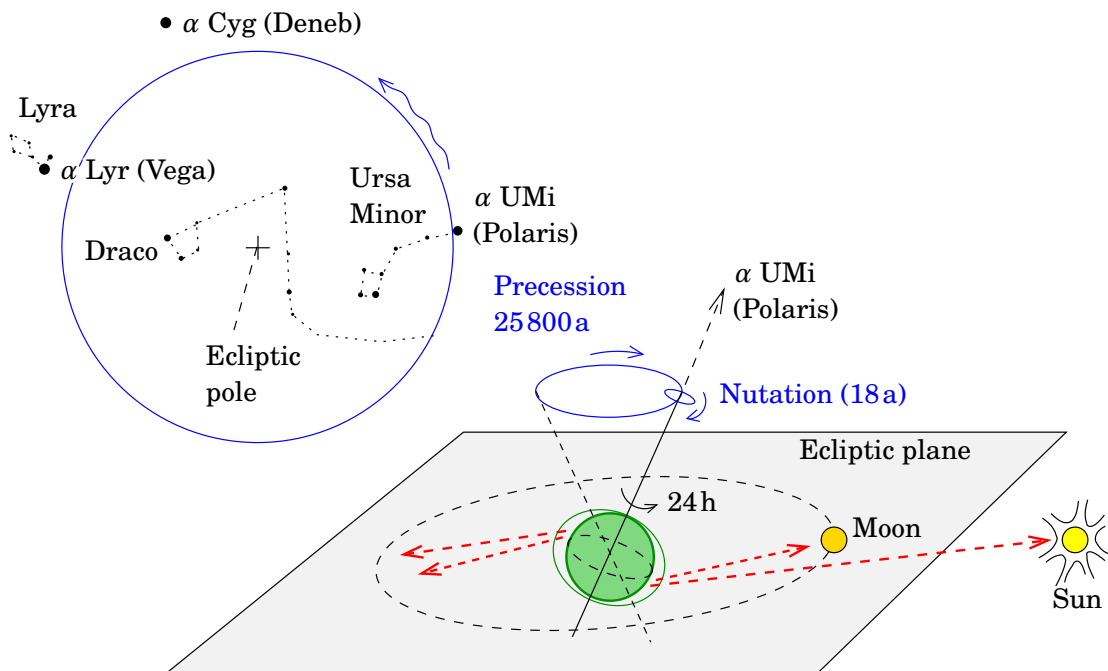


FIGURE 17.9. The Earth's precession. On the left is shown how, due to precession, the celestial pole describes a circle among the stars over a period of some 25 800 years.

circle. The nutation again causes small “wobbles” on this regular motion. Whereas today, the North Star (Polaris, α UMi) is close to the celestial pole, it will in the remote future be Vega (α Lyr), as it also was in the remote past.

Precession and nutation are, unlike polar motion, motions of the Earth's rotation axis with respect to the *stars*.

17.6 Space weather

The Sun is a star, which produces in its interior nuclear power by “burning” hydrogen to helium. In the centre of the Sun, the temperature is about 15 million kelvins. The thermal energy in the solar core travels very slowly to the surface through many gas layers, first as radiation,¹⁰ closer to the surface carried by convection currents, and finally again leaving the visible surface or *photosphere* as radiation, light, of which a vanishingly small fraction arrives also on Earth.

The outer layer of the Sun, of a thickness of about 27% of the solar radius, is in a permanent state of convective “bubbling”. Because of this,

¹⁰Thermal radiation, in the X-ray range due to the high temperatures.





FIGURE 17.10. The corona of the Sun during a total eclipse (image processed), Adler (2017). The patterns formed by the magnetic field lines in the corona and the solar wind escaping to space are visible.

energy is continuously being sent up to the Sun's highest layer, made up of thin plasma, the *corona*, heating this layer to millions of degrees.¹¹ 11 The precise form of this energy stream — acoustic, magnetic, . . . — is the subject of active investigation.

The corona, which is visible only during a total Solar eclipse, is so hot that it leaks continuously to space. This plasma or particle flow can be observed near the Earth as the *solar wind*.

Inside the convection layer of the Sun “lives” a complicated magnetic field, somewhat in the same way as in the liquid outer core of the Earth. The Sun does not rotate as a solid body: the higher-latitude zones lag clearly behind the equatorial zone. This, together with the magnetic field, acts as a complicated natural dynamo. Like the Earth's magnetic field, the Sun's magnetic field can also flip its direction, which happens periodically, approximately every 11 years.

The Sun's gas is an ionised *plasma*, which conducts electric currents almost like a superconductor. Therefore, the gas and the magnetic field

suprajohdin

¹¹For comparison, the temperature of the Sun's visual surface, or photosphere, is “only” approximately 5700K.

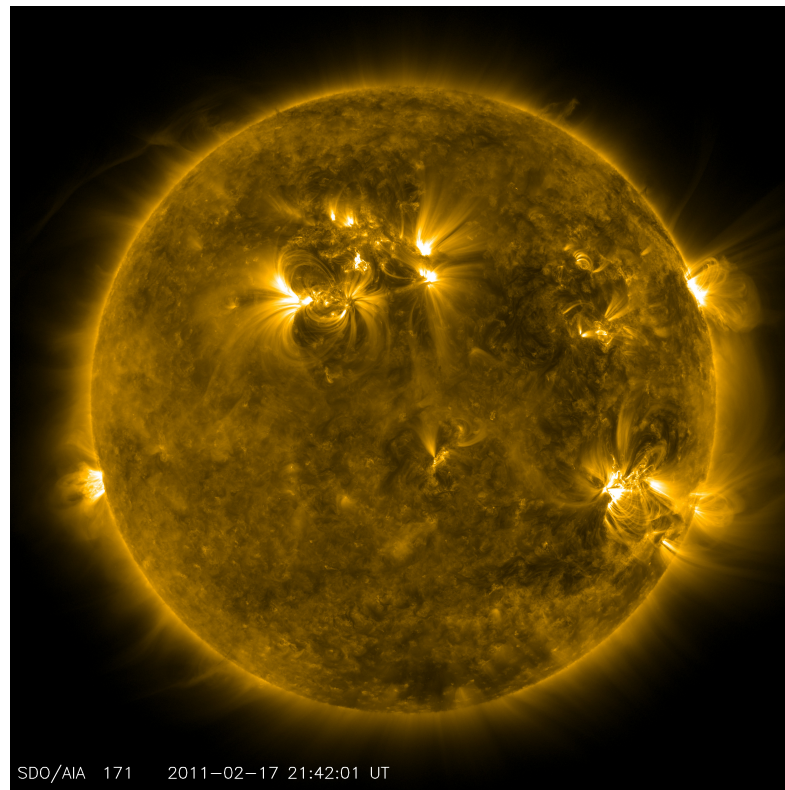


FIGURE 17.11. Sunspots and their magnetic field lines in UV light above the solar surface. Sunspots always occur in pairs, as a magnetic north and a magnetic south pole. [NASA Solar Dynamics Observatory, Atmospheric Imaging Assembly](#).



are entangled in each other (“*frozen-in*”) in an inseparable way. The subject is studied by the discipline of *magnetohydrodynamics* (MHD), to which a short introduction is given in appendix B.

auringonpilkkku

During times of an active Sun, *sunspots* are seen on its surface, areas of high magnetic field strength where the field of the convection layer breaks out to the surface. The spots are born and always show up in pairs, magnetic north and south poles. The magnetism in the spots inhibits the natural convection, and thus prevents energy from reaching the solar surface. In the centres of sunspots, temperatures are even a couple of thousand degrees lower than on the solar surface on average,¹² about 3000–4000 K.

12

The magnetic field of the sunspots extends into the space above the spots, and affects the motion of the plasma there. With special imaging

¹²However, even a single large sunspot transferred to the night sky would still outshine the full Moon!



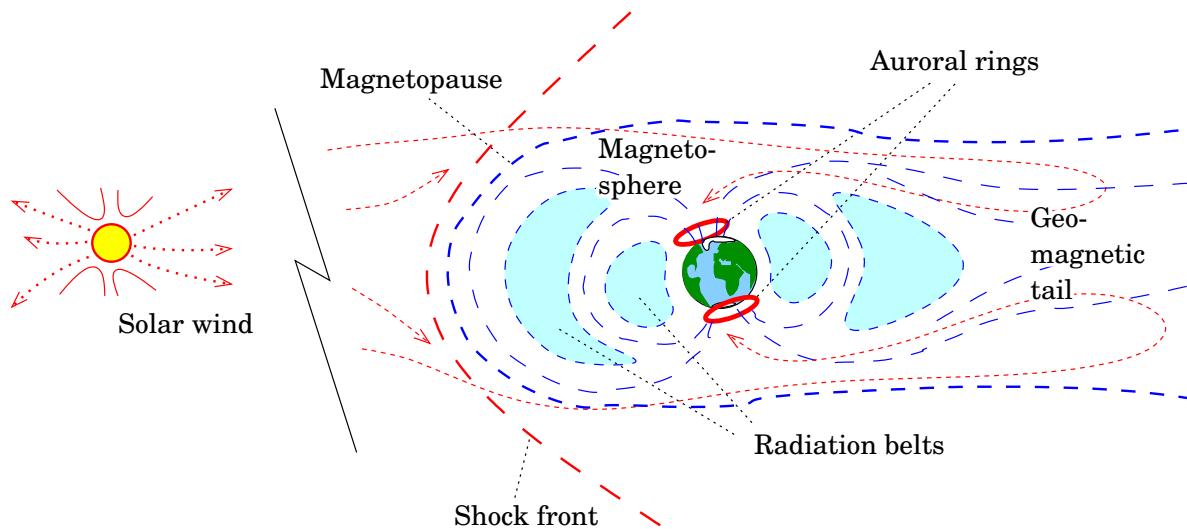


FIGURE 17.12. Space weather, the magnetosphere and aurorae.

equipment,¹³ the field's lines of force can be observed as bright swirls of gas. The topology of the field may suddenly change (*reconnection*), causing the energy released from the field to throw the hot plasma into space. During the solar flare, an excess of ultraviolet radiation and X-rays is generated, causing extra ionisation in the lowest layers of the Earth's ionosphere. This adds to the absorption of short-wave radio, disrupting radio traffic.

About a day later, the plasma itself arrives in the neighbourhood of the Earth, and its interaction with the Earth's own magnetic field and ionosphere causes aurorae. These, too, affect radio traffic and the quality of GPS observations. During an eruption, it may be impossible to collect useable GPS measurements. Generally during a solar maximum, the quality of GPS observations is poorer than during a quiet Sun.

The Earth's radiation belts also consist of hot plasma, fast, electrically charged particles, which the Earth's magnetic field keeps contained in a "magnetic bottle" (Wikipedia, Magnetosphere).

13

voimaviiva

revontuli

14

17.7 Satellite orbital motion

In the same way as the Earth orbits the Sun, the orbital motion of satellites around the Earth also follows the laws of Kepler.¹⁴ We already touched upon the matter regarding GPS, section 12.10. Kepler's laws are:

¹³Beautiful imagery and videos are found: [SOHO images](#); [SOHO MPEG movies](#).

¹⁴Johannes Kepler (1571–1630) was a German astronomer, mathematician and mystic.



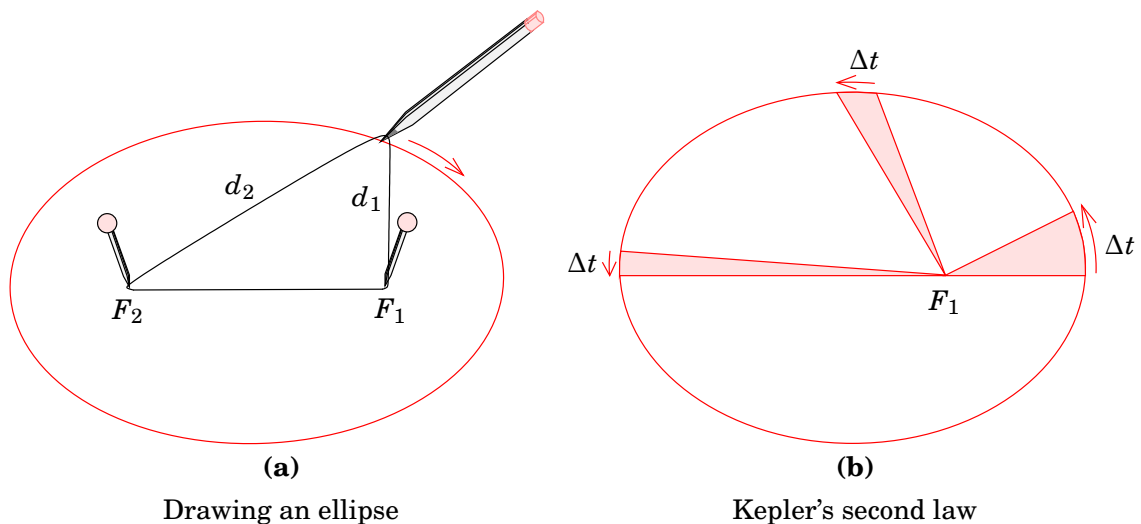


FIGURE 17.13. ← An ellipse is the set of points for which the sum of the distances $d_1 + d_2$ from two focal points F_1 and F_2 is constant. This property — a consequence of the ellipse being a conic section — is most easily proven using Dandelin spheres (Wikipedia, [Dandelin spheres](#)). It also means that if you place a lamp in one focus of an elliptical mirror, an image of the lamp will appear in the other focus. This is what gave the focal points their name. → According to Kepler's second law, the radius vector of a planet sweeps in the same time Δt over an always same-sized area.



- I** The satellite moves around the Earth in an *elliptic* orbit in a *plane*.
 polttopiste The centre of mass of the Earth is located in one of the focal points of the orbital ellipse.
- II** The radius vector between the satellite and the centre of mass of the Earth always sweeps in the same amount of time over the same surface area (*law of areas*).
 pintalaki
- III** The *squares* of the periods of different satellites stand in the same ratios as the *cubes* of the semi-major axes of their orbital ellipses. See below, equation 17.1.
 isoakselin puolikas
- rata-alkiot A satellite orbit is described by *six Kepler orbital elements*, figure 17.14. The orbital elements are described in more detail in appendix C. The angle θ_0 is Greenwich sidereal time, which describes the orientation of the Earth with respect to the stars.

Kepler's laws apply only approximately, especially for low orbits. The uneven distribution of the Earth's masses, and especially her flattening, ratahäiriö cause *orbit perturbations*. These are exploited for studying the internal



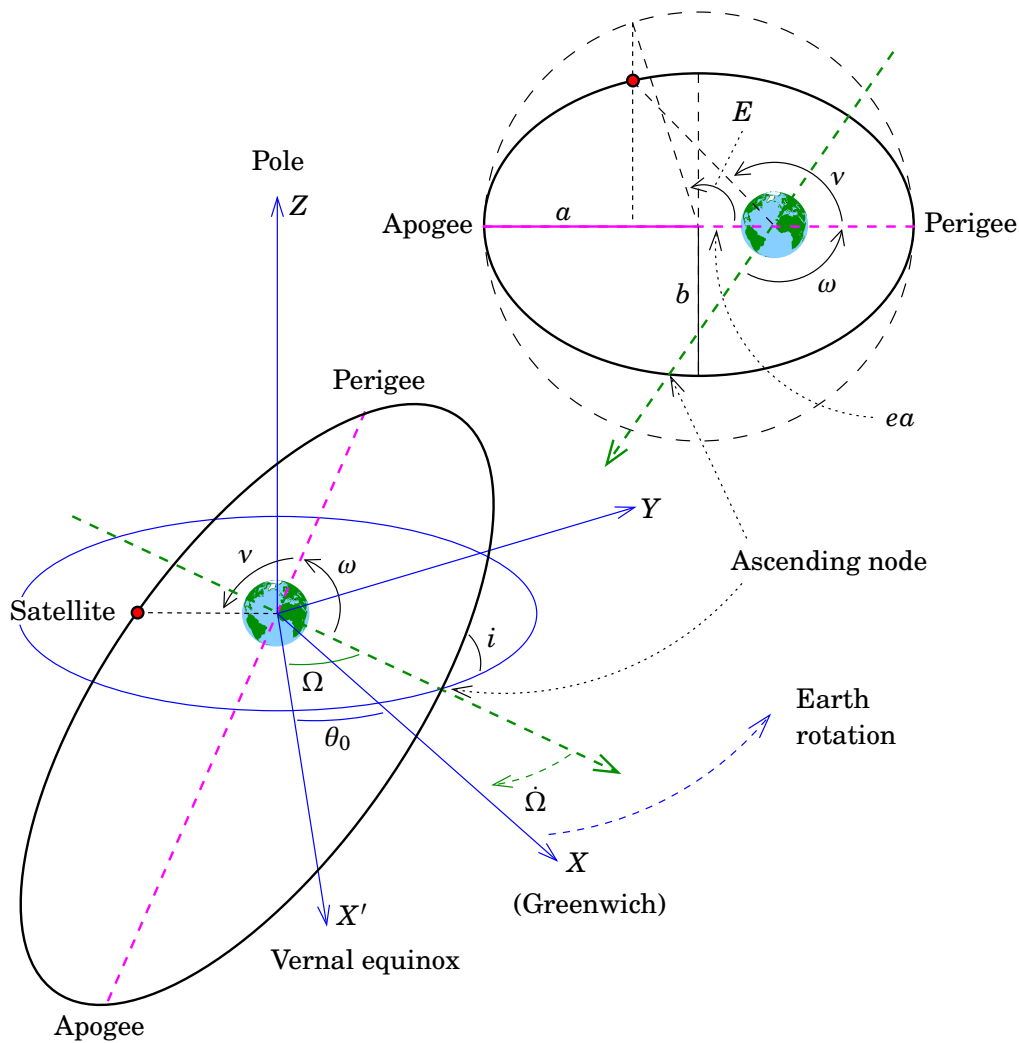


FIGURE 17.14. Kepler's orbital elements.

mass distribution of the Earth. This is how satellite geodesy has become an essential tool for studying the solid Earth.

17.8 Choosing a satellite orbit

The tilt of a satellite orbital plane relative to the equatorial plane, or its *inclination*, is an important parameter from the viewpoint of the intended use of the satellite. The inclination is in practice the same as the greatest possible northern or southern latitude over which the satellite can fly. If, for example, it is given that the inclination of some satellite is 55° , one may conclude that the satellite will never come down over Finland.¹⁵ 15

¹⁵This is assuming that the satellite is not capable of aerodynamic flight upon atmospheric entry.





TABLE 17.1. Kepler's third law for Earth satellites.

Height (km)	Period	Remark
0	84 ^m 29 ^s	Schuler period
400	92 ^m 34 ^s	
800	100 ^m 52 ^s	
20 183	11 ^h 58 ^m	GPS
35 785	23 ^h 56 ^m	Geostationary
376 603	27 ^d 07 ^h	Moon

From the viewpoint of low-flying weather and remote-sensing satellites, the importance of the inclination is that the area that a satellite can properly map lies approximately between these maximum latitudes.

This limitation does not apply for high-flying satellites. For example, geostationary satellites may perfectly well map the Nordic area. The imaging angle is not, however, good.

The choice of the height of a satellite orbit is made using Kepler's laws of orbital motion. Kepler's third law says:

$$GMP^2 = 4\pi^2 a^3, \quad (17.1)$$

in which $a = a_e + h$ is the semi-major axis of the satellite orbit, the mean distance from the centre of the Earth. The quantity h again is called the mean height of the satellite. P is the orbital period, the time taken to go around the Earth.

The perigee and apogee *heights* from the Earth's surface are formally calculated as follows:

$$h_P = (1 - e)a - a_e, \quad h_A = (1 + e)a - a_e \quad \implies h_A - h_P = 2ea.$$

In this, a_e is the equatorial radius of the Earth, according to the [GRS80](#) reference ellipsoid, 6378 137 m.



17.9 Satellite orbital precession, Sun-synchronous orbit

dynaaminen
litistyneisyys
pallofunktio-
kerroin

The figure of the Earth affects satellite orbital motion. For example, the quantity J_2 , the *dynamic flattening*, the value of which is $J_2 = 1082.6267 \cdot 10^{-6}$, is the largest of those many spherical-harmonic coefficients that together describe the figure of the Earth and that affect satellite motion. In the case of J_2 , one important effect is, that the satellite orbital plane turns at a certain rate around the rotation axis of the



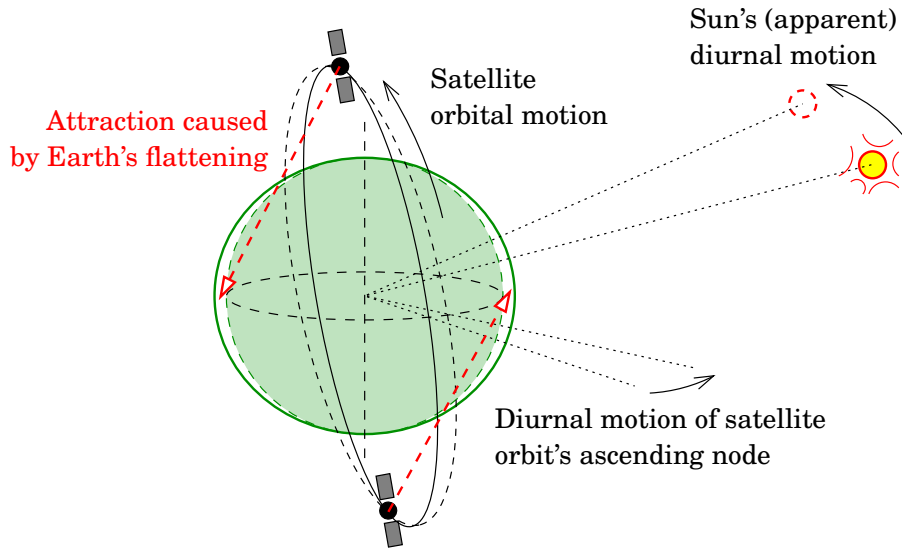


FIGURE 17.15. Sun-synchronous orbit.

Earth, a phenomenon called *nodal precession*. Because of this, if the satellite flies over the same place day after day, it will do so every day a number of minutes earlier. The rate of precession, for a circular orbit of radius a , is described by the equation

ratatason
prekessio

$$\dot{\Omega} = \frac{d\Omega}{dt} = -\frac{3}{2} \sqrt{\frac{GM}{a^3}} \left(\frac{a_e}{a}\right)^2 J_2 \cos i,$$

in which a_e is the equatorial radius of the Earth and i the inclination angle of the orbit with respect to the equator. Substituting numerical values yields

$$\frac{d\Omega}{dt} = -1.31894 \cdot 10^{18} [\text{m}^{3.5} \text{s}^{-1}] \cdot \frac{\cos i}{a^{3.5}}.$$

If, as a calculation example, we substitute into this as the satellite height

$$h = 800 \text{ km} \implies a = 6378137 \text{ m} + 800000 \text{ m} = 7178137 \text{ m},$$

we obtain

$$\frac{d\Omega}{dt} = -1.33102 \cdot 10^{-6} [\text{rads}^{-1}] \cdot \cos i = -6.589 [\text{day}^{-1}] \cdot \cos i. \quad (17.2)$$

For practical reasons (solar panels!) we often choose the satellite orbit so, that the orbital plane turns with the apparent annual motion of the Sun:

$$\frac{360^\circ}{365.25 \text{ days}} = 0.9856 [\text{day}^{-1}].$$

If the inclination of the orbital plane is chosen in the range 96° – 102° , depending on the height, the dynamic flattening J_2 of the Earth will cause just the suitable rotation of the orbital plane (“no-shadow orbit”, Sun-synchronous orbit).



Height (km)	Critical inclination
500	97° 4
750	98° 4
1000	99° 5
1500	102° 0



Self-test questions

1. A sidereal day is $23^{\text{h}}56^{\text{m}}4^{\text{s}}$. Why is it shorter than a civil day, 24h?
2. What is the ecliptic? How did it get its name?
3. What is the vernal equinox?
4. What mechanism causes the four seasons?
5. What is precession, and what is nutation? What causes them?
6. How did the Tropics of Cancer and of Capricorn get their names?
7. What geometrical quantity does Greenwich Apparent Sidereal Time (GAST) represent?
8. Describe the co-ordinates on the celestial sphere, right ascension and declination.
9. What is the hour angle of a celestial object? How is it related to its right ascension?
10. What do we mean by the upper and lower culmination of a star? Can they both always be observed (assuming clear skies)?
11. What is a meridian transit, and how is it observed?
12. Explain Väisälä's stellar triangulation.
13. Describe the components of polar motion and variations in length of day. How are they observed, historically and today?
14. How are precession and nutation different from polar motion?
15. What is a plasma?
16. What is the solar wind? Is it a sub- or supersonic flow? Why?
17. What is a *Carrington event*? Ask Google.
18. Could the wreck of the International Space Station (assuming it is not brought down in a controlled way) ever impact Finnish territory? Why / why not?
19. How does one engineer a Sun-synchronous orbit? Why is it useful?





Geodesy and geophysics

18

Satellites have revolutionized oceanography. This is not so much because of the instrument packages (remarkable as they are) but the ability to sample adequately, and to sample globally (two different things). Take the US–French altimetry mission TOPEX/POSEIDON that sampled the topography of the sea surface at about 7 km intervals to an astonishing precision of one-inch [...]. When you go over the list of accomplishments, you find that what really made the difference was the sampling. I consider this the most successful ocean experiment of all times. [...]

Walter Munk (2002)^a

^aWalter Heinrich Munk (1917–2019) was an influential American physical oceanographer.



18.1 Geodynamics

Geodynamics is the field of study within geophysics that studies the motions taking place in the Earth, like plate tectonics, post-glacial land uplift, and other motions of the Earth's crust, local, global, natural or human-caused. The study of the Earth's rotation is also normally included with geodynamics.

Of the observation techniques suitable for geodynamic research, we have already discussed satellite positioning. All GNSS methods — GPS, GLONASS, BeiDou and Galileo — are suitable for precise geodynamics measurements.

However, in scientific research we always want to use as many as possible, as independent as possible techniques for the study of the same

phenomenon, in as versatile a fashion as at all possible. Therefore we use also some very precise, but also very expensive, research methods in geodynamics research:

laser-
etäisyysmittaus

- satellite laser-ranging
- laser-ranging to the Moon

pitkäkanta-
interferometria

- VLBI, very long baseline interferometry with radio telescopes
- DORIS, Doppler Orbitography and Radiopositioning Integrated by Satellite (The IAG's International DORIS Service), a French satellite orbit determination system that serves geodetic and geodynamic research.

köyhdytetty
uraani
kuutioprisma

The satellite laser-ranging and lunar laser-ranging techniques are very similar. Laser satellites are heavy, massive spheres manufactured from high-density materials like bronze or depleted uranium, studded with reflective corner-cube prisms. The Apollo astronauts left several reflective panels covered in prisms on the Moon serving the same purpose. The Soviet vehicles deposited on the lunar surface in unmanned flights, Lunokhod 1 and 2, also carry reflective prisms.

The distance to satellites or to the Moon is determined by measuring the two-way travel time of a light pulse, multiplying it by the speed of light, and dividing by two. The influence of the atmosphere on signal propagation must be taken into account. Unlike GNSS, the laser-ranging technique measures *real distances*, without clock unknowns, not pseudoranges. Due to this, the global laser-station network is somewhat stronger geometrically than the global GNSS network. For this reason it gives a valuable boost to the global monitoring network's geometric strength.

¹ VLBI, very long baseline interferometry, uses the radio signals of remote radio sources, *quasars*. Quasars are so far away and compact¹ that in practice they are point sources. Therefore the wave form of the noise-like radio signal that they transmit will be precisely the same, no matter where on Earth it is received.

VLBI observations are carefully planned. Many radio telescopes around the globe participate in the campaigns. All participants execute an agreed programme, in which the same objects are observed simultane-

¹The dominant theory is that quasars are supermassive black holes in the cores of remote, thus young, galaxies. The radiation is generated when the hole sucks up matter from its surroundings. Our own Milky Way galaxy also has a black hole at its centre, Sagittarius A*, which however radiates much more weakly.



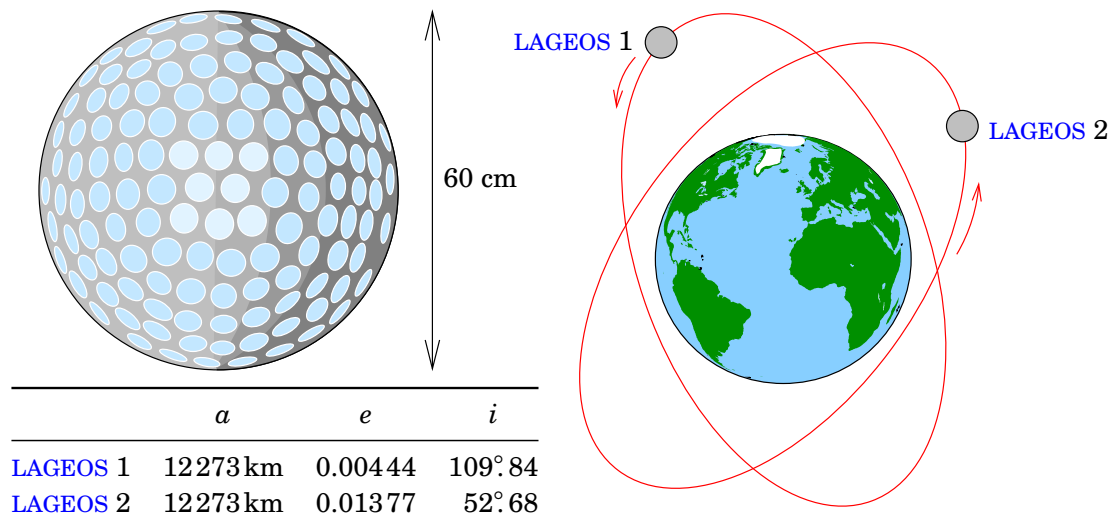


FIGURE 18.1. A LAGEOS satellite. The diameter of the satellite is 60 cm, it consists of an exterior aluminium sphere to which are attached 426 corner-cube prism reflectors (see subsection 7.4.3), and an interior sphere made of bronze. The mass of the satellite is about 400 kg. The LAGEOS 1 satellite also carries a map designed by Carl Sagan, describing the current locations of the continents, as a message to future, possibly alien, finders (Wikipedia, The LAGEOS time capsule). On the right, the orbits of the LAGEOS 1 and 2 satellites.



ously within a certain frequency band.

There are various solutions and generations of instrumentation for storing the signal. “Mark II” used common-or-garden VHS video tapes. The signal was converted by special equipment into a video signal and stored to tape. Today, the recording is often done digitally onto a hard-drive pack, and direct transfer over the Internet, using the UDP/IP protocol, is also becoming popular.

”mäyräkoira”

The recordings are read and correlated with each other, finding the time offset that makes the correspondence between the sequences apparent, just like in comparing the year rings of two pieces of wood (figure 7.7). This correlation is a compute-intensive process which earlier required specialised equipment. As a result of the correlation process, we obtain the difference in reception times Δt at the two radio telescopes, which is thus the observable of this technology. Its corresponding distance $c \Delta t$ may also be interpreted as the projection of the vector \mathbf{R}_{AB} connecting the radio telescopes A and B on the direction vector to the quasar, see figure 18.2.

vuosilusto



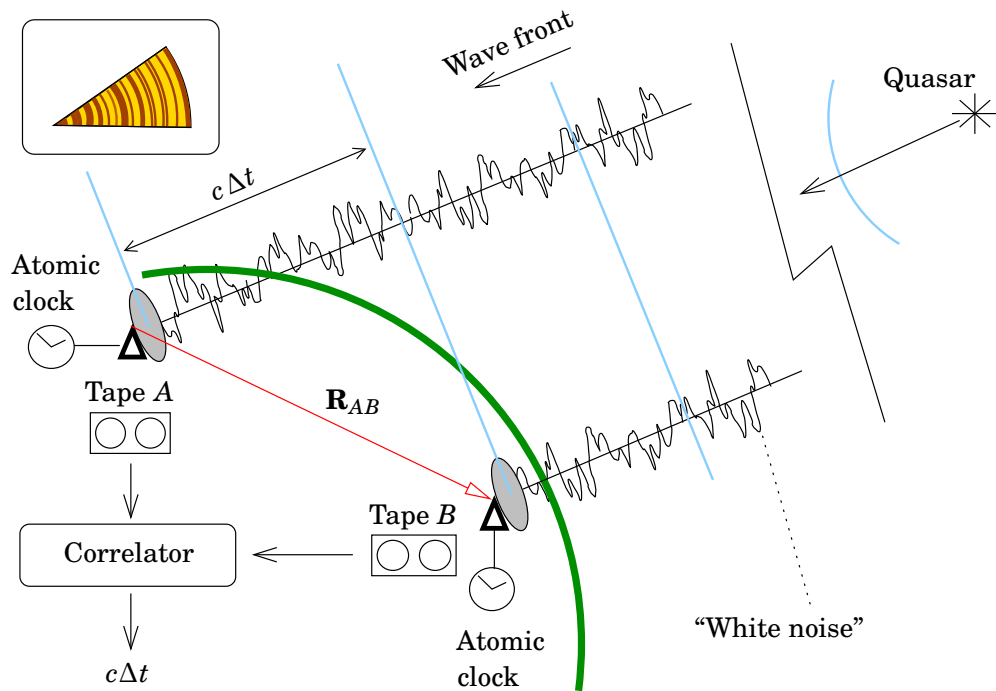


FIGURE 18.2. Principle of operation of very long baseline interferometry.

In order to work, like the famous Väisälä interferometry the method requires that the signal is random, that is, *white noise*: the coherence length has to be short. It is the same with GPS code measurement, for which the noise is generated artificially, as *pseudo-random code*.

By carrying out measurements to many different quasars, one may solve for the vector \mathbf{R}_{AB} between radio telescopes A and B in a geocentric — either inertial or celestial, tied to the stars, or co-rotating with the Earth — co-ordinate reference frame. Just like satellite laser observations, VLBI observations also strengthen the global geodynamic monitoring networks. Geodetic VLBI observations are also carried out in Finland, with the Metsähovi radio telescope, figure 18.3.

One essential difference between, on the one hand, VLBI observations, and on the other, both laser and GNSS observations, is that the centre of mass of the Earth does not come along in the VLBI observation equations. Quasars are at such huge distances, that any small change in the location of the centre of mass of the Earth would affect the observations by the radio telescopes at both ends of the VLBI vector in precisely the same way, and the effect on the end result would be zero.

On the other hand, the momentaneous *direction* of the *rotation axis* of the Earth is present in the measurement geometry, and VLBI has become, besides GNSS, a favourite means of monitoring the Earth's rotation, both





FIGURE 18.3. The radio telescope of the National Land Survey's Finnish Geospatial Research Institute (FGI) at the Metsähovi research station, used for geodetic very long baseline interferometric (VLBI) observations. The telescope, aperture 13.2m, belongs to the international VLBI Global Observing System (VGOS).



polar motion and variations in the length of the day.

vuorokauden
pituus

Close to the Metsähovi research station, three kilometres distant in Sjökulla, shielded by the landscape from the radio telescope, is a DORIS beacon containing an active radio transmitter. The French DORIS (Doppler Orbitography by Radiopositioning Integrated on Satellite) is an unusual system: *the observation stations are active* and the satellites passively collect their signals. An advantage of this solution is the centralisation of data collection. Around the world there are 60 DORIS stations, with stations on all major continental plates.



18.2 Plate tectonics

The German Alfred Wegener² proposed as early as in 1912 that the continents, fragments of the Earth's crust, were moving slowly over the soft interior of the Earth, like ice floes on the sea. As evidence he offered



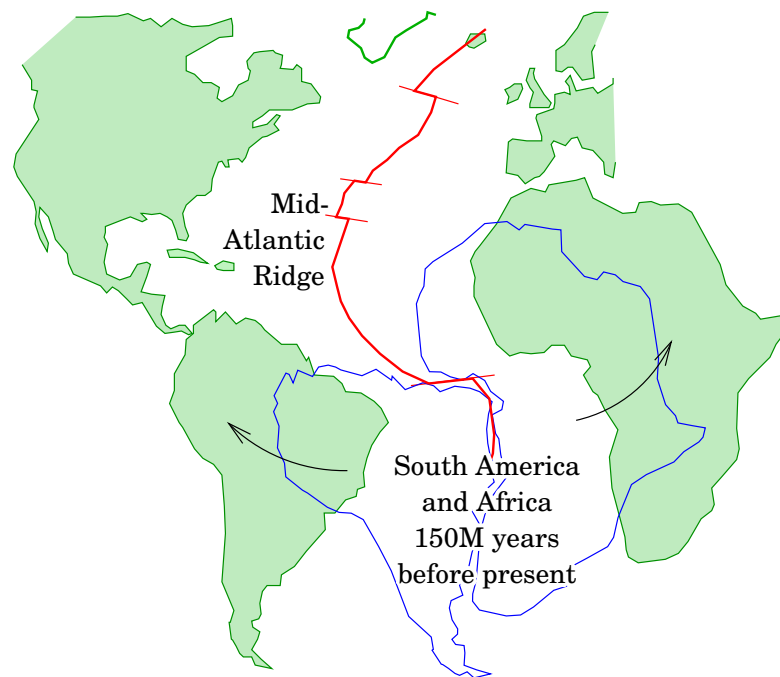


FIGURE 18.4. Alfred Wegener's continental drift theory and the Mid-Atlantic Ridge.



the shapes of the continental coast lines, which fit together remarkably well (figure 18.4) and rock types and fossils, which are often very similar in corresponding coastal locations.

poikittainen aaltoliike Nobody believed him at the time. It did not help matters that seismology showed the Earth's mantle to be hard as steel: transversal — sideways oscillating — waves, *S waves*, only travel in solids, and they travel very well through the Earth's mantle.

Not until the 1960s did the theory receive more support, especially based on research into the Earth's magnetic field. Back then, it was already possible to measure the local magnetic field from an aircraft, military technology developed for detecting submarines. The minute variations in the field were mapped, and *parallel magnetisation stripes* appeared everywhere on the maps, see figure 18.6. The stripes run in the direction of the Mid-Atlantic Ridge, or similar ridges in other oceans. Their pattern is the same everywhere, even if sometimes broader, sometimes narrower — like the tree rings from the same period in different trees, tableau 12.2.

²Alfred Lothar Wegener (1880–1930) was a German meteorologist, geophysicist and Greenland explorer, where he perished at the age of fifty and was buried into the continental ice sheet.



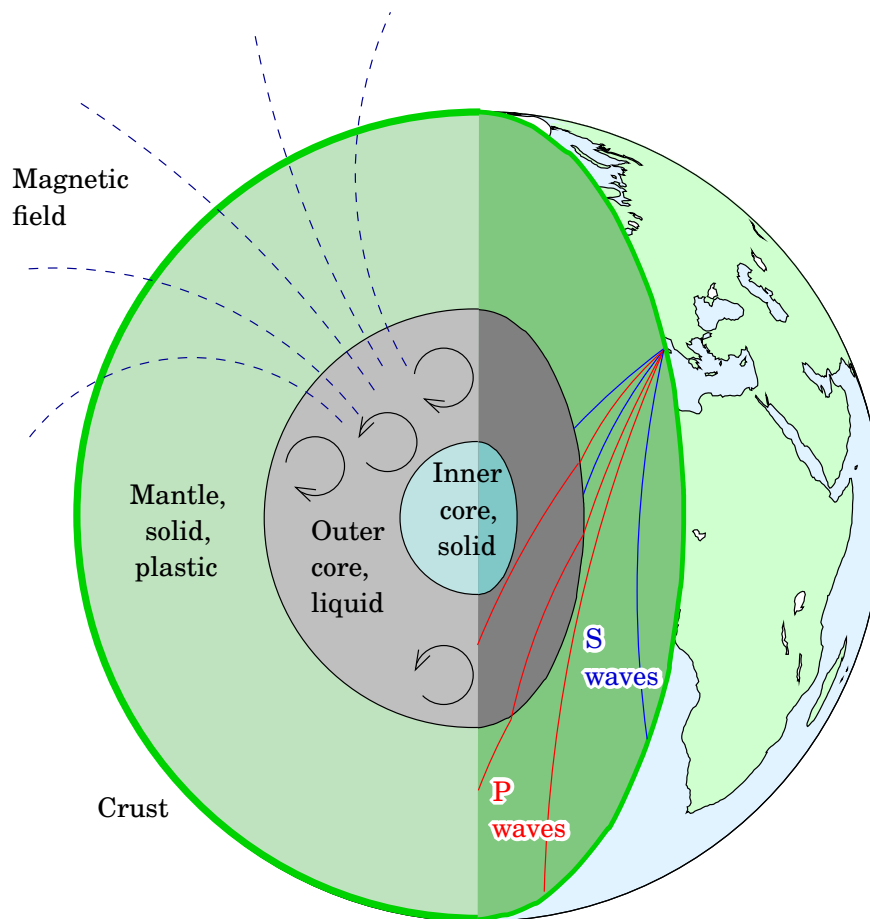


FIGURE 18.5. The internal structure of the Earth. Convection currents in the outer core generate the electric currents that maintain the Earth's magnetic field — the natural-dynamo theory. The core consists fairly certainly of an iron-nickel alloy with lighter-element impurities.



The theoretical explanation is that new Earth crust — sea floor — is being formed all the time at the Mid-Atlantic Ridge: hot, liquid magma rises up, cools down and solidifies, and the iron-ore particles in the magma turn themselves permanently along the direction of the magnetic field at that point in time.

merenpohja
Atlantin
keskiselänne

At this moment, the Earth's northern magnetic pole — in Canada — is physics-wise a south pole, *S*; the southern magnetic pole, near Tasmania, is an *N* type pole. These roles, however, swap at irregular intervals, of the order of a million years. These alternating directions are recorded into the sea floor in the same way that sound is recorded on the magnetic tape of a tape recorder.

The Earth's internal structure according to our best current knowledge



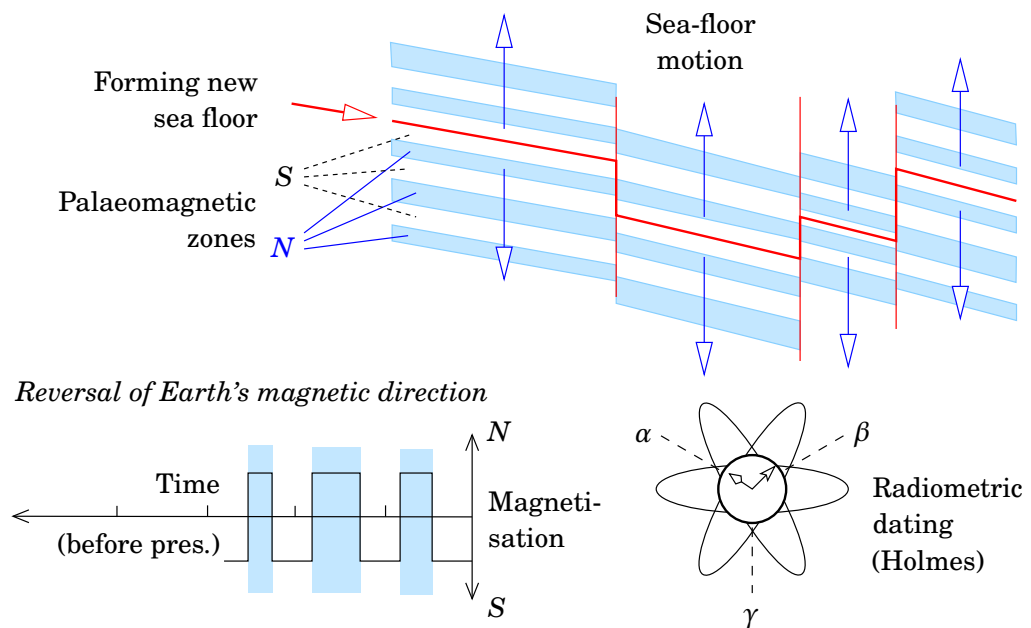


FIGURE 18.6. Palaeomagnetism and sea-floor spreading.

is presented in figure 18.5.

³ At the same time, the British geologist Arthur Holmes³ developed *radiometric dating*, which exploits the uniform decay rate of radioactive isotopes of various elements occurring in the Earth's crust as a geological clock (Lewis, 2000). Among the useful decay processes are potassium 40 → argon 40, half-life 1.25 billion years, uranium 238 → lead 206, 4.47 billion years, and uranium 235 → lead 207, 710 million years.

By measuring the concentrations of the decay product and the original isotope, one may infer how much time has elapsed since the solidification of the rock or mineral. The ages determined in this way for different places on the sea floor agree very well with the pattern of the magnetic stripes: the longer the distance from the central ridge, the older the sea floor.

We say that the *continental* plates move, but the plates include sea floor as well. We speak of the Eurasian plate, although it also contains the whole North Atlantic sea floor north of the Azores and east of the Mid-Atlantic Ridge. What is happening is a very slow *convective motion*, which carries the plates of the Earth's crust along with it. The rates of motion vary from a few centimetres per year to as much as ten centimetres. The precise form of the convective pattern is still unclear: does it take place

³Arthur Holmes [FRS FRSE](#) (1890–1965) was a British geologist, a pioneer of radiometric dating and our understanding of the mechanisms of plate tectonics.



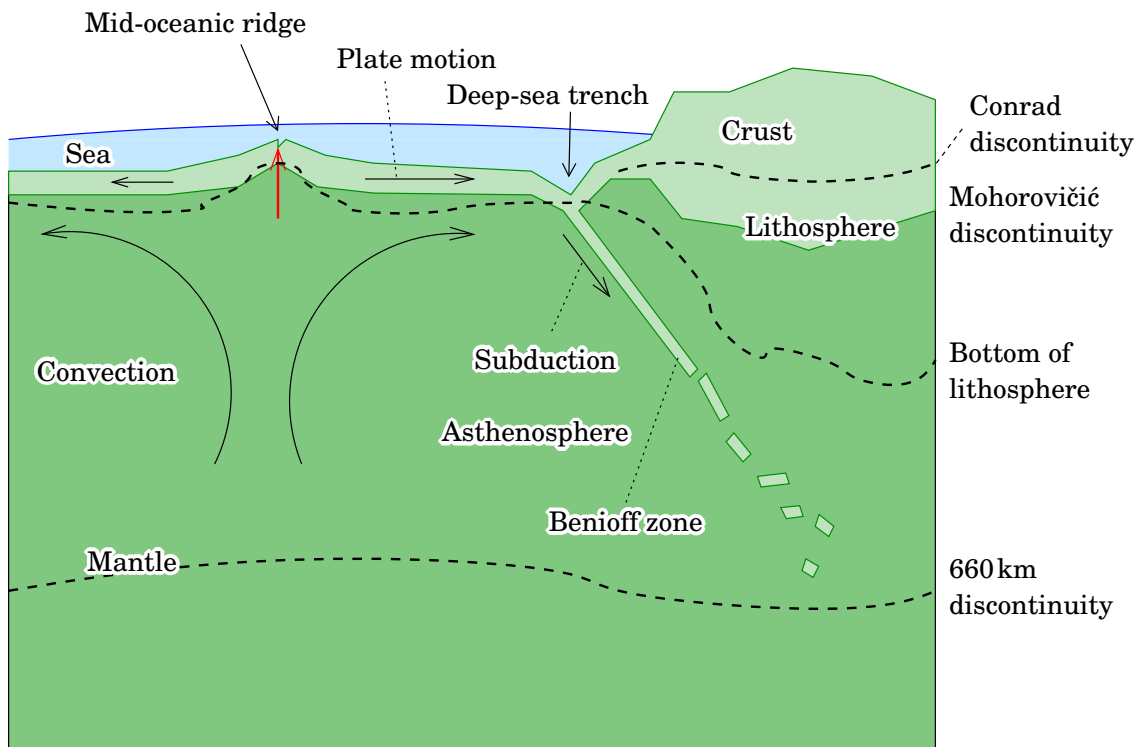


FIGURE 18.7. Global plate tectonics.

throughout the mantle, or in two layers?

The thermal energy that maintains the convection originates from two sources: radioactive decay, and “primordial heat”, the ongoing cooling down and stratification after the formation of the Earth. Both parts are believed to be of similar magnitude (Wikipedia, Earth’s internal heat budget). The heat produced by the Earth’s core is entirely primordial: the solid inner core continues to grow at the expense of the liquid outer core.

alkulämpö
kerrostuminen

The convection theory is presented in figure 18.7, and more details on the mechanism of continental motion in figure 18.8. At the Mid-Atlantic Ridge (and the Mid-Pacific, Mid-Indian etc. ridges) new sea-floor crust is being formed out of the magma rising from the mantle (however, as already said, the Earth’s mantle itself is *not liquid*). At the edges of the oceans again are found *deep-sea trenches*, under which the oceanic crust dives down into the mantle (*subduction*).

syvänmeren
hauta
alityöntö

In precise geodetic work, the continental-plate motion must be taken into account. When computing the velocity of motion, the **NUVEL** models may be used, which tabulate the motion of each plate as a *rotation* around a given *pole* at a given rotation rate. The **NUVEL-1A** model (DeMets et al., 1994) has been computed from geological research material. The results

kiertoliike



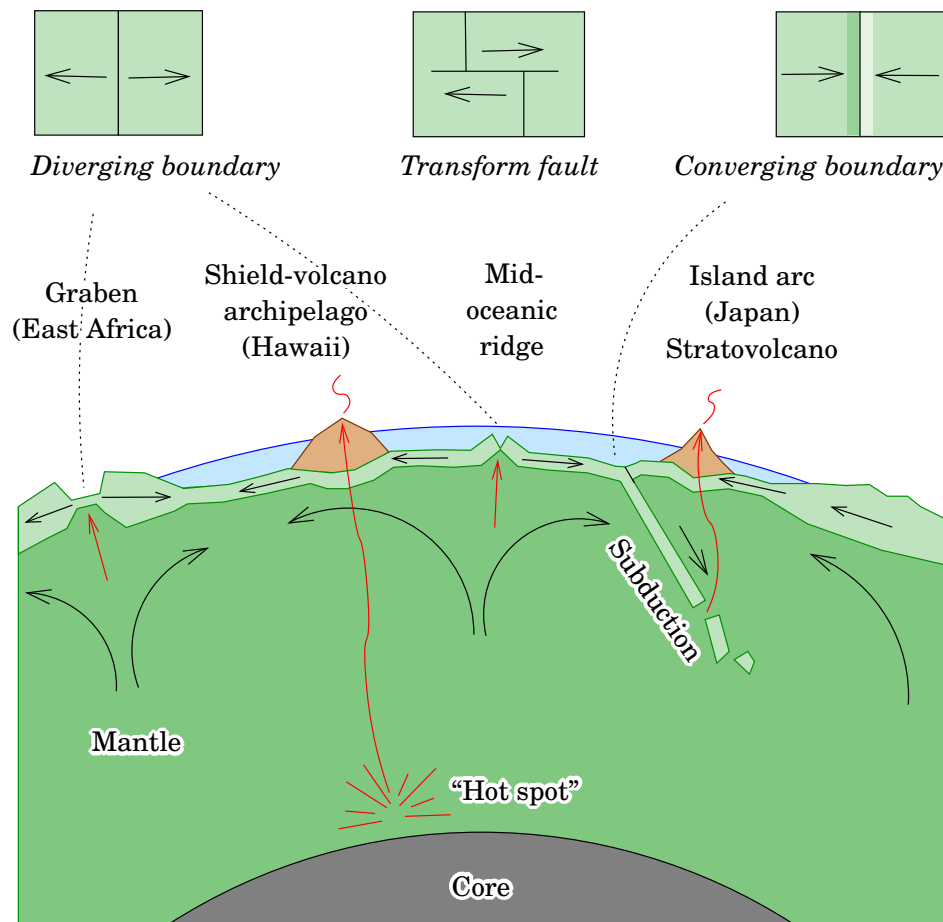


FIGURE 18.8. Mechanisms of plate tectonics, plate boundaries and volcanism. The thickness of the Earth's crust is exaggerated. The "deep hot spot" hypothesis is not generally accepted among geophysicists.



agree, within their measurement uncertainties, with the results obtained by [GPS](#) and other methods.

Over the years, improved plate-motion models have been computed, see [UNAVCO](#), [plate-motion calculator](#).



18.3 Glacial isostatic adjustment (GIA)

maan
vajoaminen
pullistuma

In many Arctic areas, the Earth's surface has been rising slowly after the end of the last ice age: *glacial isostatic adjustment* ([GIA](#)). Outside the Arctic area, the phenomenon takes the form of subsidence in the *periglacial bulge*. And the changes in sea level associated with the varying volume of land ice have a global impact.

When modelling the response of the solid Earth to varying ice load, we need to consider the following modes of response:



Elastic The object responds immediately, and returns to its original shape immediately when the load vanishes. For example, a spring, a tennis ball.

Plastic The object responds slowly and continuously. It does not return to its original shape when the load vanishes. For example, syrup, modelling clay. The resistance offered to deformation by a plastic substance is characterised by a property called *viscosity*.

Viscosity The more solid a substance, the higher the value of its *viscosity*. Unit: Pas (pascal second) or Ns/m^2 .

The physical character of post-glacial land uplift is one of *plastic* rebound. The phenomenon is studied because it offers a possibility to *determine viscosity values for the Earth's mantle*. The results obtained point to the following structure:

1. the "lid" or *lithosphere*, thickness 50–100 km, which has a high (in practice, infinite) viscosity and responds purely elastically⁴
2. under the lithosphere, the *asthenosphere*, a layer with a relatively low viscosity, 10^{20} – 10^{21} Pas, thickness several hundred km
3. under the asthenosphere, the lower mantle, which has a relatively high viscosity, order 10^{22} Pas or even higher.

The numerical values given above are highly uncertain and may well change as a result of ongoing research.

For comparison, the viscosity of running water at a temperature of 20° C is 0.001 Pas; that of liquid sodium — used as coolant in breeding reactors — 0.0007 Pas at its melting point 98° C. The viscosity of *pitch*, again, was determined in a famous, and still ongoing, experiment (Edgeworth et al., 1984) as $(2.3 \pm 0.5) \cdot 10^8$ Pas.

Example: the Fennoscandian land uplift

The land is rising in Finland, as it is in Sweden, Norway, Denmark, Scotland and Canada. All these vertical movements are caused by the melting, after the last ice age, of continental glaciers or *ice sheets*, some 11 000 years ago. In Fennoscandia a continental ice sheet, of a thickness of up to two-three kilometres, pressed the Earth's crust down. The phenomenon is called *isostatic adjustment*: although both the Earth's crust and the mantle consist of solid

⁴ ... until the loading becomes too heavy and the rock fractures: it is *frangible*.



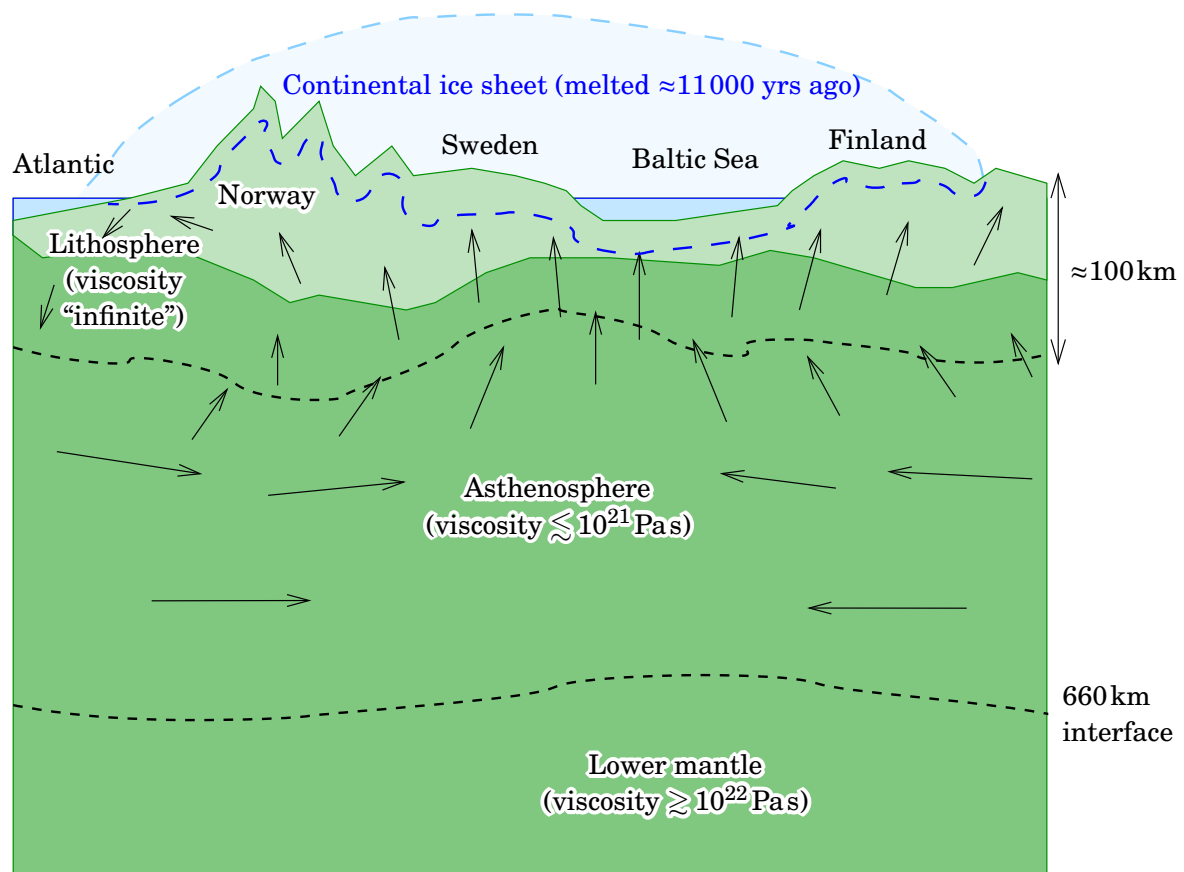


FIGURE 18.9. Post-glacial land uplift in Fennoscandia. The figure gives a vertical cross-section of the area; in fact, the phenomenon is three-dimensional. Topography and ice-sheet thickness are exaggerated.



matter, they nevertheless give way slowly and *plastically* under great loads.

The post-glacial land uplift in Fennoscandia is taking place in an area in the shape of an ellipse, the centre of which, the land-uplift maximum, is located at the narrow (“Kvarken”) of the Gulf of Bothnia, on the Swedish side, where the uplift amounts to over 9 mm/a . Going outwards from here, the rate diminishes in all directions. At the Finnish south-eastern border, only 3 mm/a remains. The zero land-uplift line runs through Denmark and Northern Poland, curving on the Lithuanian and Russian territories to the north-east. Outside, that is to the south of, the line, the land is slowly subsiding.

Merenkurkku

Fennoscandia is not the only area in the world where the Earth’s crust is rising as a result of the termination of the last ice age. We speak of **GIA**,



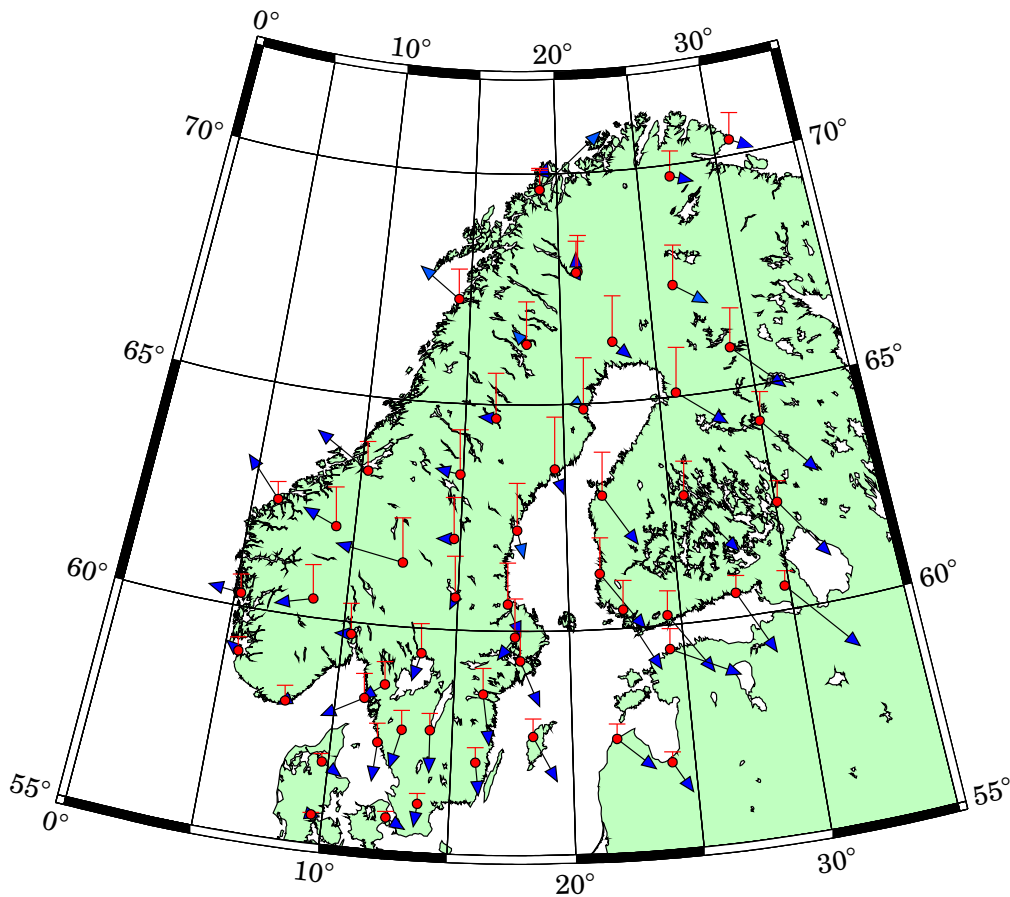


FIGURE 18.10. Horizontal and vertical motions in Fennoscandia as determined by the BIFROST project. The horizontal arrows and vertical bars are not on the same scale. Source Lidberg et al. (2009).

glacial isostatic adjustment. Other similar extended areas are:

- The Laurentide (Northern Canada) land uplift area. Much more extensive than the Fennoscandian one.
- The West Siberian land uplift area. Little studied (Ehlers et al., 2015).

When the Earth's crust rises, material flows slowly inwards in the asthenosphere towards the uplift centre, into the void under the crust that would otherwise be created. The land uplift is not only a vertical motion: horizontal motions also occur, as GNSS monitoring over extended periods has shown.

Techniques for studying and measuring the post-glacial land uplift include

- Regularly repeated (at intervals of decades) precise levelling.
- Monitoring sea level with respect to the Earth's crust by using



mareographs, also called tide gauges, at the coast.

- Monitoring changes in gravity, for example the Nordic gravity profiles (Mäkinen et al., 2010). The changes in gravity have *two causes*:
 - The point of measurement shifts with the land uplift away from the Earth's centre, and gravity diminishes.
 - Due to the flow of mass under the Earth's crust, the amount of mass under the point increases and gravity grows stronger.

The observed change in gravity is the net result of both effects.

- GNSS monitoring in three dimensions. This activity started in the 1990s, when sufficiently precise GNSS receivers and processing methods became available.



18.4 Local geodynamics



18.4.1 Anthropogenic motions of the Earth's crust

We humans, by our activities, often cause motions of the Earth's crust, *anthropogenic* motions. For example, building a reservoir dam and the filling of its reservoir causes an additional loading of the local Earth's crust, and may even cause tiny earthquakes.

In Venezuela, the pumping of oil in the Caracas region has caused a very noticeable local subsidence of the land. There are many other similar areas in the world. In the Netherlands, in Groningen, the pumping of natural gas has caused a subsidence, for which farmers are being financially compensated. For this reason, the motions are measured with geodetic precision, at regular intervals.

The pumping of drinking and irrigation water from porous layers — *aquifers* — can also cause subsidence of the land by as much as metres. This is a very common problem everywhere, but especially in developing countries (Wikipedia, Groundwater-related subsidence).



18.4.2 Natural motions of the Earth's crust

Local movements of the Earth's crust often happen close to geological faults, like the San Andreas fault in California, the boundary between two tectonic plates, where the plates move slowly with respect to each other. When the movement of the plates gets stuck, stresses inside the

Earth's crust build up, and may after years be released destructively in the form of an earthquake.

Various geodetic methods are used nowadays to monitor motions of the Earth's crust. At the mid-ocean ridges, like in Iceland, too, all kinds of geodetic measurement activity takes place in the service of geophysical research. valtameren
keskiselänne


The continental plates are not necessarily completely rigid and of one piece; rather, inside them there are also all kinds of faults along which tectonic movements have taken place — and surely still take place. In addition to observing microseismicity, one could try to observe these with geodetic monitoring techniques (Ahola, 2001).

Variations in sea level, like the tides, may be reflected in the level of the Earth's crust in coastal areas. One speaks of *ocean and atmospheric loading*. The motion caused by this loading may even be a couple of centimetres, but peters out quickly going in-land. Only in recent years has it been possible to measure this motion using GNSS, but the uncertainties are large. This tidal loading is also visible in long gravimetric monitoring time series. It is one way of studying the local elastic properties of the solid Earth.

The effect of the atmosphere, mostly variations in air pressure, should also be visible in this way. The phenomenon is, however, very weak and hard to observe with confidence. The problem with gravimetric techniques again is the difficulty of separating the effect of loading from the many other effects of the atmosphere on the measurement device and its surroundings.

18.5 Deformation monitoring

Both traditional (total-station measurement, precise levelling) and modern techniques like GNSS monitoring can be used to monitor local deformations in the Earth's crust. takymetri

In earthquake-prone areas (for example Japan, California) GNSS monitoring networks made up of many hundreds of continuously operating receivers have been built. In Japan, the Sendai earthquake of 2011 was recorded by 1200 stations covering the country, as shown in the video. 

However, more traditional techniques, like monitoring with automatic total stations, are also used. Collecting and processing the materials is done in real time.

Detecting deformations from the observational material is a similar task as detecting gross errors: the same kind of *statistical testing* (section 15.3) may be used, with a suitably chosen alternative hypothesis H_a .

The SAR (synthetic-aperture radar) technique from satellites (Field-synteettinen aukko ing, 2017) is also used for deformation monitoring. “Synthetic aperture” means that the radar images taken during different satellite overpasses are computationally combined in such a way that a “virtual objective” the size of the distance between overpasses is created. The method is interferometric (InSAR). After suitable processing, the deformations show up in the images as interference fringes.



18.6 Studying the Earth’s gravity field from orbit

During the decade 2000–2010, three satellite missions were launched to study the fine structure of the Earth’s gravity field or geopotential; in other words, to draft a global geoid map.



18.6.1 CHAMP

CHAMP (Challenging Minisatellite Payload for Geophysical Research and Applications, CHAMP Mission) was launched from the Plesetsk cosmodrome in Russia on 15 July 2000. The orbital height of CHAMP was initially only 450 km, which dropped to 350 km during the flight (for comparison, the orbital height of the GPS satellites is 20 000 km).

CHAMP carried a GPS receiver for precise orbit determination of the satellite. From GPS data, one may compute the precise location $\mathbf{x}(t)$ of the satellite as a function of time. From this, one may calculate the geometric acceleration $\mathbf{a}(t)$ by differentiation:

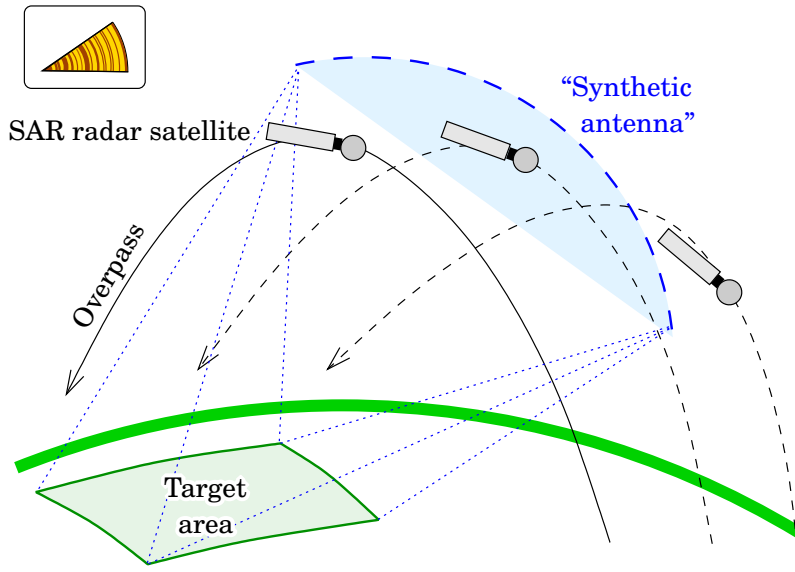
$$\mathbf{a}(t) = \frac{d^2}{dt^2} \mathbf{x}(t).$$

kiihtyvyyss-
mittari

The satellite also carried an accelerometer, which served to measure the satellite accelerations caused by the atmosphere’s aerodynamic forces acting on the satellite. After eliminating these, what is left is only the accelerations caused by the Earth’s gravitational field, from which a precise geopotential or *geoid model* may be computed.

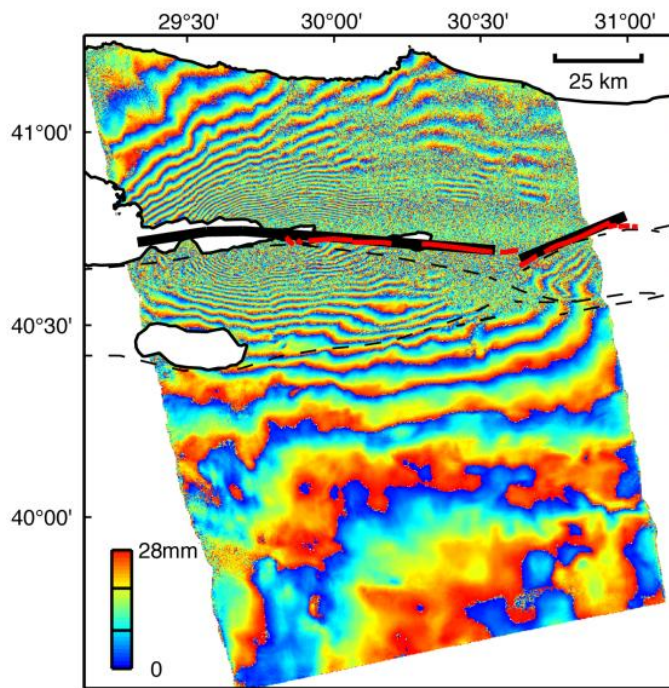
The data collected by CHAMP has been used to compute global geopotential models. The geographical resolution of the models has been modest, of an order of 1000 km. The satellite returned into the atmosphere on 19 September 2010 and burned up after 58 277 orbits.





(a)

The geometry of SAR imaging



(b)

Example of an InSAR image

FIGURE 18.11. InSAR image. Earthquake of 17 August 1999 in Izmit, Turkey. The deformation interval between interference fringes of the same colour is some 7 cm of horizontal motion. The thin red lines are faults. NASA / JPL-Caltech.



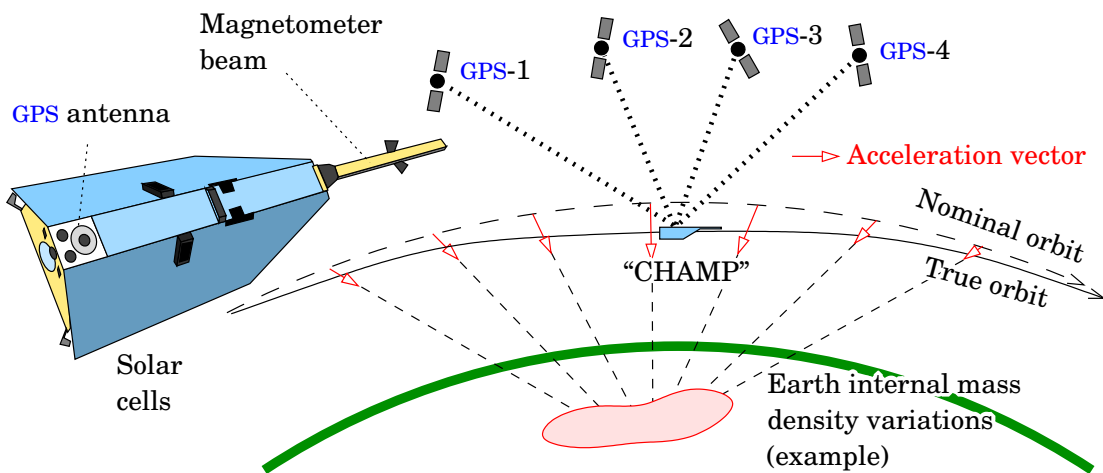


FIGURE 18.12. Determining the Earth's gravitational field by tracking the orbit of a low-flying satellite using a GPS receiver.



18.6.2 GRACE

merenpohjan
paine

GRACE (Gravity Recovery And Climate Experiment, **GRACE Mission**) measured the *temporal changes* in the Earth's gravity field, very precisely, at a time resolution of about a month, at a fairly crude spatial resolution. These temporal changes are mostly caused by movements of the “blue film”, the atmosphere and hydrosphere. The quantity measured is also called the *sea-floor pressure*, perhaps somewhat surprisingly. The explanation is, that the quantity is proportional to all of the mass contained in a column of air and water.

GRACE was a *satellite pair*: the satellites (“Tom and Jerry”) flew in a tandem formation at some 450 km height, at an average separation of 220 km. A microwave link measured changes in the distance between the satellite at an accuracy of $1 \mu\text{m/s}$. Both satellites also carried sensitive accelerometers to measure and eliminate atmospheric drag.

The measurement system was so sensitive, that changes in a water layer of even a millimetre's thickness could be noticed, if they extended over an area the size of a continent, some 500 km.

The successful launch took place in 2001. The data has been a treasure for hydrologists (**GRACE Mission, hydrology**). It was not until October 2017 when one of the satellites developed a malfunction that measurements came to a close. In the following months, the satellites re-entered the atmosphere.

A **GRACE Follow-On Mission** (GRACE-FO) was launched in 2018.



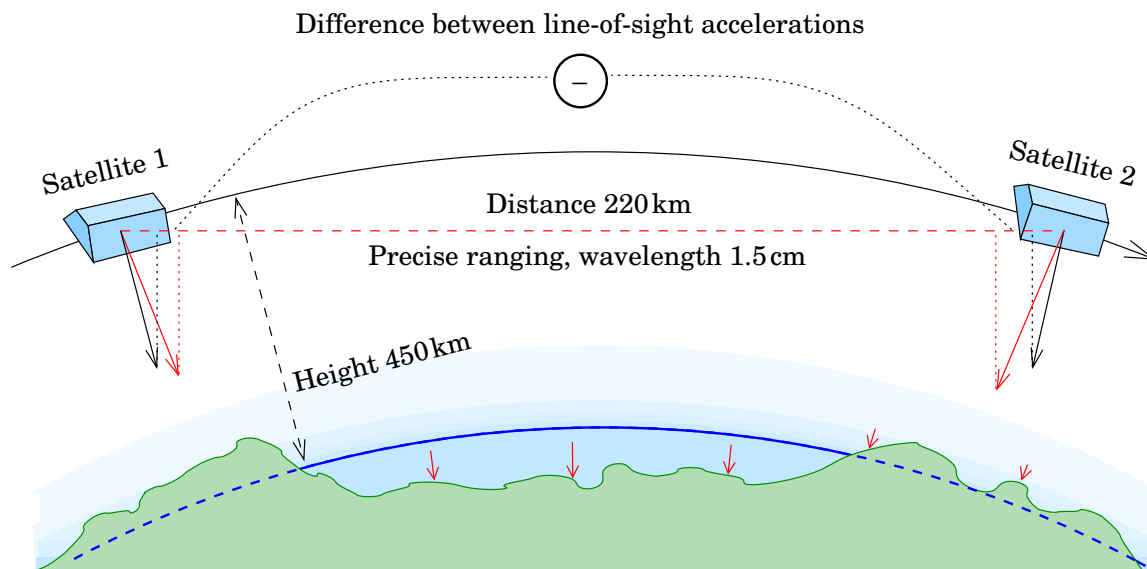


FIGURE 18.13. Basic idea of the **GRACE** satellite pair: measuring the tiny temporal changes in the Earth's gravity field using **SST**, satellite-to-satellite tracking. The satellites measure mass shifts in the Earth's "blue film" — atmosphere, hydrosphere — variations of "total sea-floor pressure" ↓.



18.6.3 **GOCE**

GOCE (Gravity Field and Steady State Ocean Circulation Explorer) was the most ambitious of the satellite missions. The satellite was launched from Plesetsk, Russia, on 17 March 2009. The orbital height was only 250 km, and the satellite carried a rocket engine (an ionic engine) and a stock of propellant (xenon) for orbit maintenance against atmospheric drag. The **GOCE** payload contained a *gravitational gradiometer*, an instrument for precisely measuring components of the *gradient* of the Earth's attraction, the dependence of the attraction on the three co-ordinates of place. The gradiometer consisted of several extremely sensitive accelerometers mounted on a frame.

ajoine

GOCE has worked well. However, in July 2010 a serious malfunction occurred in the telemetry link, which was repaired in August. The mission ended on the 21th of October 2013, and on the 11th of November, the satellite returned into the atmosphere and was seen burning up over the Falkland Islands.

It has been theoretically established that gradiometry is the best way to measure the very local features of the gravity field, better than orbital tracking with **GNSS**. The smallest details in the geoid map that can be



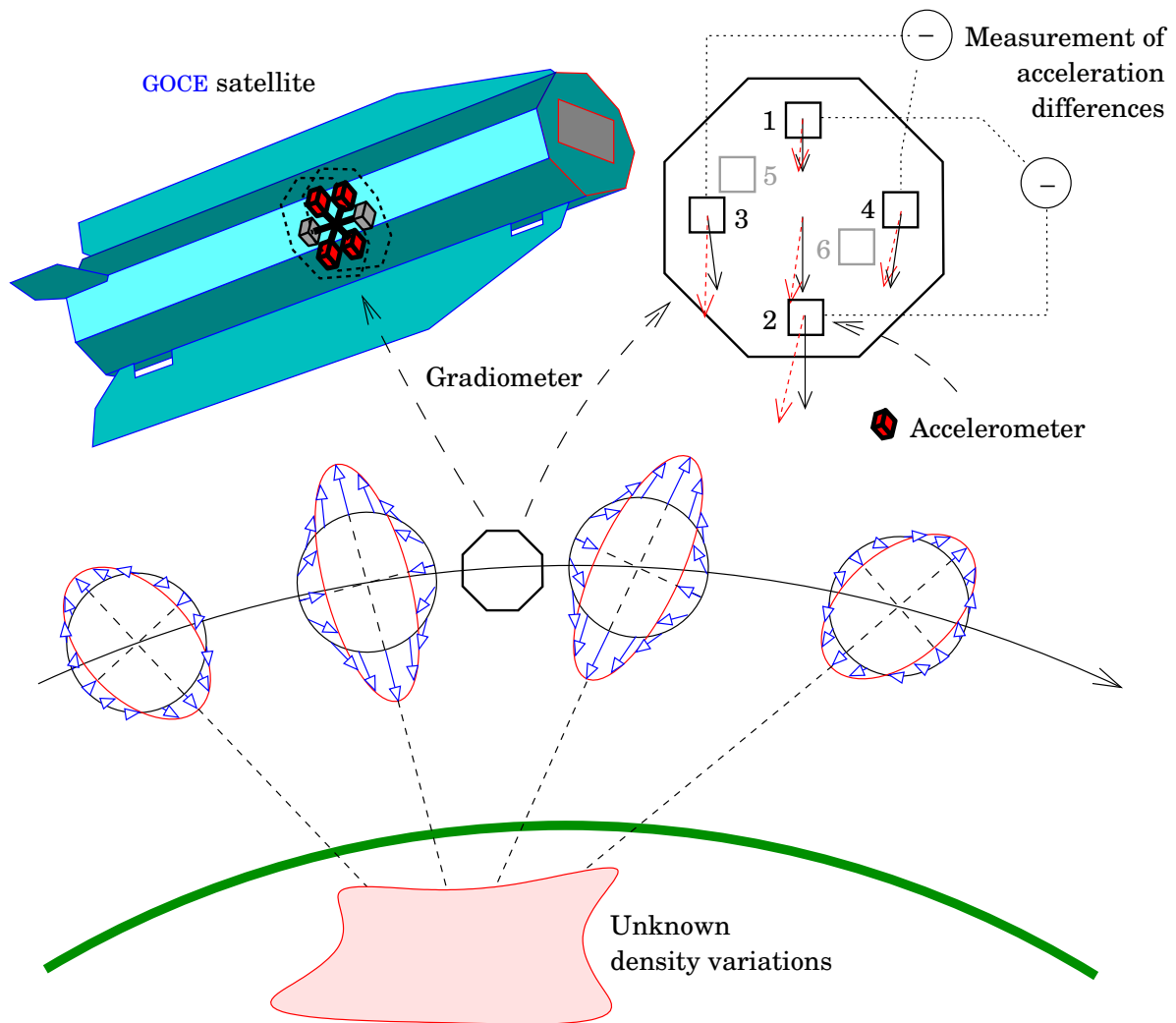


FIGURE 18.14. Determining the Earth's gravitational field with the gravitational gradiometer on-board the GOCE satellite.

seen in the GOCE material are 100 km in diameter, their precision being as good as ± 2 cm.

meritopografia One important application for a global geoid map this precise is mapping the deviations of the mean sea surface from the geoid, an equipotential surface, with similar precision. The true location in space of sea level is obtained by satellite radar altimetry, also at a precision level of several centimetres. This difference in level between the mean sea surface and an equipotential surface, the *sea-surface topography*, may be *inverted* into a map of ocean currents. The theory behind this is explained in subsection 18.10.2. See figure 18.15.

The name of the GOCE mission was inspired by this possibility.



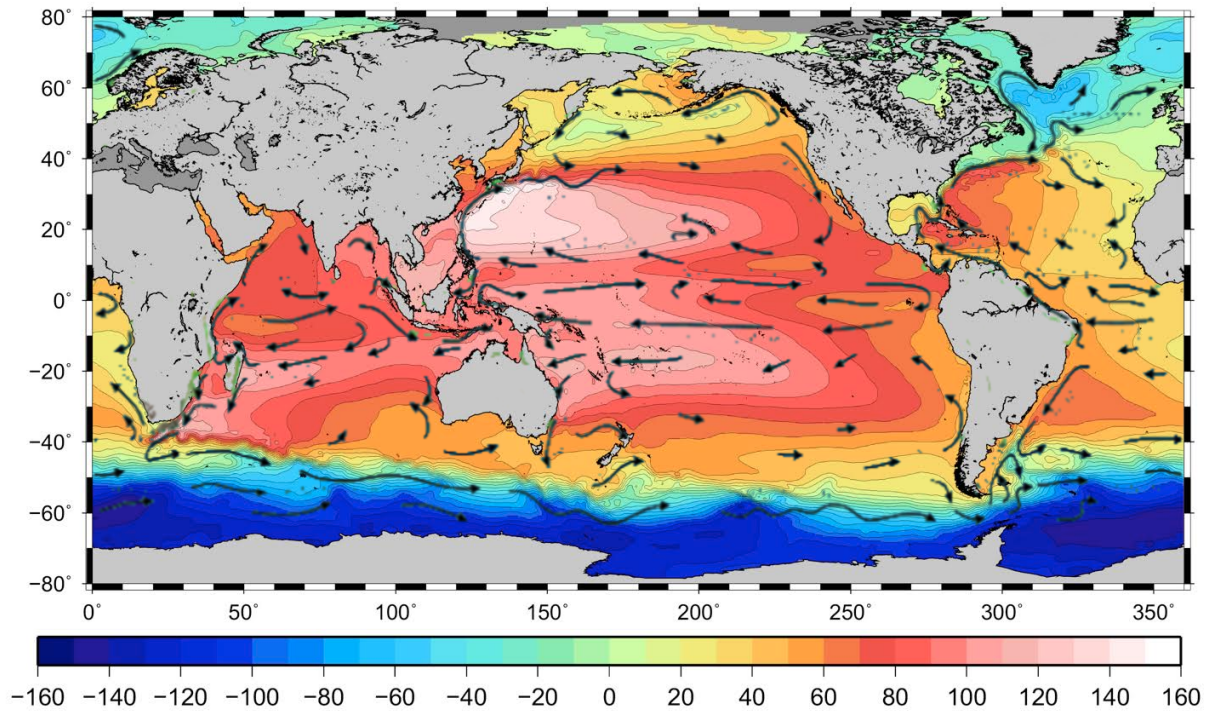


FIGURE 18.15. A sea-surface topography map produced by the [GOCE](#) mission © European Space Agency. Unit cm. Ocean surface currents drawn on top, [NOAA](#) / Rick Lumpkin ([NOAA](#), [Ocean currents](#)). Compare with figure [18.21](#)!



18.7 Atmospheric research and GNSS

18.7.1 Water-vapour values from a GNSS network

The atmosphere affects the propagation of radio waves, and thus of the [GNSS](#) signal. As can be seen from the refractive-index equation for [taitekerroin](#) microwaves [7.6](#):

$$N_M = 10^6 \cdot (n_M - 1) = \frac{77.624 \text{ K/hPa}}{T} (p - e) + \frac{64.70 \text{ K/hPa}}{T} \left(1 + \frac{5748 \text{ K}}{T} \right) e,$$

both the total air pressure p and the partial pressure of water vapour, or *absolute humidity*, e affect propagation, but in different ways. If we substitute $T \approx 285 \text{ K}$ (approximately $12 \text{ }^\circ\text{C}$), we obtain

$$N_M = 0.27(p - e) + 4.8e.$$

Write this into the form

$$N_M = ap_{\text{dry}} + bp_{\text{wet}},$$

in which $p_{\text{dry}} = p - e$ is the total of the partial pressures of the dry constituents of the atmosphere, and $p_{\text{wet}} = e$ is the partial pressure of



water vapour. The coefficients a and b describe the influences of dry air (mostly nitrogen, oxygen, argon, and carbon dioxide) and of water vapour. The values $a \approx 0.27 \text{ hPa}^{-1}$ and $b \approx 4.8 \text{ hPa}^{-1}$ depend somewhat on temperature. However, b is some 18 times a : *water vapour affects the propagation of the GNSS signal some 18 times more strongly than dry air.*

poolisuus Behind this phenomenon is the *chemical polarity* of the water molecule; see section 7.5.

The partial pressure of water vapour is a less-used way to describe how much water vapour there is in the atmosphere. The more-used way is *relative humidity* in percentages, the amount of water vapour compared to the amount at *saturation*, when there is as much water vapour in the air as there can be at a certain temperature, before it starts condensing out as liquid. The saturation partial pressure e_{sat} is given in fair approximation by the equation of *Clausius*⁵ and *Clapeyron*:⁶

$$e_{\text{sat}}(T) = e_{\text{sat}}(T_0) \exp\left(-4895 \text{ K} \cdot \left(\frac{1}{T} - \frac{1}{T_0}\right)\right).$$

When we know that the temperature $T_0 = 100 \text{ }^\circ\text{C} = 373.15 \text{ K}$ is the boiling point of water: $e_{\text{sat}}(T_0) = 1 \text{ atmosphere} = 1013.25 \text{ hPa}$, it follows that

$$e_{\text{sat}}(T) = 1013.25 \text{ hPa} \cdot \exp\left(-4895 \text{ K} \cdot \left(\frac{1}{T} - \frac{1}{373.15 \text{ K}}\right)\right). \quad (18.1)$$

The lower the satellite is in the local sky, the longer the path through the atmosphere. If the zenith angle is ζ , we may describe the signal propagation delay in the atmosphere by the following equation (\tilde{a} and \tilde{b} are the integrals of the above-mentioned coefficients a and b over the full height of the atmosphere, divided by 10^6):

$$d_{\text{trop}} = \frac{\tilde{a}p_{\text{dry}} + \tilde{b}p_{\text{wet}}}{\cos \zeta} = \frac{\tilde{a}(p - e) + \tilde{b}e}{\cos \zeta} = \frac{d_{\text{zenith}}}{\cos \zeta}. \quad (18.2)$$

verkkotasoitus In a GNSS network adjustment, we may solve the station values d_{zenith} , **zeniitin** zenith propagation delays, as unknowns, by substituting equation 18.2 **kulkuviive** into the GNSS observation equations 12.1 — and with the identity $D_{\text{trop}} = d_{\text{trop}}$ into equations 12.2. (In a small area, one can only resolve zenith-delay *differences* between stations.)

ilmapuntari If we also measure, at the same station, the total air pressure with a barometer:

⁵Rudolf Julius Emanuel Clausius (1822–1888) was a German physicist and mathematician, one of the founders of thermodynamics.

⁶Benoît Paul Émile Clapeyron (1799–1864) was a French engineer and physicist, one of the 72 names on the Eiffel Tower (**Eiffel Tower, 72 names**).



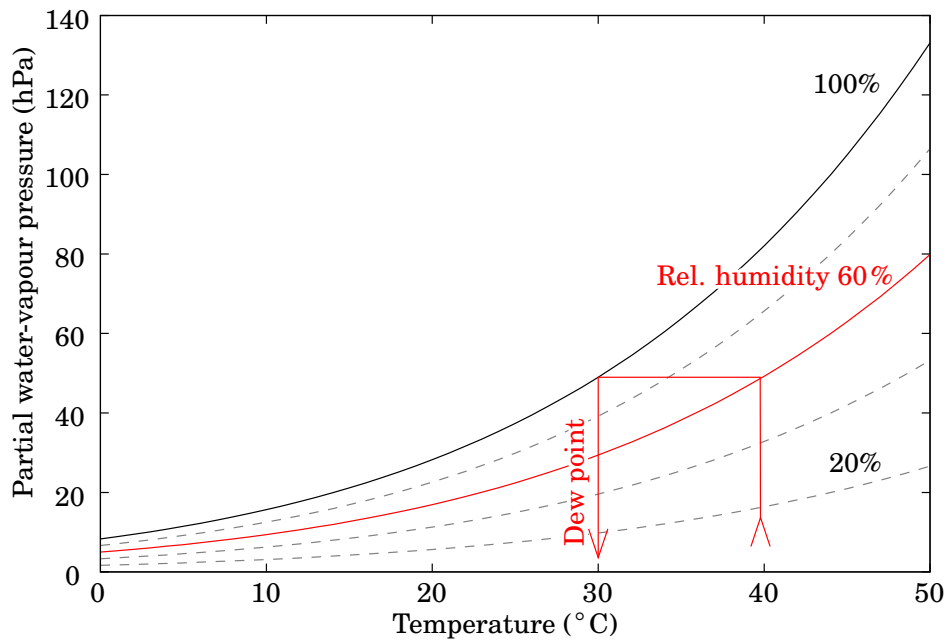


FIGURE 18.16. Saturation partial pressure e_{sat} and partial pressures e of water vapour at various temperatures and relative humidities according to equation 18.1. The red path shows how the dew point (30° C) follows from temperature (40° C) and relative humidity (60%). This means that temperature and dew point together can be used to determine relative humidity. Many hygrometers — devices for measuring the humidity of air — are based on this principle.



$$p = p_{dry} + p_{wet},$$

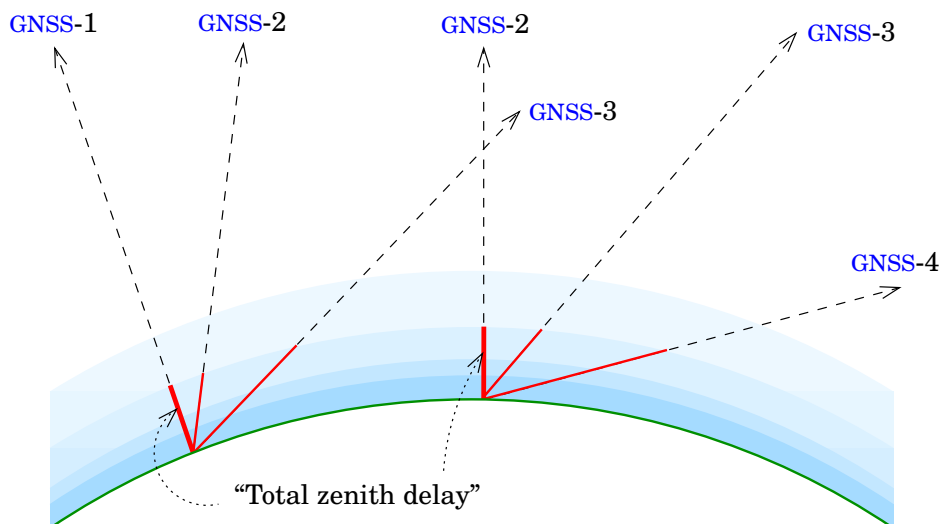


FIGURE 18.17. Use of GNSS for studying the troposphere.



we may solve separately for p_{dry} and p_{wet} .

The latter quantity, converted to amount of matter in an air column, is also called the *integrated water-vapour content* or *total precipitable water-vapour content* (converted to liquid water, unit mm). It is a quantity that weather and climate researchers are very much interested in!



18.7.2 GNSS radio occultation or “limbsounding”

Using low-flying satellites, it is possible to measure how the radio signal from a GNSS satellite is slowed down as the lowest point of the ray dives deeper and deeper into the atmosphere. This technique is called GNSS limbsounding or radio occultation. The first satellite to exploit this technique was GPS/MET. The Danish Ørsted satellite and the aforementioned CHAMP also used the technique.

The technique requires having a GNSS receiver on board the satellite. As there are already some 30 satellites in the GPS alone, and the low-orbit satellite goes around the Earth once every 1.5 hours, a quite substantial amount of information is collected every 24 hours.

The technique is important because it allows the determination of temperatures high up in the atmosphere, above the tropopause, where there is not much water vapour left. This is the layer in which climatic global warming would become most visible.⁷ The ozone layer is located at the top of the measurement range, and processes taking place there also affect the temperature.

The technique is the following: the delay caused by the atmosphere is proportional to air density, which is a piecewise exponential function of height.⁸

$$\rho(H) = \rho(H_0) \exp\left(-\frac{H - H_0}{S(t, k, g)}\right),$$

in which $S(t, k, g)$ is the *scale height*, a function of air temperature t , air composition k , and gravity g . If we assume that k ⁹ and g are known, we may calculate temperature t from scale height S .

⁷Climate models predict surprisingly, and observations confirm, that, while temperatures at the Earth’s surface and in the troposphere go up, a compensating lowering of temperatures is expected for the stratosphere. The cause is enhanced radiative cooling by increasing carbon dioxide concentrations.

⁸It is a little more complicated than this, as the GNSS signal travels through many atmospheric layers, not only the layer of closest approach to the Earth’s surface.

⁹The water vapour content at these heights is practically zero due to the low temperature.



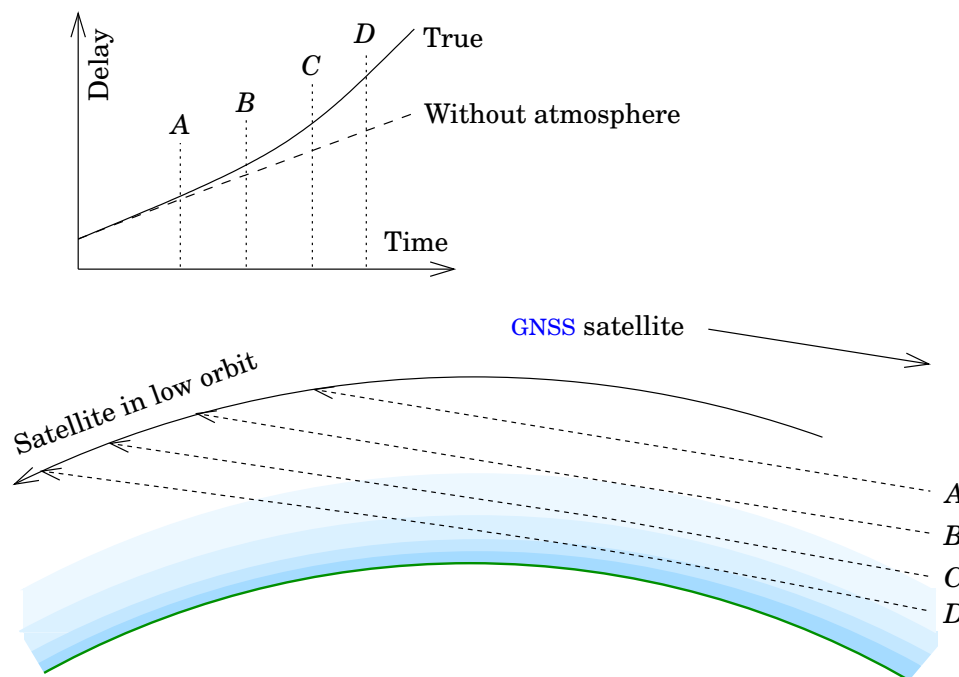


FIGURE 18.18. How the GNSS radio occultation technique works.



18.7.3 Ionosphere sounding

From GPS measurements at two frequencies, L_1 and L_2 , we may compute *ionospheric models*, in which the ionospheric electron density is mapped, as a function of place (φ, λ, h) and time. From this can then be calculated **TEC**, the total electron content, integrated along the path between satellite and receiver.

These models have been computed, nearly in real time, using the global GNSS network, since 1998 by the IGS, the International GNSS Service. A data format, **IONEX**, Ionosphere Map Exchange Format, has been developed in support of this activity.



18.8 Long-term variations in Earth rotation axis and orbit

Milanković¹⁰ proposed as early as in 1941 a hypothesis that variations in the *insolation* — the radiative power coming from the Sun — at the edges of the large continental ice sheets cause their growth and retreat, and that these variations would be caused by astronomical factors. The work of Milanković is based on earlier work by, among others, James Croll,¹¹ to explain the ice ages and interglacials.

¹⁰Milutin Milanković (1879–1958) was a Serbian polymath, engineer and climatologist.

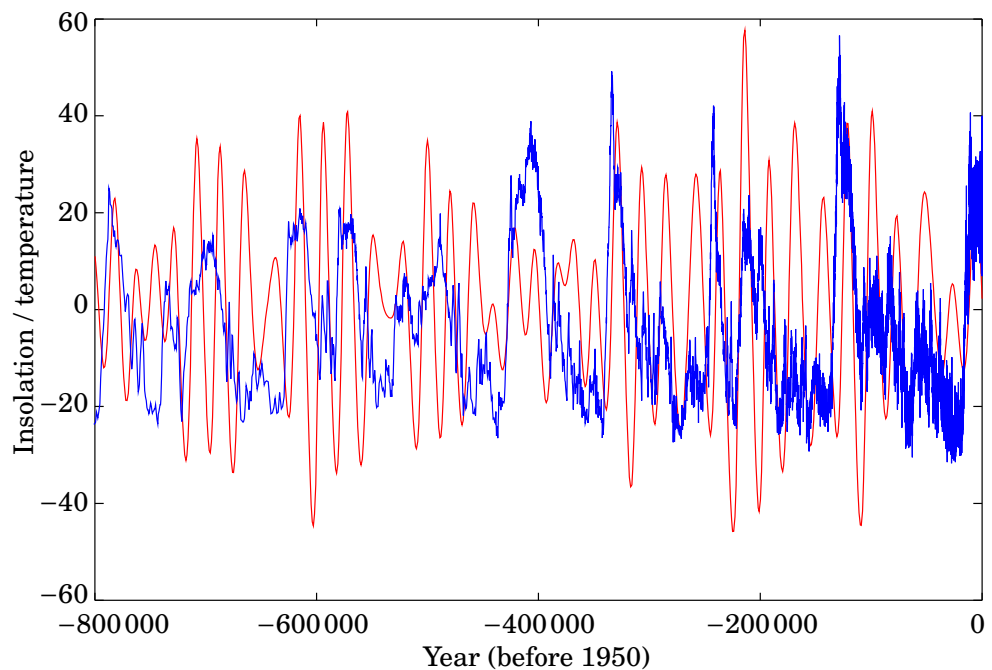


FIGURE 18.19. Milanković cycles over the past 800 000 years on both hemispheres. *Red*, theoretically computed summer insolation — from July 21 to August 20 — at 65° northern latitude (outside the atmosphere). *Blue*, Antarctic temperature estimate, computed using the deuterium isotope. Arbitrary scales. Sources: [Laskar et al. \(2004\)](#); [Jouzel and Masson-Delmotte \(2007\)](#), see [IMCCE, Milanković app](#) and the *Science* article [Jouzel et al. \(2007\)](#).



18.8.1 Tilt of the rotation axis

Besides the precessional motion, the obliquity, or tilt, of the Earth's axis of rotation with respect to the ecliptic also varies slowly. Currently (epoch 2000.0) the tilt angle is $23^\circ 26' 21''$ and is slowly diminishing, by about $0''.47$ per year.

Actually the variation is periodic, between $22^\circ 1'$ and $24^\circ 5'$, with a period of some 41 000 years. This variation is related to orbit perturbations caused by the planets on the Earth's orbital plane.



18.8.2 Climatological precession

Astronomical precession is the earlier described phenomenon, that the rotation axis of the Earth turns around in a conical figure in a period of about 25 800 years. In addition, the major axis of the Earth's orbital

isoakseli

¹¹James Croll [FRS](#) (1821–1890) was a Scottish autodidact physicist, astronomer and climatologist.

ellipse also rotates in the orbital plane, the period being 112000 years (apsidal precession). This motion is caused by orbit perturbations by the other planets.

The net result of both processes is *climatological precession*, the period of which is

$$\left(\frac{1}{112000} + \frac{1}{25800} \right)^{-1} \text{ a} = 21000 \text{ a.}$$

This is the time period in which the season when the Earth is closest to the Sun cycles through the calendar. Currently the Earth is closest to the Sun at the beginning of January. Only 11000 years ago, at the termination of the last ice age, the Earth was closest to the Sun in July. The result was 7% more solar radiation power for the Arctic summers.

The direct impacts of climatological precession are *opposite* in the northern and southern polar areas, because the boreal (northern) summer is the austral (southern) winter, and the austral summer is the boreal winter.

18.8.3 Variation of the orbital eccentricity

The eccentricity of the Earth's orbit is also variable, with important periods being ~ 100000 and 413000 years, between the values 0 and 0.068. This variation, caused by orbit perturbations from the other planets, modulates the effect of climatological precession.

18.9 Land-ice research and climate change

Much geophysical research is aimed at continental ice sheets, where, thanks to *polar amplification* — the enhanced warming of the polar regions — the warming of our planet's climate is first becoming visible. A continental ice sheet is a *glacier*: it grows above the snow line from falling snow that slowly compacts, first into firn and then into ice, and flows slowly and plastically to the coasts, where it loses mass through melting and formation (“calving”) of icebergs.

poikastuminen

The thicknesses of both the Greenland and the Antarctic ice sheet have been monitored by satellite techniques, among them radar altimetry. Radar mapping from orbit (SAR, synthetic-aperture radar) has also been used for this. Observing the small changes of course requires long time series, and this is still a shortcoming of the satellite methods.

GRACE (subsection 18.6.2) has also been used to study the change in mass of these continental ice sheets. This change represents directly the



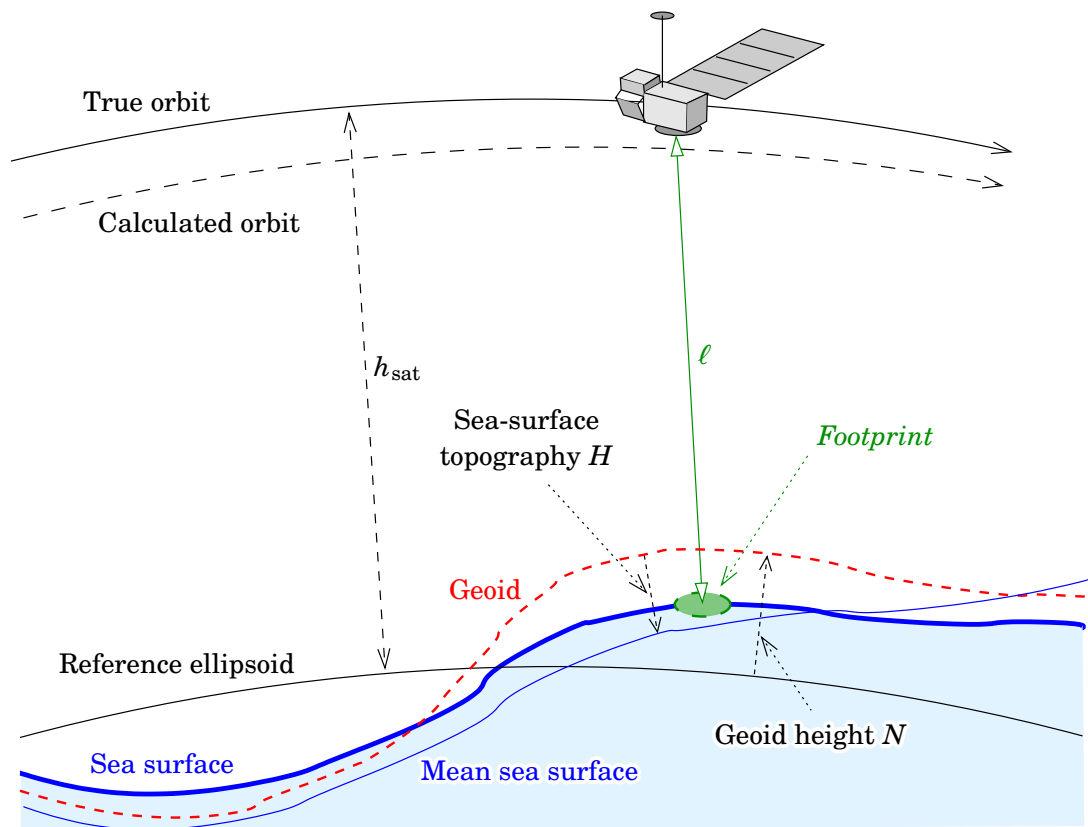


FIGURE 18.20. The measurement geometry of satellite radar altimetry.

contribution of these ice sheets to sea-level rise.

The GNSS technique has also been used for studying both crustal and ice motions, both in Greenland (Khan et al., 2010) and in Antarctica, for example in the surroundings of the Finnish Antarctic base Aboa in Dronning Maud Land (Koivula and Mäkinen, 2003).



18.10 Geodetic oceanography



18.10.1 Satellite altimetry and the geoid

It is possible to measure the location of the momentaneous sea surface in a geocentric system from a satellite. The instrument used is a *radar altimeter* (Rummel and Sansò, 1993) and its precision of measurement is a few centimetres. See figure 18.20.

A satellite altimetric radar works in this way: it sends a short microwave pulse straight downwards, where it is reflected back from the sea surface. The reflection is received and the shape of the return pulse is analysed. The reflection does not just come from the point straight underneath the satellite, but from a whole area called the *footprint*. This

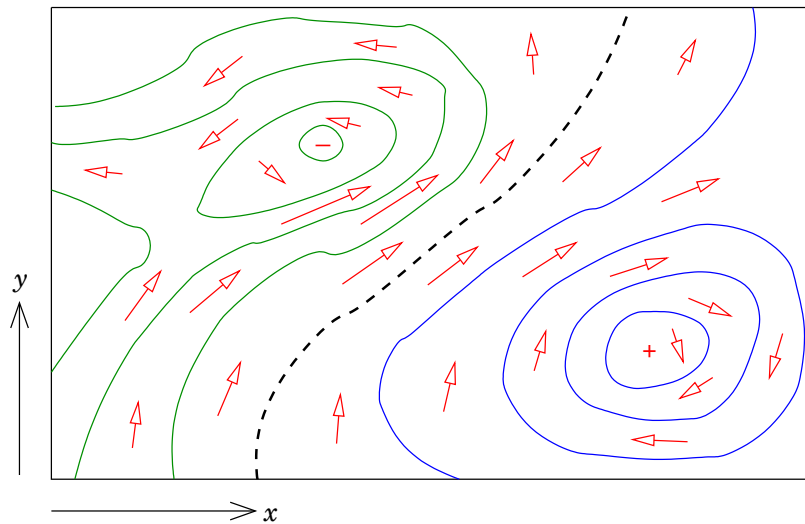


FIGURE 18.21. Theoretical link between sea-surface topography and ocean currents.

footprint may have a diameter of several kilometres, more if the sea state includes high waves. aallokko

Based on the analysis of the pulse shape, the distance between the satellite and the sea surface is inferred. If the satellite orbit is known — nowadays altimetric satellites always carry a GNSS receiver — the momentaneous, geocentric location of the sea surface may be computed. Successive measurement points form arcs running either from south to north or from north to south. rataakaari

18.10.2 Ocean currents and sea-surface topography

Mean sea level aligns pretty nicely with the geoid, an equipotential surface of the Earth's gravity field. If sea water were in a “state of rest”, an equilibrium state like in communicating vessels, the mean sea surface would be exactly the same as the geoid. This is, however, not the case. The causes of this *sea-surface topography* are yhtyvät astiat

- local variations in *temperature* and *salinity* suolaisuus
- variations in *air pressure* from place to place, to which the sea surface responds like an “inverted barometer” (IB) ylösalainen ilmapuntari
- The *drag* of winds with the sea surface, pushing water masses in certain directions
- the Coriolis force acting on ocean currents. coriolisvoima

Ocean currents respond to the rotation of the Earth though the Coriolis¹²

¹³ force, in the same way as air currents (Buys Ballot's¹³ law): currents try to turn to the right in the Northern Hemisphere, to the left in the Southern one. This causes a *transversal tilt* of ocean currents, which is proportional to the flow velocity and the sine of latitude, $\sin \varphi$. We speak of *geostrophic flow*.

The magnitude of the sea-surface topography is globally ± 1 m. The Baltic Sea also has a sea-surface topography, a tilt of the sea surface, as a result of which the bottoms of the Gulfs of Finland and Bothnia are about 20–30 cm higher than the western parts of the Baltic Sea, due to the salinity gradient. There is also a height difference across the Danish straits, caused by the push of predominantly westerly winds.

pohjukka

Earlier on, some satellite missions were mentioned that were intended to determine the precise location in space of the *geoid*. The sea-surface topography on the oceans, the difference between mean sea level and the geoid, is caused mostly by *ocean currents*. The rotation of the Earth causes currents to tilt in the transversal direction. Using this observable phenomenon, it is possible to theoretically compute the volume of the current, the amount of water transported, and from this in turn the thermal energy carried along with the water. This is an essential matter for climate research. This was one objective of the ambitious **GOCE** project (subsection 18.6.3).

In figure 18.21 we see the relationship between ocean currents and sea-surface topography. We may also describe this in equations, the *geostrophic equations*:

$$\frac{\partial H}{\partial x} = +2v_y \frac{\omega}{\gamma} \sin \varphi, \quad \frac{\partial H}{\partial y} = -2v_x \frac{\omega}{\gamma} \sin \varphi, \quad (18.3)$$

in which $H(x, y)$ is the sea-surface topography, ω the rotation rate of the Earth, γ gravity, x and y plane co-ordinates in the east and north directions, and v_x and v_y are the east and north components of the flow velocity. Because the equations contain $\sin \varphi$, the effect is opposite in the Northern and Southern Hemispheres, and in the immediate vicinity of the equator, it vanishes.

By simple partial differentiation, we may infer from the observed

¹²Gaspard-Gustave de Coriolis (1792–1843) was a French mathematician, one of the 72 names on the Eiffel Tower ([Eiffel Tower, 72 names](#)).

¹³Christoph Hendrik Diederik Buys Ballot (1817–1890) was a Dutch chemist and meteorologist.



sea-surface topography $H(x, y)$ to the flow pattern $\mathbf{v}(x, y)$, in which $\mathbf{v} = v_x \mathbf{i} + v_y \mathbf{j}$ is the current velocity vector in the (x, y) plane.

18.10.3 Mareographs, satellite altimetry, sea-level rise

Starting in 1992, the French-American satellite [TOPEX/Poseidon](#) and its follow-up satellites [Jason-1](#), [Jason-2](#) and [Jason-3](#) ([NASA JPL, TOPEX/Poseidon, Jason](#)) have been doing pioneering work monitoring the rise of the global sea level. Traditionally, the sea level has been monitored using *mareographs* or *tide gauges*; the rise of the sea level over the whole 20th century has been estimated to have been of the order of 1.3–2 mm/a, or 13–20 cm over the whole century.¹⁴

14

A mareograph measures the rise in the sea level with respect to the Earth's crust on which it has been built. A long tube connects the instrument with the open sea, in order to dampen wave motion.

This method has two problems:

- The Earth's crust itself may move. It is mostly the post-glacial land uplift, more commonly called [GIA](#) (glacial isostatic adjustment) that is important here: it affects Fennoscandia and Canada most strongly, but its effects extend to all of Europe and North America. Outside the land-uplift area there is a broad zone in which the land is slowly subsiding, the “periglacial bulge”. This phenomenon must be carefully modelled and removed from the mareograph data, if one wishes to compute global sea-level rise precisely. Nowadays there are continuously operating [GNSS](#) stations co-located with many mareographs, with the help of which the vertical land motion may be empirically determined. Unfortunately these time series are still short.
- There are only mareographs on the coasts, and there, only at individual points. There are large ocean areas where there is not a single mareograph nearby. And the further one goes back in time, the weaker the situation becomes. In the Southern Hemisphere, the situation was really weak before around 1950.

Altimetric satellites can measure the location of the sea surface *geocentrically* — using an on-board [GNSS](#) positioning device — and *everywhere* underneath their orbits. For example, the inclination angle of the orbit

¹⁴A new analysis ([Dangendorf et al., 2017](#)) proposes that the sea-level rise before 1990 was substantially slower, only some 1.1 ± 0.3 mm/a.

ratatason
prekessio

of a satellite like [TOPEX/Poseidon](#) is 66° , and the whole ocean surface between latitudes 66°S and 66°N is being mapped at an interval of some ten days. When the plane of the satellite orbit precesses, with respect to the Sun, once in 117 days, we obtain the long-term trend in the global mean sea level with at least this temporal resolution. During the latest two decades, the mean sea level has risen some $3.2 \pm 0.4 \text{ mm/a}$, clearly more than during the 20th century on average ([Sea Level Research Group](#)). The reason for the acceleration of the rise is undoubtedly the warming of the climate.



18.10.4 Height systems, mareographs, sea-surface topography

Mareographs may also be used to observe the sea-surface topography, but only on the coast. This presupposes that the mareographs are connected to the same height system. In Finland, the matter has been organised so that the 13 mareographs on the Finnish coast operated by the Finnish Institute of Marine Research, nowadays part of the Finnish Meteorological Institute, are connected to the national precise-levelling network, which the Finnish Geodetic Institute, today the Finnish National Land Survey's Geospatial Research Institute [FGI](#), maintains. The necessary connecting levellings are carried out by the geodesists at a few years' intervals.

The Finnish mareographs measure continuously the level of the local sea surface with respect to the local Earth's crust. The measurement values are transferred electronically to the headquarters of the Finnish Meteorological Institute in Helsinki, where they are processed and archived.

Near five mareographs, *pillars* have been erected by the Finnish Geodetic Institute in order to carry out regular [GNSS](#) measurements. In this way one may also obtain the local sea surface in a geocentric reference frame by means of satellite positioning.

Use of [GNSS](#) buoys is also becoming more common: these obtain the geocentric location of the momentaneous sea surface, far away from the coast, too. The sea state may also be monitored.



Self-test questions

1. Which geodetic observation techniques useful for geodynamics do you know?
2. What is the evidence showing that the Earth's outer core is liquid?



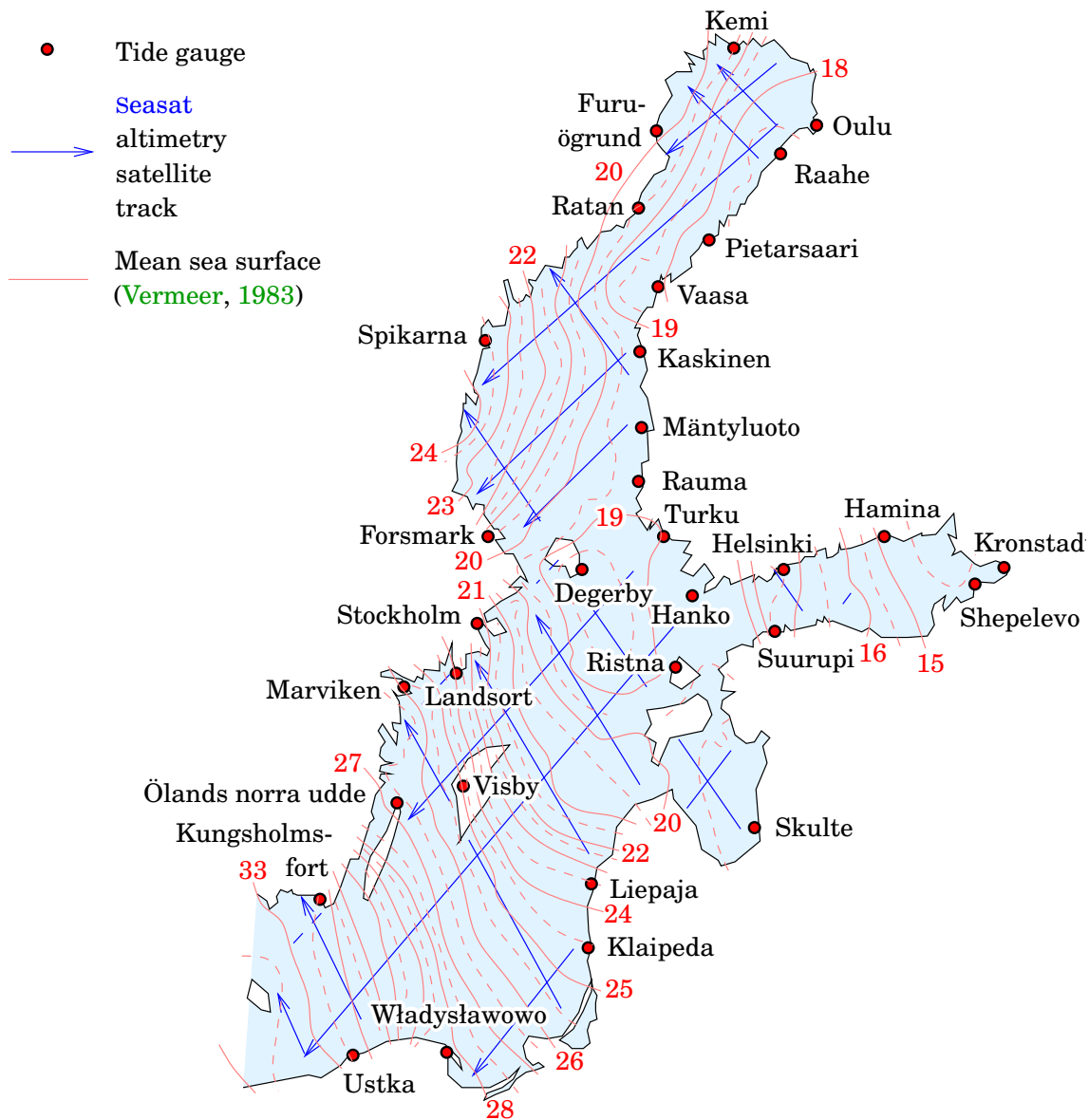


FIGURE 18.22. Tide gauges (mareographs) of the Baltic Sea, some ground tracks of the **Seasat** satellite back in 1978, and contours of a solution for the mean sea surface computed from the satellite's data.



3. Describe the polar motion and its monitoring.
4. In what way were the mission objectives of the **GRACE** and **GOCE** satellite missions different?
5. How can a network of **GNSS** stations be used to determine the water-vapour content of the atmosphere above it?
6. What is the atmospheric scale height, and how can **GNSS** radio occultation (“limbsounding”) be used to determine it?



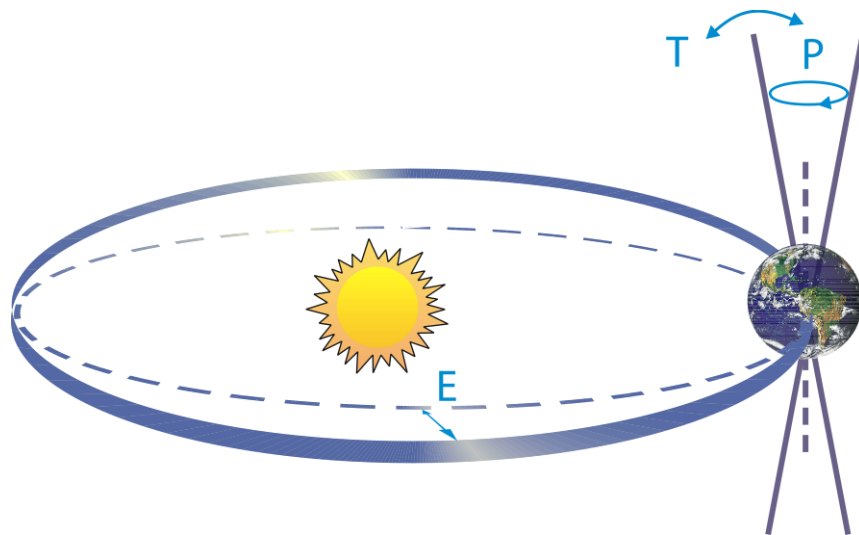


FIGURE 18.23. “Schematic of the Earth’s orbital changes (Milankovitch cycles) that drive the ice age cycles. ‘T’ denotes changes in the tilt (or obliquity) of the Earth’s axis, ‘E’ denotes changes in the eccentricity of the orbit (due to variations in the minor axis of the ellipse), and ‘P’ denotes precession, that is, changes in the direction of the axis tilt at a given point of the orbit.”



From IPCC (2007), Frequently Asked Question 6.1, page 449.

7. What are the three changes in the Earth’s rotation axis and orbital motion causing long-term climatic variations — the glacial cycle — according to the theory of Milanković?
8. See figure 18.23. After what you learned about Kepler’s laws of orbital motion, what is wrong with this picture (and caption text)?
9. According to a newspaper article, Associated Press (1988), military aircraft that crash-landed in Greenland during the Second World War were found 46 years later 260 feet deep in the ice. Based on your knowledge of how glaciers work, explain how they ended up this deep. What is the annual snowfall rate at the site of the crash?
10. What is the Coriolis force?
11. The geostrophic equations are

$$\frac{\partial H}{\partial x} = +2v_y \frac{\omega}{\gamma} \sin \varphi, \quad \frac{\partial H}{\partial y} = -2v_x \frac{\omega}{\gamma} \sin \varphi. \quad (18.3)$$

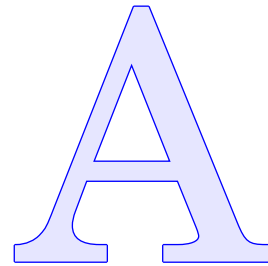
Explain each occurring symbol.

12. What was and is the mission objective of the TOPEX/Poseidon satellite and its follow-up satellites, the Jasons?





Properties of matrices



Matrices are used very widely in science to describe multidimensional phenomena, like co-ordinate transformations and mappings from one vector space to another. The matrix is written out as a two-dimensional table of numbers between parentheses or square brackets. The *symbolic* notation for a matrix is an italic¹ capital letter, like *A*, *B*, *I*. 1

Formally, matrices are a *type of number*, like real numbers and complex numbers. Many operations on numbers, like addition and subtraction, also exist for matrices: same-sized matrices may be added simply by adding together the corresponding elements. Other operations, like multiplication, do not always work between matrices: they require that the *sizes* of the matrices, the numbers of their rows and columns, are suitable. To wit, the number of columns in the left matrix must be the same as the number of rows in the right matrix, for multiplication to be even possible.



A.1 Adding and subtracting matrices

Same-sized matrices are added by adding together the corresponding elements:

$$C = A + B$$

means

$$c_{ij} = a_{ij} + b_{ij}$$

for every pair i, j , $i = 1, \dots, n$, $j = 1, \dots, m$, if the matrices A , B , and C are of size $n \times m$. For example, in the following way ($n = m = 2$):

$$\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix},$$

¹Alternatively, many authors use bold capital letters.

with the numerical example

$$\overbrace{\begin{bmatrix} 5 & 2 \\ 1 & 4 \end{bmatrix}}^C = \overbrace{\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}}^A + \overbrace{\begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix}}^B,$$

in which $c_{11} = a_{11} + b_{11}$, and so on. The same applies for subtraction.



A.1.1 Multiplying a matrix with a constant

This amounts to multiplying every element with this constant. For example,

$$k \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} ka_{11} & ka_{12} \\ ka_{21} & ka_{22} \end{bmatrix}.$$



A.2 Matrices and vectors

A matrix with only one row is called a *row vector*; a matrix with only one column, a *column vector*. For example,

$$r = \begin{bmatrix} r_1 & r_2 & r_3 & r_4 \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}.$$

Vectors are commonly denoted by small letters.



A.3 The unit matrix

In the same way that for numbers there exists a *zero element* 0, for which holds $\forall a \in \mathbb{R} : a + 0 = 0 + a = a$, and a *unit element* 1 with the property $\forall a \in \mathbb{R} : a \cdot 1 = 1 \cdot a = a$, these also exist for matrices. The unit matrix is simply

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

It is always square. Every number of dimensions has its own unit matrix, like

$$I_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

is the unit matrix for three dimensions.





A.4 Matrix multiplication

Multiplication of matrices is done following the simple scheme “row times column”. If

$$C = A \cdot B,$$

this means that for every i and k :

$$c_{ik} = \sum_{j=1}^n a_{ij} b_{jk},$$

meaning that the matrix

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1k} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{i1} & \cdots & [c_{ik}] & \cdots & c_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nk} & \cdots & c_{nn} \end{bmatrix}$$

is obtained by multiplying every row of matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ [a_{i1}] & \rightarrow & [a_{ij}] & \rightarrow & [a_{in}] \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}$$

with every column of matrix

$$B = \begin{bmatrix} b_{11} & \cdots & [b_{1k}] & \cdots & b_{1n} \\ \vdots & \ddots & \downarrow & \ddots & \vdots \\ b_{j1} & \cdots & [b_{jk}] & \cdots & b_{jn} \\ \vdots & \ddots & \downarrow & \ddots & \vdots \\ b_{n1} & \cdots & [b_{nk}] & \cdots & b_{nn} \end{bmatrix}.$$

So, one sums $c_{ik} = a_{i1} b_{1k} + a_{i2} b_{2k} + \cdots + a_{in} b_{nk}$, “row i times column k ”, for each element of the matrix C . In the above formula we have marked one row, with row number i , in the matrix A , and one column, with column number k , in the matrix B . Multiplication of this row with this column thus yields the element of C called c_{ik} .

Here it is assumed, for the sake of simplicity, that all matrices are square, $n \times n$ rows and columns. This is not inevitable; however, the number of columns of A and the number of rows in B must be the same for multiplication to be possible.



Exercise Verify that for an arbitrary square matrix A it holds that

$$A \cdot I = I \cdot A = A.$$



A.5 The transpose

The transpose of matrix A , the matrix A^T , is defined as

$$\overbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}}^A \begin{matrix} \\ \\ \\ \\ \end{matrix}^T = \overbrace{\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}}^{A^T},$$

$[n \times m]$ $[m \times n]$

or more compactly

$$(A^T)_{ij} = (A)_{ji}, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

The transpose A^T of matrix A is thus obtained by interchanging rows and columns, or, if $m = n$ and we have a square matrix, by “mirroring” all elements through the main diagonal. Trivially, if the transposition is done twice, one obtains back the original matrix:

$$(A^T)^T = A.$$

The transpose of a *vector* changes a row vector into a column vector and vice versa:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}^T = [a_1 \ a_2 \ \cdots \ a_n], \quad [a_1 \ a_2 \ \cdots \ a_n]^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

The latter notation is often used in running text to save paper when presenting a column vector.



A.6 The inverse matrix

The *inverse matrix* N^{-1} of a square matrix N is defined as the matrix K , for which $K \cdot N = N \cdot K = I$, in which I is the unit matrix already defined above. A property of the unit matrix again is $I \cdot N = N \cdot I = N$ for all square matrices N of the same size. The role of the unit matrix is the same as for real numbers the role of the number 1. For a real number



a again, the concept corresponding to the inverse matrix is the number $1/a$.

There exist good numerical algorithms for inverting a matrix. Inversion is however heavy for large matrices: the computational work required is proportional to the matrix size n to the *third* power! So, inverting a size 20×20 matrix takes about 8 times the computation time needed for inverting a size 10×10 matrix.

Manual inversion is only readily possible for matrices of size up to 3×3 — and is not worth the bother (Reagan, 2021).

For example, if

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix},$$

its inverse matrix is

$$A^{-1} = \begin{bmatrix} 0.4 & -0.1 \\ -0.2 & 0.3 \end{bmatrix},$$

which is easily verified:

$$A \cdot A^{-1} = \begin{bmatrix} 3 \cdot 0.4 - 1 \cdot 0.2 & -3 \cdot 0.1 + 1 \cdot 0.3 \\ 2 \cdot 0.4 - 4 \cdot 0.2 & -2 \cdot 0.1 + 4 \cdot 0.3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I,$$

and similarly $A^{-1} \cdot A = I$.

A valuable tool in matrix computations is the MATLAB™ software which is in widespread use within the science community. For home users, there are similar open-source offerings like octave or scilab.

In practice, the computation is based on solving systems of linear equations: if $X = A^{-1}$ is the inverse of the matrix A , it holds, for the case of a size 2×2 matrix, that

$$AX = I,$$

or

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

or

$$\left. \begin{array}{l} a_{11}x_{11} + a_{12}x_{21} = 1 \\ a_{21}x_{11} + a_{22}x_{21} = 0 \end{array} \right\} \implies x_{11}, x_{21},$$

$$\left. \begin{array}{l} a_{11}x_{12} + a_{12}x_{22} = 0 \\ a_{21}x_{12} + a_{22}x_{22} = 1 \end{array} \right\} \implies x_{12}, x_{22}.$$

If the matrix is of size $n \times n$, similarly n systems of equations are created, with in each, n equations in n unknowns.



Singularity Just like the division $1/a$ does not work for all numbers a — specifically not for $a = 0$ — also the computation of the inverse matrix does not work for all matrices. A simple example of a non-invertible matrix is

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

and the solution equations above become conflicting:

$$x_{11} + x_{21} = \text{both } 1 \text{ and } 0,$$

$$x_{12} + x_{22} = \text{the same.}$$

Such a matrix is called *singular*.



A.7 The determinant of a matrix

The *determinant* of a square matrix is a single number that characterises the nature of the mapping that a matrix represents, from one vector space into another of the same number of dimensions: how a small volume element in one vector space is mapped onto the other one, the ratio of the source and destination volumes. Also, if the determinant vanishes, that means that the mapping has no inverse mapping and the matrix is singular.

The determinant is computed recursively:

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} &= \\ &= a_{11} \det \begin{bmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n3} & \cdots & a_{nn} \end{bmatrix} + \cdots \\ &\quad \cdots + (-1)^{n-1} a_{1n} \det \begin{bmatrix} a_{21} & a_{22} & \cdots & a_{2(n-1)} \\ a_{31} & a_{32} & \cdots & a_{3(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n(n-1)} \end{bmatrix}. \end{aligned}$$

The rule is: for every element of the first row, multiply with the determinant of the matrix formed by deleting the first row and the column of



that element.² And alternate the *algebraic sign*: a plus sign for odd- j a_{1j} , a minus sign for even- j a_{1j} .

A typical property of the determinant is that it is *fully antisymmetric*: exchange, in the matrix A , any two columns, or any two rows, with each other, and the determinant of the new matrix A' will be $\det A' = -\det A$.



A.8 Vectorial products

The scalar product of two vectors \mathbf{a} and \mathbf{b} is defined as

$$\langle \mathbf{a} \cdot \mathbf{b} \rangle \stackrel{\text{def}}{=} \|\mathbf{a}\| \|\mathbf{b}\| \cos \alpha,$$

in which α is the angle between the vectors.

Let $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ be an *orthonormal basis*, that is, a set of vectors spanning an n -dimensional vector space, for which

$$\langle \mathbf{e}_i \cdot \mathbf{e}_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Write

$$\mathbf{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \dots + a_n \mathbf{e}_n, \quad \mathbf{b} = b_1 \mathbf{e}_1 + b_2 \mathbf{e}_2 + \dots + b_n \mathbf{e}_n,$$

and

$$\bar{\mathbf{a}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad \bar{\mathbf{b}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Now, the scalar product of the vectors is calculated as a matrix product:

$$\langle \mathbf{a} \cdot \mathbf{b} \rangle \stackrel{\text{def}}{=} \bar{\mathbf{a}}^T \bar{\mathbf{b}} = \bar{\mathbf{b}}^T \bar{\mathbf{a}} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n.$$

It is seen that the scalar product is symmetric:

$$\langle \mathbf{a} \cdot \mathbf{b} \rangle = \langle \mathbf{b} \cdot \mathbf{a} \rangle.$$

The *vectorial product* of vectors is more complicated. In two dimensions it is a scalar:

$$\langle \mathbf{a} \times \mathbf{b} \rangle \stackrel{\text{def}}{=} \|\mathbf{a}\| \|\mathbf{b}\| \sin \alpha,$$

²This determinant is called the *cofactor* of that matrix element.



with α the *oriented* angle between the two vectors.

In three dimensions the vectorial product is itself a vector, of length

$$\|\langle \mathbf{a} \times \mathbf{b} \rangle\| = \|\mathbf{a}\| \|\mathbf{b}\| \sin \alpha,$$

which is perpendicular to both vectors \mathbf{a} and \mathbf{b} :

$$\langle \mathbf{a} \times \mathbf{b} \rangle \perp \mathbf{a}, \quad \langle \mathbf{a} \times \mathbf{b} \rangle \perp \mathbf{b}, \quad \iff \quad \langle \langle \mathbf{a} \times \mathbf{b} \rangle \cdot \mathbf{a} \rangle = \langle \langle \mathbf{a} \times \mathbf{b} \rangle \cdot \mathbf{b} \rangle = 0.$$

The vectorial product is antisymmetric:

$$\langle \mathbf{a} \times \mathbf{b} \rangle = -\langle \mathbf{b} \times \mathbf{a} \rangle,$$

and thus $\langle \mathbf{a} \times \mathbf{a} \rangle = 0$. It needs an *orientation*, a corkscrew rule: if you turn a corkscrew from \mathbf{a} to \mathbf{b} , it will move forward in the direction of $\langle \mathbf{a} \times \mathbf{b} \rangle$.

For the orthonormal basis in three dimensions $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ this means

$$\begin{aligned} \langle \mathbf{e}_1 \times \mathbf{e}_1 \rangle &= \langle \mathbf{e}_2 \times \mathbf{e}_2 \rangle = \langle \mathbf{e}_3 \times \mathbf{e}_3 \rangle = 0, \\ \langle \mathbf{e}_1 \times \mathbf{e}_2 \rangle &= \mathbf{e}_3, & \langle \mathbf{e}_2 \times \mathbf{e}_3 \rangle &= \mathbf{e}_1, & \langle \mathbf{e}_3 \times \mathbf{e}_1 \rangle &= \mathbf{e}_2, \end{aligned}$$

assuming that the basis vectors span a right-handed co-ordinate frame.

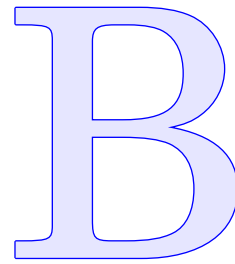
Calculation is done using determinants:

$$\begin{aligned} \langle \mathbf{a} \times \mathbf{b} \rangle &= \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = \\ &= \mathbf{e}_1 \det \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} + \mathbf{e}_2 \det \begin{bmatrix} a_3 & a_1 \\ b_3 & b_1 \end{bmatrix} + \mathbf{e}_3 \det \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} = \\ &= (a_2 b_3 - a_3 b_2) \mathbf{e}_1 + (a_3 b_1 - a_1 b_3) \mathbf{e}_2 + (a_1 b_2 - a_2 b_1) \mathbf{e}_3. \end{aligned}$$





A short introduction to magnetohydrodynamics



B.1 Plasma

A *plasma* is a gas in which the atoms have been stripped of part of their electrons: *ionisation*. The negatively charged electrons are moving around freely within the plasma, among the remaining, positively charged atom remnants, or *ions*. In space, plasmas are commonly hot, with temperatures varying from thousands to billions of degrees.

Plasmas in space are so extended and rarefied, that they behave like *superconductors*: electric currents run through them for a long time without noticeable weakening. For this reason, a magnetic field also cannot move through a plasma: matter and field lines move hand-in-hand. This is called *frozen-in* magnetism or Alfvén's theorem. If a plasma is compressed, the magnetic field contained in it will grow stronger: similarly if the plasma is stretched and rolled up like puff pastry, as is happening inside the Sun's convection layer. This is how natural dynamos work.

ohut
suprajohdin

lehtitaikina



B.2 Maxwell's equations

Maxwell's field equations are¹

$$\begin{aligned} \langle \nabla \times \mathbf{B} \rangle &= \mu_0 \mathbf{j} + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}, & \langle \nabla \times \mathbf{E} \rangle &= -\frac{\partial \mathbf{B}}{\partial t}, \\ \langle \nabla \cdot \mathbf{B} \rangle &= 0, & \langle \nabla \cdot \mathbf{E} \rangle &= \frac{\rho}{\epsilon_0}. \end{aligned} \quad (\text{B.1})$$

(\mathbf{E} electric field vector, \mathbf{B} magnetic field vector, \mathbf{j} vector of electric current density, ρ density of electric charge, ϵ_0 , μ_0 constants of nature.)

¹This compact form of the equations we owe to Oliver Heaviside (1850–1925). Heaviside was a telegraph researcher who developed the theory of electric circuits. He also proposed the existence of the *ionosphere* as an explanation for the long range — beyond the horizon — of radio waves.

The partial differential operators $\nabla \times$ and $\nabla \cdot$ are known in vector calculus as *curl* and *divergence* operators, curl and div. They are defined as follows:

$$\begin{aligned} \text{curl } \mathbf{E} &= \langle \nabla \times \mathbf{E} \rangle \stackrel{\text{def}}{=} \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{bmatrix} = \\ &= \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) \mathbf{i} + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right) \mathbf{j} + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right) \mathbf{k}, \\ \text{div } \mathbf{E} &= \langle \nabla \cdot \mathbf{E} \rangle \stackrel{\text{def}}{=} \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}, \end{aligned}$$

in which $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is the orthonormal basis of the (x, y, z) co-ordinate frame, and

$$\mathbf{E} = E_x \mathbf{i} + E_y \mathbf{j} + E_z \mathbf{k}.$$



B.3 “Frozen-in” magnetic field

In a superconductor there can be no net electric field integrated around a loop — as it would cause an infinite electric current! So

$$\oint_{\partial S} \langle \mathbf{E} \cdot \mathbf{t} \rangle ds = 0,$$

in which \mathbf{t} is the tangent vector of loop ∂S . The *Stokes loop integral theorem* ([Wikipedia, Kelvin–Stokes theorem](#)) — also called the Kelvin–

²Stokes theorem because Lord Kelvin² apparently discovered it — says that this integral is the same as the surface integral of the *curl* $\langle \nabla \times \mathbf{E} \rangle$ of \mathbf{E} over the surface S , the edge of which is ∂S . That integral must thus also vanish:

$$\iint_S \langle \langle \nabla \times \mathbf{E} \rangle \cdot \mathbf{n} \rangle dS = \oint_{\partial S} \langle \mathbf{E} \cdot \mathbf{t} \rangle ds = 0.$$

In this, \mathbf{n} is the normal on the surface S . From the *Maxwell–Faraday* equation [B.1](#) follows now

$$\iint_S \langle \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} \rangle dS = 0 \implies \iint_S \frac{\partial}{\partial t} \langle \mathbf{B} \cdot \mathbf{n} \rangle dS = \iint_S \langle \mathbf{B} \cdot \frac{\partial \mathbf{n}}{\partial t} \rangle dS.$$

Here, the integrand on the right, $\langle \mathbf{B} \cdot \frac{\partial \mathbf{n}}{\partial t} \rangle$, vanishes if the loop surface S is chosen to be fixed, as then $\frac{\partial \mathbf{n}}{\partial t} = 0$. Then, if also ∂S is constant:

$$\iint_S \frac{\partial}{\partial t} \langle \mathbf{B} \cdot \mathbf{n} \rangle dS = 0 \implies \frac{\partial}{\partial t} \iint_S \langle \mathbf{B} \cdot \mathbf{n} \rangle dS = 0.$$

²Sir William Thomson (1824–1907) [FRS](#) [FRSE](#) (Lord Kelvin) was a British physicist, engineer and inventor. He was ennobled in 1866 for his work on the trans-Atlantic telegraph cable.



The integral $\iint_S \langle \mathbf{B} \cdot \mathbf{n} \rangle dS$ is the magnetic *flux* through the loop surface S , the total number of field lines passing through the loop, which thus must be constant in time and cannot change. This is why a superconducting sphere floats, or levitates, on a magnetic field: the field cannot penetrate the sphere. vuo

In a moving plasma, the assumption that ∂S is constant no longer holds. We assert without proof ([Wikipedia, Alfvén's theorem](#)) that the above applies also in a plasma for an arbitrary loop connected with the plasma: the plasma and the field are moving hand-in-hand — though the plasma can freely flow along the field lines. If the plasma is compressed, the magnetic field strength increases. ³ 3



B.4 History of the field

The term “plasma” was invented by Nobel laureate Irving Langmuir⁴ (1927). The scientific discipline studying plasmas is called *magnetohydrodynamics* (MHD). It has been estimated that, in the universe, over 99% of normal — not “dark” — matter is plasma. 4

A well-known magnetohydrodynamicist was the Swedish Nobel laureate Hannes Alfvén, 1908–1995. He described in 1942 how in a superconductor the magnetic field lines move along with the matter.



Self-test questions

1. In a vacuum, both electric charge $\rho = 0$ and electric current $\mathbf{j} = 0$. The first two Maxwell equations [B.1](#) become

$$\langle \nabla \times \mathbf{B} \rangle = \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}, \quad \langle \nabla \times \mathbf{E} \rangle = -\frac{\partial \mathbf{B}}{\partial t}. \quad (\text{B.2})$$

Letting the operator $\nabla \times$ act from the left followed by substitution yields

$$\langle \nabla \times \langle \nabla \times \mathbf{B} \rangle \rangle = \epsilon_0 \mu_0 \frac{\partial \langle \nabla \times \mathbf{E} \rangle}{\partial t} = -\epsilon_0 \mu_0 \frac{\partial^2 \mathbf{B}}{\partial t^2}. \quad (\text{B.3})$$

³A spectacular example of this is the magnetic field of a neutron star or pulsar ([Wikipedia, Pulsar](#)), the collapsed, still rapidly spinning core of a star.

⁴Irving Langmuir (1881–1957) was an American chemist and physicist known for his research on matter.



We now use a theorem from vector calculus,

$$\langle \mathbf{a} \times \langle \mathbf{b} \times \mathbf{c} \rangle \rangle = \mathbf{b} \langle \mathbf{a} \cdot \mathbf{c} \rangle - \mathbf{c} \langle \mathbf{a} \cdot \mathbf{b} \rangle,$$

which in our case means

$$\langle \nabla \times \langle \nabla \times \mathbf{B} \rangle \rangle = \nabla \langle \nabla \cdot \mathbf{B} \rangle - \langle \nabla \cdot \nabla \rangle \mathbf{B}.$$

Here, according to the third Maxwell equation B.1, $\langle \nabla \cdot \mathbf{B} \rangle = 0$, so

$$\langle \nabla \times \langle \nabla \times \mathbf{B} \rangle \rangle = -\langle \nabla \cdot \nabla \rangle \mathbf{B} = -\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \mathbf{B},$$

and, with equation B.3:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \epsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \right) \mathbf{B} = 0. \quad (\text{B.4})$$

- Show by substitution that $\mathbf{B}(u)$ is a solution to this equation if $u = x - ct$ or $u = x + ct$. Use the chain rule.
- What is the value of parameter c ?
- What other solutions can you think of?
- Now let equations B.2 be acted upon by the operator $\nabla \times$ and substitute the other way around to find a similar equation for \mathbf{E} .
- Alternatively, assume $\mathbf{B} = \mathbf{B}(x, t)$ and decompose equation B.4:

$$\begin{aligned} \left(\frac{\partial^2}{\partial x^2} - \epsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \right) \mathbf{B} &= 0 \\ \implies \left(\frac{\partial}{\partial x} - \sqrt{\epsilon_0 \mu_0} \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial x} + \sqrt{\epsilon_0 \mu_0} \frac{\partial}{\partial t} \right) \mathbf{B} &= 0, \end{aligned}$$

to which the separate equations

$$\left(\frac{\partial}{\partial x} - \sqrt{\epsilon_0 \mu_0} \frac{\partial}{\partial t} \right) \mathbf{B} = 0, \quad \left(\frac{\partial}{\partial x} + \sqrt{\epsilon_0 \mu_0} \frac{\partial}{\partial t} \right) \mathbf{B} = 0$$

lead to the same solutions.

- [Advanced] Still alternatively, try separation of variables: write

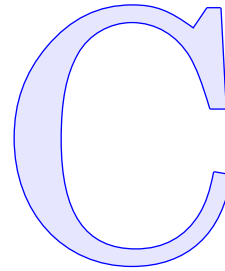
$$\mathbf{B}(x, t) = \mathbf{B}_0 \cdot X(x) \cdot T(t),$$

and derive separate differential equations for the scalar functions $X(x)$ and $T(t)$. Don't be confused by the vector constant \mathbf{B}_0 .





The Kepler orbital elements for satellites



C.1 Angular elements describing the *orientation* in space of the orbit

- Ω The *right ascension*, or astronomical longitude, of the ascending node. The zero point of this longitude is the place among the stars where the plane of the ecliptic and the plane of the equator intersect, and where the Sun crosses the equator at the beginning of spring: the *vernal equinox point*. rektaskensio
- i The inclination, the *tilt angle* of the orbital plane with respect to the equatorial plane. For GPS satellites $i = 55^\circ$. This is also the highest latitude (north and south) where the satellite moves through the zenith. kevät-päiväntasaus
- ω The argument of perigee. The angular distance, seen from the Earth's centre, between the ascending node and the perigee of the satellite orbit. nouseva solmu



C.2 Elements describing the orbit's *size and shape*

- a The semi-major axis of the satellite orbit, the “mean radius” of the orbit. The mean height of the satellite is obtained from this by subtracting the Earth's radius. isoakselin puolikas
- e The eccentricity of the satellite orbit, $e = \sqrt{a^2 - b^2} / a$, in which b is the semi-minor axis. This describes how much the Earth's centre is “to the side” from the centre of the orbital ellipse, in other words, how large is the height difference between the perigee (the lowest point of the orbit) and apogee (the highest point). pikkuakselin puolikas



C.3 Elements describing the satellite's place in its orbit, its "time table"

There are three alternatives:

luonnollinen
anomaliala

$v(t)$ true anomaly. The direction angle of the satellite in the orbital plane as seen from the Earth's centre, reckoned from the perigee in the direction of motion.

$E(t)$ eccentric anomaly. The direction angle of the satellite in the orbital plane as seen from the *centre of the ellipse*, after the orbital ellipse has been inflated into a circle.

$M(t)$ mean anomaly. The time that has elapsed since the last passage through the perigee, in units of the satellite's orbital period P .

Between these exist the following relationships, which may be used to convert them into each other:

$$\frac{\tan\left(\frac{1}{2}v(t)\right)}{\tan\left(\frac{1}{2}E(t)\right)} = \sqrt{\frac{1+e}{1-e}}, \quad E(t) = M(t) + e \sin E(t).$$

Study figure 17.14. E and v are geometric quantities. The mean anomaly M , on the other hand, is a mere measure of time, scaled to the period P of the satellite, and referenced to its time of passage τ through the perigee:

$$M(t) \stackrel{\text{def}}{=} 2\pi \frac{t - \tau}{P}.$$



Self-test questions

1. From the definition of inclination: "This is also the highest latitude (north and south) where the satellite moves through the zenith." Does this precisely apply on a flattened Earth?

Bibliography

ABCDEFGHIJKLMNOPRSTUVW

A

Michael S. Adler. Wikimedia Commons, Total solar eclipse of 2017, 2017. URL https://commons.wikimedia.org/wiki/File:Total_Solar_Eclipse_8-21-17.jpg. © 2017 (CC BY-SA 4.0). Accessed 19th May, 2019. 471

Joel Ahola. Nuottavaaran siirroksen deformaatiotutkimus GPS-mittausten avulla. Master's thesis, Helsinki University of Technology, Department of Surveying, Espoo, February 2001. 493

Associated Press. World War II planes found in Greenland in ice 260 feet deep. *The New York Times (archive)*, 4th August 1988. URL <https://www.nytimes.com/1988/08/04/us/world-war-ii-planes-found-in-greenland-in-ice-260-feet-deep.html>. Accessed 12th April, 2019. 512

AUSPOS – Online GPS Processing Service. Geoscience Australia. URL <http://www.ga.gov.au/scientific-topics/positioning-navigation/geodesy/auspos>. Accessed 10th April, 2019. 355

B

Willem Baarda. *A testing procedure for use in geodetic networks*, volume 2 number 5 of *Publications on geodesy, new series*. Netherlands Geodetic Commission, Delft, 1968. URL <https://www.ncgeo.nl/downloads/09Baarda.pdf>. Accessed 14th May, 2019. 167, 405

Willem Baarda. *S-transformations and criterion matrices*, volume 5 number 1 of *Publications on geodesy, new series*. Netherlands Geodetic Commission, Delft, second revised edition, 1981. URL <https://www.ncgeo.nl/downloads/18Baarda.pdf>. Accessed 2nd October, 2021. 365

BakerNunn.org. URL <http://bakernunn.org>. Accessed 19th May, 2019. 465

- Silvio A. Bedini. Along came a spider – Spinning silk for cross-hairs. The search for cross-hairs for scientific instrumentation, Part 1. *The American Surveyor*, March/April 2005a. URL https://archive.amerisurv.com/PDF/TheAmericanSurveyor_BediniAlongCameASpider-SpinningSilkForCross-HairsPart1_March-April2005.pdf. Accessed 14th May, 2019. 99
- Silvio A. Bedini. Along came a spider – Spinning silk for cross-hairs. The search for cross-hairs for scientific instrumentation, Part 2. *The American Surveyor*, May 2005b. URL https://archive.amerisurv.com/PDF/TheAmericanSurveyor_r_BediniAlongCameASpiderPart2_May2005.pdf. Accessed 14th May, 2019. 99
- Mirjam Bilker-Koivula and Matti Ollikainen. Suomen geoidimallit ja niiden käyttäminen korkeuden muunnoksissa. Research note (in Finnish) 29, Finnish Geodetic Institute, 2009. URL <https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/fgi/GLtiedote29.pdf>. Accessed 14th May, 2019. 279
- BIPM, SI base units. Bureau International des Poids et Mesures. URL <https://www.bipm.org/en/measurement-units/si-base-units>. Accessed 14th May, 2019. 24
- Boller and Chivens, Baker Nunn. URL https://bollerandchivens.com/?page_id=1455. Accessed 14th May, 2019. 465
- Ilmari Bonsdorff. Geodeettisen laitoksen vuosikertomus ajalta heinäkuu 1918 – joulukuu 1919. Annual report, Finnish Geodetic Institute, 1920. © 1920 Geodeettinen laitos. 211, 215
- Kai Borre. GPS Easy Suite II — A MATLAB Companion. *Inside GNSS*, March–April 2009. URL <http://www.insidegnss.com/auto/marapr09-borre.pdf>. Accessed 7th April, 2019. 313
- Claude Boucher and Zuheir Altamimi. Memo : Specifications for reference frame fixing in the analysis of a EUREF GPS campaign, 2007. URL <http://users.auth.gr/kvek/20070327-MEMO-ver6.pdf>. Accessed 11th March, 2019. 281
- Nathaniel Bowditch, LL. D. *American Practical Navigator*. National Geospatial-Intelligence Agency, Springfield, Virginia, 2017. URL https://msi.nga.mil/NGAPortal/MSI.portal?_nfpb=true&_pageLabel=msi_portal_page_62&pubCode=0002. Accessed 14th May, 2019. 157, 170
- Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer Science and Business Media, New York, 2013. ISBN 978-1-4757-2917-7. 395



C

- Bradley Darren Carter, Michael C. B. Ashley, Yin-Sheng Sun, and John W. V. Storey. Redesigning a Baker-Nunn Camera for CCD imaging. *Proceedings Astronomical Society of Australia*, (1):74–76, 1992. URL <https://newt.phys.unsw.edu.au/~mcba/pubs/carter92.pdf>. Accessed 14th May, 2019. 465
- CHAMP Mission. CHAMP – CHallenging Minisatellite Payload. Deutsches Geoforschungszentrum, Helmholtz-Zentrum Potsdam. URL <https://www.gfz-potsdam.de/champ/>. Accessed 12th April, 2019. 494
- Roger Cicala. Iceland spar: The rock that discovered optics. *Lensrentals*, 11th December 2013. URL <https://www.lensrentals.com/blog/2013/12/iceland-spar-the-rock-that-discovered-optics/>. Accessed 8th April, 2019. 191
- Michael Alan Ralph Cooper. *Modern Theodolites and Levels*. Granada Publishing Ltd, London, second edition, 1982. ISBN 978-0-2461-1502-7. 170
- Michael Alan Ralph Cooper. *Control Surveys in Civil Engineering*. Collins, London, 1987. ISBN 978-0-8939-7272-1. 418, 420, 423
- Nicholas Crane. *Mercator – The Man who Mapped the Planet*. Weidenfeld and Nicholson, London, 2002. ISBN 978-0-8050-6624-1. 3
- Wesley G. Crawford, RPLS. Back to basics: Quick setup with a laser plummet. *Point of Beginning*, 1st January 2009. URL <https://www.pobonline.com/articles/93629-back-to-basics-quick-setup-with-a-laser-plummet>. Accessed 14th May, 2019. 134
- Arpád Czobor and Zsuzsanna Németh. Balloon-borne beacons for stellar triangulation. *Advances in Space Research*, 1(11):29–34, 1981. URL [https://doi.org/10.1016/0273-1177\(81\)90445-2](https://doi.org/10.1016/0273-1177(81)90445-2). Accessed 14th May, 2019. 462

D

- Sönke Dangendorf, Marta Marcos, Guy Wöppelmann, Clinton P. Conrad, Thomas Frederikse, and Riccardo Riva. Reassessment of 20th century global mean sea level rise. *Proceedings of the National Academy of Sciences*, 114(23): 5946–5951, 2017. URL <https://doi.org/10.1073/pnas.1616007114>. Accessed 14th May, 2019. 509
- Albert Dastre. *Questions scientifiques - L'Heure légale*, volume 4e période, tome 148. 1898. URL https://fr.wikisource.org/wiki/Questions_scientifiques_-_L%E2%80%99Heure_l%C3%A9gale/01. Accessed 4th March, 2020. 53
- Charles DeMets, Richard G. Gordon, Donald F. Argus, and Seth Stein. Effect of recent revisions to the geomagnetic reversal time scale on estimates of



current plate motions. *Geophysical Research Letters*, 21(20):2191–2194, 1994. URL <https://doi.org/10.1029/94GL02118>. Accessed 14th May, 2019. 487

Deutsche Fotothek. Rechenbrett, Vermessung des Inhalts eines Fasses, 1523. URL <http://www.deutschefotothek.de/documents/obj/88963434>. Accessed 18th November, 2019. 251

John Louis Emil Dreyer. The Well of Eratosthenes. *The Observatory*, 37: 352–353, 1914. URL <http://articles.adsabs.harvard.edu/full/1914Obs...37..352D/0000352.000.html>. Accessed 14th May, 2019. 3

DuPont Teijin Films. Mylar polyester film, Safety in Handling and Use. URL https://web.archive.org/web/20201027150459/https://stenbacka.fi/wp-content/uploads/sites/3/2016/07/mylar_a_perustieto.pdf. Accessed 19th November, 2021. 464

E

Valerie J. Easton and John H. McColl. STEPS Statistics Glossary v1.1. URL http://www.stats.gla.ac.uk/steps/glossary/probability_distributions.html. Accessed 14th April, 2019. 36

R. Edgeworth, B. J. Dalton, and Thomas Parnell. The pitch drop experiment. *European Journal of Physics*, pages 198–200, 1984. URL <https://smp.uq.edu.au/files/109/outreach-pitch-drop-abstract.pdf>. Accessed 12th August, 2019. 489

Jürgen Ehlers, Philip Hughes, and Philip L. Gibbard. *The Ice Age*. Wiley-Blackwell, 2015. ISBN 978-1-118-50781-0. 491

Eiffel Tower, 72 names. List of the 72 names on the Eiffel Tower. URL https://en.wikipedia.org/wiki/List_of_the_72_names_on_the_Eiffel_Tower. Accessed 7th April, 2019. 194, 330, 362, 453, 500, 508

Anon. *Elements of surveying*. Headquarters, Department of the Army, 1971. URL <https://play.google.com/books/reader?id=564XAAAAYAAJ>. Accessed 14th March, 2019. 138

Loránd Eötvös. *Three Fundamental Papers of Loránd Eötvös*. Loránd Eötvös Geophysical Institute of Hungary, 1998. ISBN 963-7135-02-2. Ed. Zoltán Szabó. 452

EUREF. EUREF, Reference Frame Sub Commission for Europe. URL <http://www.euref.eu/>. Accessed 14th April, 2019. 57



F

- Eric Fielding. Introduction to SAR Interferometry, 2017. URL https://appliedsciences.nasa.gov/sites/default/files/Session4-SAR-English_0.pdf. Accessed 22nd September, 2021. 494
- Finnish National Land Survey, About positioning services. URL <https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/positioning-services>. Accessed 10th April, 2019. 353
- Finnish National Land Survey, Elevation model 10 m. URL <https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/product-descriptions/elevation-model-10-m>. Accessed 9th April, 2019. 252
- Finnish National Land Survey, Elevation model 2 m. URL <https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/product-descriptions/elevation-model-2-m>. Accessed 9th April, 2019. 252
- Finnish National Land Survey, Paikkatietoikkuna. URL <https://kartta.paikkatietoikkuna.fi/?lang=en>. Accessed 29th September, 2021. 86, 87
- Finnish National Land Survey, RINEX service. URL <https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/positioning-services/rinex-palvelu>. Accessed 10th April, 2019. 354
- Octavi Fors, Jorge Núñez, José Luis Muiños, Francisco Javier Montojo, Roberto Baena-Gallé, Jaime Boloix, Ricardo Morcillo, María Teresa Merino, Elwood C. Downey, and Michael J. Mazur. Telescope Fabra ROA Montsec: A new robotic wide field Baker-Nunn facility. *Publications of the Astronomical Society of the Pacific*, 125(927):522–538, May 2013. URL <https://doi.org/10.1086/670941>. 465
- Clemens Franz. Wikimedia Commons, Der Meridianstein in Hammerfest, Norwegen, 2005. URL https://commons.wikimedia.org/wiki/File:Hammerfest_Meridianstein.jpg. © 2005 Clemens Franz (CC BY-SA 3.0). Accessed 13th April, 2019. 10
- Joshua Fraser, Anoop Haridas, Guna Seetharaman, Raghuveer M. Rao, and Kannappan Palaniappan. KOLAM: a cross-platform architecture for scalable visualization and tracking in wide-area imagery. *Proceedings of the SPIE*, 8747, 2013. URL <https://doi.org/10.1117/12.2018162>. Accessed 17th September, 2021. 255
- Jean-Pierre Friedelmeyer. Du côté des lettres (2) : une lettre de Sophie Germain à Carl Friedrich Gauss (20 février 1807), et la réponse de celui-ci (30 avril



1807), 2014. URL <https://images.math.cnrs.fr/Du-cote-des-lettres-une-lettre-de-Sophie-Germain-a-Carl-Friedrich-Gauss-20>. Accessed 14th May, 2019. **357**

G

GDGPS APPS. Automatic Precise Positioning Service of the Global Differential GPS System. URL https://apps.gdgps.net/apps_file_upload.php. Accessed 10th April, 2019. **355**

Geotrim, Trimnet. URL <https://geotrim.fi/palvelut/trimnet-vrs/>. Accessed 1st March, 2020. **352**

GLOBE Task Team and others. The Global Land One-km Base Elevation Project, 1999. URL <https://www.ngdc.noaa.gov/mgg/topo/globe.html>. Accessed 9th April, 2019. **252**

GRACE Follow-On Mission. Jet Propulsion Laboratory. URL <https://gracefo.jpl.nasa.gov/mission/overview/>. Accessed 12th April, 2019. **496**

GRACE Mission. Measuring Earth's Surface Mass and Water Changes. Jet Propulsion Laboratory. URL <https://grace.jpl.nasa.gov/>. Accessed 12th April, 2019. **496**

GRACE Mission, hydrology. NASA. URL https://commons.wikimedia.org/wiki/File:Global_Gravity_Anomaly_Animation_over_LAND.gif. Accessed 12th April, 2019. **496**

Erik W. Grafarend, Horst Kremers, Juhani Kakkuri, and Martin Vermeer. Modelling vertical refraction in the SW Finland Triangular Network TAGNET adjustment. Report 87:2, Finnish Geodetic Institute, Helsinki, 1987. **180, 182**

Erik W. Grafarend, Rey-Jer You, and Rainer Syffus. *Map Projections: Cartographic Information Systems*. Springer Verlag, second edition, 2014. ISBN 978-3642364938. **55**

Richard S. Gross. The excitation of the Chandler wobble. *Geophysical Research Letters*, 27:2329–2332, 2000. URL <https://doi.org/10.1029/2000GL011450>. Accessed 14th May, 2019. **466**

Thomas Gruber, Reiner Rummel, Johannes Ihde, Gunter Liebsch, Axel Rülke, Uwe Schäfer, Michael Sideris, Elena Rangelova, and Philip Woodworth. Scientific Roadmap towards Height System Unification with GOCE. Slide show, GOCE 5th International GOCE User Workshop, Paris, 2014. URL <https://slideplayer.com/slide/10319716/>. Accessed 17th September, 2021. **95**



- GUM. Evaluation of measurement data – Guide to the expression of uncertainty in measurement. Report JGCM 100:2008, Working Group 1, Joint Committee for Guides in Metrology, 2008. URL https://www.bipm.org/utis/common/documents/jcgm/JCGM_100_2008_E.pdf. Accessed 17th August, 2019. 30
- Werner Gurtner and Lou Estey. RINEX, The Receiver Independent Exchange Format v. 3.00. 28th November 2007. URL <https://epic.awi.de/id/eprint/29985/1/Gur2007a.pdf>. Accessed 9th April, 2019. 302, 353
- ## H
- Pasi Häkli, Jyrki Puupponen, Hannu Koivula, and Markku Poutanen. Suomen geodeettiset koordinaatistot ja niiden väliset muunnokset. Research note (in Finnish) 30, Finnish Geodetic Institute, Masala, 2009. URL <https://www.maamittauslaitos.fi/sites/maanmittauslaitos.fi/files/fgi/GLtiedote30.pdf>. Accessed 14th May, 2019. 57, 218, 280
- Jonathan Hardis. Why is a second defined as "the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom."? Why not 9,192,632,000 or other numbers? Quora answer, 2018. URL <https://www.quora.com/Why-is-a-second-defined-as-the-duration-of-9-192-631-770-periods-of-the-radiation-corresponding-to-the-transition-between-the-two-hyperfine-levels-of-the-ground-state-of-the-caesium-133-atom-Why-not-9-192-632-000-or-other-numbers>. Accessed 8th April, 2019. 26
- Paul Harnesk and Åke Davidsson. *Vem är Vem? Stor-Stockholm 1962*. Project Runeberg, 1962. URL <http://runeberg.org/vemarvem/sthlm62/0152.html>. Accessed 17th November, 2019. 185
- Markku Heikkinen. Solving the shape of the Earth by using digital density models. Report 81:2, Finnish Geodetic Institute, Helsinki, 1981. 431
- Veikko A. Heiskanen. *Kenttämittaus ja kartoitus*. Werner Söderström Oy, second, expanded edition, 1943. 140
- Veikko A. Heiskanen and Seppo Härmälä. *Maastomittaus ja kartoitus*. Werner Söderström Oy, fifth, unchanged edition, 1963. 16, 18, 95
- Weikko A. Heiskanen and Helmut Moritz. *Physical Geodesy*. W. H. Freeman and Company, San Francisco, 1967. i, 13, 267, 435, 437, 444
- Weikko A. Heiskanen and Felix A. Vening Meinesz. *The Earth and its Gravity Field*. McGraw-Hill, New York, 1958. 435



- Hexagon Geosystems, HxGN SmartNet. URL <https://www.hxgnsmartnet.com>. Accessed 4th March, 2020. **353**
- Nicolàs de Hilster. Geodetic instruments. URL http://dehilster.info/geodetic_instruments/1980_sat-sagem_AGA-Minilir.php. © Nicolàs de Hilster. Accessed 9th April, 2019. **248**
- Bernard Hofmann-Wellenhof, Herbert Lichtenegger, and James Collins. *GPS Theory and Practice*. Springer-Verlag, Vienna, fifth corrected edition, 2001. ISBN 978-3-2118-3534-0. **i, 283, 298, 326, 328**
- Humboldt University Berlin. Friedrich Robert Helmert, 2017. URL https://www.researchgate.net/publication/318994932_Friedrich_Robert_Helmert_founder_of_modern_geodesy_on_the_occasion_of_the_centenary_of_his_death. © 2017 Humboldt-Universität zu Berlin, Universitätsbibliothek (CC BY 3.0). Accessed 19th May, 2019. **72**

I

- Anon. IAG Resolutions adopted at the XXIIth General Assembly in Birmingham, 1999. URL https://iag.dgfi.tum.de/fileadmin/IAG-docs/IAG_Resolutions_1999.pdf. Resolution 3. Accessed 14th May, 2019. **202**
- IGS Central Bureau. URL <https://igs.org/governance-management#central-bureau>. Accessed 17th September, 2021. **331**
- IMCCE, Milanković app. IMCCE, Paris Observatory. URL <http://vo.imcce.fr/insola/earth/online/earth/online/index.php>. Accessed 12th April, 2019. **504**
- INSPIRE Knowledge Base. European Commission. URL <https://inspire.ec.europa.eu>. Accessed 9th April, 2019. **252**
- The Institute of Australian Culture. The Old Survey [song, 1905], 2012. URL <http://www.australianculture.org/the-old-survey-1905/>. Accessed 19th August, 2019. **119**
- International DORIS Service. URL <https://ids-doris.org/>. Accessed 12th April, 2019. **480**
- IPCC. Climate Change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In Susan Solomon, Dahe Qin, Martin R. Manning, Zhenlin Chen, Melinda Marquis, Kirsten B. Averyt, Melinda M. B. Tignor, and Henry L. Miller, editors, *Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge Univ. Press, Cambridge UK, 2007. URL <https://www.ipcc.ch/report/ar4/wg1/>. **512**



IUGG, GGOS. Global Geodetic Observing System. URL <https://ggos.org>. Accessed 9th April, 2019. 265

J

Haseeb Jamal. Compound curves in engineering survey, 2017. URL <https://www.aboutcivil.org/compound-curves-in-engineering-surveying.html>. Accessed 14th May, 2019. 241, 242

Harold Jeffreys. *The Theory of Probability*. Oxford Classic Texts in the Physical Sciences. Oxford University Press, 1998. ISBN 978-0-1915-8967-6. Originally published in © 1939. 395

Jean Jouzel and Valerie Masson-Delmotte. EPICA Dome C Ice Core 800KYr deuterium data and temperature estimates. IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2007-091. NOAA/NCDC Paleoclimatology Program, Boulder CO, USA., 2007. URL <https://doi.pangaea.de/10.1594/PANGAEA.683655>. Accessed 11th March, 2019. 504

Jean Jouzel, Valerie Masson-Delmotte, Olivier Cattani, Gabrielle B. Dreyfus, Sonia Falourd, Georg Hoffmann, Bénédicte Minster, Julius Nouet, Jean-Marc Barnola, Jérôme Chappellaz, Hubertus Fischer, Jean-Charles Gallet, Sigfus Johnsen, Markus Leuenberger, Laetitia Loulergue, Dieter Luethi, Frederic Parrenin Hans Oerter, Grant M. Raisbeck, Dominique Raynaud, Adrian Schilt, Jakob Schwander, Enricomaria Selmo, Roland Souchez, Renato Spahni, Bernhard Stauffer, Jørgen Peder Steffensen, Barbara Stenni, Thomas F. Stocker, Jean-Louis Tison, Martin Werner, and Eric W. Wolff. Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, 317(5839):793–796, 2007. URL <https://doi.org/10.1126/science.1141038>. Accessed 4th March, 2020. 504

JPL, The Global Differential GPS System. URL <https://www.gdgps.net>. Accessed 9th April, 2019. 353

JUHTA. JHS 163 Finland's height system N2000. JHS Recommendation, Advisory Committee on Information Management in Public Administration, 2010. URL <https://www.suomidigi.fi/ohjeet-ja-tuki/jhs-suositukset/jhs-163-suomen-korkeusjarjestelma-n2000>. 213

JUHTA. JHS 178 Interface of geographic information services for local authorities. JHS Recommendation, Advisory Committee on Information Management in Public Administration, 2012a. URL <https://www.suomidigi.fi/ohjeet-ja-tuki/jhs-suositukset/jhs-178-kunnan-paikkatietopalvelurajapinta-vanhentunut>. 213



- JUHTA. JHS 184 Point measurement in the EUREF-FIN coordinate reference system. JHS Recommendation, Advisory Committee on Information Management in Public Administration, 2012b. URL <https://www.suomidigi.fi/ohjeet-ja-tuki/jhs-suositukset/jhs-184-kiintopistemittaus-euref-fin-koordinaattijarjestelmassa>. 213, 216, 218
- JUHTA. JHS 185 Composing base maps for the city. JHS Recommendation, Advisory Committee on Information Management in Public Administration, 2014. URL <https://www.suomidigi.fi/ohjeet-ja-tuki/jhs-suositukset/jhs-185-asemakaavan-pohjakartan-laatiminen>. 213, 235
- JUHTA. JHS 196 EUREF-FIN coordinates in Finland. JHS Recommendation, Advisory Committee on Information Management in Public Administration, 2016a. URL <https://www.suomidigi.fi/ohjeet-ja-tuki/jhs-suositukset/jhs-196-euref-fin-jarjestelman-mukaiset-koordinaatit-suomessa>. 213, 217
- JUHTA. JHS 197 EUREF-FIN coordinate systems, related conversions and map sheet distribution. JHS Recommendation, Advisory Committee on Information Management in Public Administration, 2016b. URL <https://www.suomidigi.fi/ohjeet-ja-tuki/jhs-suositukset/jhs-197-euref-fin-koordinaattijarjestelmat-niihin-liittyvat-muunnokset-ja-karttalehtijako>. 61, 63, 213
- K**
- Erkki Kääriäinen. The second levelling of Finland in 1935-1955. Publication 61, Finnish Geodetic Institute, Helsinki, 1966. 112
- Anon. Kaavoitusmittausohjeet 2003. Publication 94, Finnish National Land Survey, 31st January, 2003. URL https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/old/kaavoitusmittausohjeet_2003_0.pdf. Accessed 22nd September, 2021. 213
- Heribert Kahmen and Wolfgang Faig. *Surveying*. Walter de Gruyter, Berlin – New York, 1988. ISBN 978-3-1100-8303-3. i, 27, 106, 124, 134, 140, 148, 180, 187, 222, 224, 256, 260, 390
- Juhani Kakkuri. *Stellar Triangulation with Balloon-borne Beacons and Satellites*. PhD thesis, University of Helsinki, 1973. 462
- Juhani Kakkuri, Teuvo Parm, Vidal Ashkenazi, and S. A. Crane. Accuracy analysis of the Finnish laser geodimeter traverse. In *Proc., Int. Symp. on Geodetic Networks and Computations*, volume 5, pages 85–93, Munich, 1981. Verlag der Bayerischen Akademie der Wissenschaften. 206
- Ulla Kallio. *Tasoituslasku*. TKY 587. Otatieto, third edition, 1998. ISBN 978-9-5167-2267-5. i



- Kari-Pekka Karlsson. Maastotietojärjestelmä 2030 -selvitys. Final report, Ministry of Agriculture and Forestry, 2015. URL http://mmm.fi/documents/1410837/1801204/Maastotietojarjestelma_2030_selvitys_Loppuraportti_28.2.2015.pdf. Accessed 24th November, 2021. 231
- Charles Karney. Online Geoid Calculator. URL <https://geographiclib.sourceforge.io/cgi-bin/GeoidEval>. Accessed 14th April, 2019. 118
- Shfaqat Abbas Khan, Lin Liu, John Wahr, Ian Howat, Ian Joughin, Tonie van Dam, and Kevin Fleming. GPS measurements of crustal uplift near Jakobshavn Isbræ due to glacial ice mass loss. *Journal of Geophysical Research: Solid Earth*, 115(B9):B09405, 2010. URL <http://dx.doi.org/10.1029/2010JB007490>. Accessed 14th May, 2019. 506
- Hannu Koivula and Jaakko Mäkinen. Geodetic activities at Finnish Antarctic research station Aboa. In *Fifth International Antarctic Geodesy Symposium AGS'03, Lviv, Ukraine, Sept. 15–17, 2003*. URL https://www.researchgate.net/publication/237467751_GEODETTIC_ACTIVITIES_AT_FINNISH_ANTARCTIC_RESEARCH_STATION_ABOA. Accessed 14th May, 2019. 506
- Hannu Koivula, Jaakko Kuokkanen, Simo Marila, Sonja Lahtinen, and Tuukka Mattila. Assessment of sparse GNSS network for network RTK. *Journal of Geodetic Science*, 8(1):136–144, 2018. URL <https://doi.org/10.1515/jogs-2018-0014>. Accessed 15th August, 2019. 351
- Yoshihide Kozai. The motion of a close Earth satellite. *Astronomical Journal*, 64(1274):367–377, 1959. URL <https://doi.org/10.1086/107957>. Accessed 14th May, 2019. 328
- T. J. Kukkamäki and Pekka Lehmoskoski. Influence of the earth magnetic field on Zeiss Ni 002 levels. Report 84:1, Finnish Geodetic Institute, Helsinki, 1984. 106

L

- Jacques Laskar, Philippe Robutel, Frédéric Joutel, Mickael Gastineau, Alexandre C. M. Correia, and Benjamin Levrard. A long term numerical solution for the insolation quantities of the Earth. *Astronomy and Astrophysics*, 428:261–285, 2004. URL <https://doi.org/10.1051/0004-6361:20041335>. Accessed 14th May, 2019. 504
- Javier Arribas Lázaro. *GNSS Array-based Acquisition: Theory and Implementation*. PhD thesis, Universitat Politècnica de Catalunya, 2012. URL <https://theses.eurasip.org/theses/449/gnss-array-based-acquisition-theory-and/download/>. Accessed 14th May, 2019. 300



Rachel Leah. Rapper B.o.B. starts GoFundMe to prove earth is flat. *Salon*, 25th September, 2017. URL <https://www.salon.com/2017/09/25/b-o-b-flat-earth/>. Accessed 14th May, 2019. 20

Leica. *TPS - System 1000 – Quick Start*. Leica Geosystems AG, Heerbrugg, Switzerland, 1997. URL http://www.engineeringsurveyor.com/software/leica/tps2000_quick_start_en.pdf. Accessed 1st December, 2021. 152

Cherry L. E. Lewis. *The Dating Game*. Cambridge University Press, 2000. ISBN 978-1-1391-9713-7. 486

Martin Lidberg, Jan M. Johansson, Hans-Georg Scherneck, and Glenn A. Milne. Recent results based on continuous GPS observations of the GIA process in Fennoscandia from BIFROST. *Journal of Geodynamics*, 50(1):8–18, July 2009. URL <https://doi.org/10.1016/j.jog.2009.11.010>. Accessed 14th May, 2019. 491

M

Jaakko Mäkinen, Andreas Engfeldt, Linda Engman, Bjørn Geir Harsson, Tõnis Oja, Sven Rekkedal, Knut Røthing, Paavo Rouhiainen, Hannu Ruotsalainen, H. Skatt, Gabriel Strykowski, Heikki Virtanen, Karin Wiczerkowski, and Detlef Wolf. The Fennoscandian Land Uplift Gravity Lines: Comparison of observed gravity change with observed vertical motion and with GIA models. Report, *Nordiska Kommissionen för Geodesi*, 2010. URL https://www.nordicgeodeticcommission.com/wp-content/uploads/2014/10/1-Makinen_et_al_land_uplift_gravity_lines.pdf. Accessed 11th March, 2019. 492

Stephen Malys, John H. Seago, Nikolaos K. Pavlis, P. Kenneth Seidelmann, and George H. Kaplan. Why the Greenwich meridian moved. *Journal of Geodesy*, 89(12):1263–1272, 2015. URL <https://doi.org/10.1007/s00190-015-0844-y>. Accessed 14th May, 2019.

Stephen Malys, Robert Wong, and Scott A. True. The WGS84 Terrestrial Reference Frame in 2016. Presented at eleventh meeting of the International Committee on GNSS, November 2016, Sochi, Russia, 2016. URL <https://www.unoosa.org/pdf/icg/2016/icg11/wgd/02wgd.pdf>. Accessed 17th September, 2021. 58

Pierre Louis Moreau de Maupertuis. *La Figure de la Terre, Déterminée par les observations. . . .* Jean Catuffe, Amsterdam, 1738. URL <https://www.doria.fi/bitstream/handle/10024/69470/001.pdf>. Accessed 7th April, 2019. 1, 8

Ministry of the Environment, Legislation on land use and building. URL <https://ym.fi/en/land-use-and-building-act>. Accessed 17th September, 2021. 235



- Patrap Misra and Per Enge. *Global Positioning System – Signals, Measurements, and Performance*. Ganga-Jamuna Press, revised second edition, 2010. ISBN 978-0-9709-5442-8. 289
- Soumen Mondal, K. G. Gupta, Sneha Lata, Biman J. Medhi, Tarun Bangia, T. S. Kumar, Shobhit Yadav, and S. K. Singh. Development of ARIES Baker-Nunn camera to a wide-field Imaging Telescope with CCD. *arXiv*, 2009. URL <https://arxiv.org/pdf/0905.0361.pdf>. Accessed 14th May, 2019. 465
- Oliver Montenbruck, Peter Steigenberger, and André Hauschild. Broadcast versus precise ephemerides: A multi-GNSS perspective. *GPS Solutions*, 19(2): 321–333, 2015. URL <https://doi.org/10.1007/s10291-014-0390-8>. Accessed 14th May, 2019. 341
- Walter Heinrich Munk. The U.S. Commission on Ocean Policy, Testimony, 12th April, 2002. URL https://govinfo.library.unt.edu/oceancommission/meetings/apr18_19_02/munk_statement.pdf. Accessed 14th May, 2019. 479
- N**
- NASA JPL, TOPEX/Poseidon, Jason. URL <https://sealevel.jpl.nasa.gov/>. Accessed 12th April, 2019. 509
- NASA Solar Dynamics Observatory, Atmospheric Imaging Assembly. URL https://sdo.gsfc.nasa.gov/assets/img/browse/2013/07/15/20130715_001424_4096_0171.jpg. Accessed 19th May, 2019. 472
- NGS, Computation utilities. URL https://geodesy.noaa.gov/TOOLS/Inv_Fwd/Inv_Fwd.html. Accessed 14th September, 2020. 68, 86
- NOAA et al. ETOPO1 Global Relief Model, 2008. URL <http://dx.doi.org/10.7289/V5C8276M>. Accessed 14th May, 2019. 252
- NOAA, Ocean currents. How does the ocean affect climate and weather on land? URL <https://oceanexplorer.noaa.gov/facts/climate.html>. © NOAA. Accessed 12th November, 2019. 499
- O**
- GNU Octave. URL <https://www.gnu.org/software/octave/>. Accessed 13th April, 2019. 333
- OGC, Catalogue Service. Open Geospatial Consortium. URL <https://www.ogc.org/standards/cat>. Accessed 9th April, 2019. 231
- Victor Rafael Ölander. Trigonometriset korkeusmittaukset pitkillä välimatkoilla ja niiden tarkkuus. *Maanmittaus*, 1932. 182



- Matti Ollikainen. GPS-koordinaattien muuntaminen kartastokoordinaateiksi. Research note (in Finnish) 8, Finnish Geodetic Institute, Helsinki, 1993. 279, 280
- Boudewijn Overgaauw, Boudewijn A. C. Ambrosius, and Karel F. Wakker. Analysis of the EUREF-89 GPS data from the SLR/VLBI sites. *Bulletin Géodésique*, 68:19–28, 1994. URL <https://doi.org/10.1007/BF00806749>. Accessed 14th May, 2019. 57
- P**
- Teuvo Parm. Kansallisen koordinaattijärjestelmän luominen Suomessa. *Maanmittaus*, 63(1), 1988. 56
- Jerry Penry and David Lee Ingram. The American theodolite. *The American Surveyor*, 10(9), 2013. URL https://archive.amerisurv.com/PDF/TheAmericanSurveyor_PenryIngram-TheAmericanTheodolite_Vol10No9.pdf. Accessed 22nd September, 2021. 137
- William B. Penzes. Time Line for the Definition of the Meter, undated. URL <https://www.nist.gov/system/files/documents/pml/div683/museum-timeline.pdf>. Accessed 17th September, 2021. 26
- Mika Pitkä. Sataman tehokkuuden osatekijät – tarkastelussa lastauksen ja purun automatisointi. Bachelor's thesis, Kymenlaakso University of Applied Sciences, 2009. 247
- Porvoo. Koordinaatti- ja korkeusjärjestelmä. URL <https://www.porvoo.fi/koordinaatti-ja-korkeusjarjestelma>. © City of Porvoo. Accessed 26th November, 2019. 392
- Robert Pound and Glen A. Rebka, Jr. Gravitational red-shift in nuclear resonance. *Physical Review Letters*, 3(9):439–441, 1959. URL <https://doi.org/10.1103/PhysRevLett.3.439>. Accessed 14th May, 2019. 190
- Markku Poutanen. *GPS-paikanmääritys*. Ursa, Helsinki, 1998. ISBN 951-9269-89-4. Ursan julkaisuja 64. i
- Markku Poutanen. *Satelliittipaikannus*. Ursa, Helsinki, 2017. ISBN 978-9-5259-8541-2. i, 283, 298, 326, 328
- Jyrki Puupponen and Jaakko Järvinen. Historical video documentation of triangulation measurements in Finland. In *Integrating Generations, FIG Working Week*, Stockholm, Sweden, 14th – 19th June, 2008. URL https://www.fig.net/resources/proceedings/fig_proceedings/fig2008/papers/hs05/hs05_02_puupponen_jarvinen_2934.pdf. Accessed 8th April, 2019. 57



R

RAKLI ry. Kiinteistöalan yhteiskunnallinen ja kansantaloudellinen merkitys. Report, Finnish Association of Building Owners and Construction Clients RAKLI, 2014. URL <https://kti.fi/wp-content/uploads/Kiinteist%C3%B6alan-yhteiskunnallinen-ja-kansantaloudellinen-merkitys-2014.pdf>. Accessed 17th September, 2021. 16

Reginald Reagan. One Matrix, Inverted, 26th July 2021. URL <https://raginrayguns.tumblr.com/post/657824795686207489/david-salsburg-the-lady-tasting-tea-during>. Accessed 17th September, 2021. 517

Jan Reijnoudt. Het hele IJsselmeer waterpas, 1996. URL https://www.digibron.nl/viewer/collectie/Digibron/id/tag:RD.nl,19960213:newsm1_9f3bdeec998e02ec9e49004d5e149c1661831c48735609a683bd383e. Accessed 22nd September, 2021. 95

ROB, EUREF Permanent GNSS Network. Royal Observatory of Belgium. URL http://www.epncb.oma.be/_networkdata/stationlist.php. Accessed 10th April, 2019. 86, 354

Rovaniemi, The Degree Measurement Expedition. The degree measurements by de Maupertuis in the Tornionlaakso valley 1736–1737. City of Rovaniemi. URL http://lapinkavijat.rovaniemi.fi/maupertuis/index_eng.html. Accessed 16th May, 2019. 8

The R Project for Statistical Computing. R Foundation. URL <https://www.r-project.org/>. Accessed 13th April, 2019. 333

Jean M. Rüeger. *Electronic Distance Measurement – An Introduction*. Springer-Verlag, Berlin Heidelberg, fourth edition, 1996. 202, 204, 205, 307

Jean M. Rüeger. Refractive indices of light, infrared and radio waves in the atmosphere. *UNISURV report S 68*, School of Surveying and Spatial Information Systems, University of New South Wales, 2002. URL https://www.sage.unsw.edu.au/sites/sage/files/SAGE_collection/SpecialSeries/s68.pdf. Accessed 4th October, 2021. 204, 307

Reiner Rummel and Fernando Sansò, editors. *Satellite Altimetry in Geodesy and Oceanography*, volume 50 of *Lecture Notes in Earth Sciences*, 1993. Springer-Verlag. 506

S

Hannu Salmenperä. Runko- ja kartoitusmittaukset. Report 4, Tampere University of Technology, Rakennustekniikan osasto, Geoinformatiikan laboratorio, Tampere, 1998. i, 213, 234



- Sanastokeskus TSK. Geoinformatiikan sanasto. TSK 51, Finnish National Land Survey, 2018. URL <http://www.tsk.fi/tiedostot/pdf/GeoinformatiikanSanasto.pdf>. Accessed 14th May, 2019. 214
- Gavin Schrock. CERN. xyHt *magazine*, 2014. URL <https://www.xyht.com/civiltransportation/cern-2/>. Accessed 2nd July, 2019. 114
- Scilab Enterprises. URL <https://www.scilab.org/>. Accessed 13th April, 2019. 333
- Sea Level Research Group. University of Colorado. URL <https://sealevel.colorado.edu/>. Accessed 12th April, 2019. 510
- Günter Seeber. *Satellitengeodäsie. Grundlagen, Methoden und Anwendungen*. De Gruyter, Berlin New York, 1989. ISBN 978-3-1101-0082-2. 306
- Shuttle Radar Topography Mission, 2014. URL <https://www2.jpl.nasa.gov/srtm/>. Accessed 22nd September, 2021. 252
- Juho Simonen. Maalaserkeilaus infrarakentamisen mittauksissa. Master's thesis, Aalto-yliopisto, 2012. 138
- Dava Sobel. *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. Walker and Company, New York, 1998. ISBN 978-1-8570-2571-2. Paperback. 461, 462
- SOHO images. ESA and NASA. URL <https://sohowww.nascom.nasa.gov/data/realtime-images.html>. Accessed 5th March, 2020. 473
- SOHO MPEG movies. ESA and NASA. URL <https://sohowww.nascom.nasa.gov/data/realtime/mpeg/>. Accessed 5th March, 2020. 473
- SOPAC, GPS Explorer. Scripps Orbit and Permanent Array Center. URL <http://sopac.ucsd.edu/dataBrowser.shtml>. Accessed 5th March, 2020. 354
- SOPAC, Hatanaka file compression. Scripps Orbit and Permanent Array Center. URL <http://sopac.ucsd.edu/hatanaka.shtml>. Accessed 5th March, 2020. 354
- Govert Strang van Hees. Stokes' formula using fast Fourier techniques. *Manuscripta geodaetica*, 15:235–239, 1990. URL https://doi.org/10.1007/978-1-4612-3104-2_47. Condensed version. Accessed 14th May, 2019. 14



T

- Mikko Takalo. Computer-supported Finnish EDM height traversing. Part I: technical description. *Surveying Science in Finland*, 13(2):56–99, 1995. **95, 182**
- Mikko Takalo and Paavo Rouhiainen. On system calibration of digital level. In *Ingenieurvermessung 2004, 14th International Conference on Engineering Surveying*, Zürich, 15th – 19th March, 2004. URL https://www.researchgate.net/publication/254398461_Development_of_a_System_Calibration_Comparator_for_Digital_Levels_in_Finland. Accessed 22nd September, 2021. **108**
- Mikko Takalo, Paavo Rouhiainen, Pekka Lehmuskoski, and Veikko Saaranen. On calibration of Zeiss DiNi12. In *FIG Working Week 2001*, Seoul, Korea, 6th – 11th May, 2001. FIG Commission 5. URL https://www.researchgate.net/publication/237495166_ON_CALIBRATION_OF_ZEISS_DINI12. Accessed 12th March, 2019. **108**
- Martti Tikka. *Käytännön geodesia III. Mittausmenetelmät*. Otakustantamo, Espoo, 1985. **i**
- Martti Tikka. *Käytännön geodesia I. Mittaustekniikan perusteet ja rakennusteknilliset sovellutukset*. Otatieto, Helsinki, third revised edition, 1991. **i**
- Waldo R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods, 1970. URL <https://www.jstor.org/stable/143141>. Accessed 22nd September, 2021. **335**
- Wolfgang Torge. *Geodesy*. Walter de Gruyter, Berlin – New York, third completely revised and extended edition, 2001. ISBN 978-3-11-087995-7. **i, 3, 14, 15**
- U
- UNAVCO, EGM96 geoid calculator. URL <https://www.unavco.org/software/geodetic-utilities/geoid-height-calculator/geoid-height-calculator.html>. Accessed 21st September, 2021. **118**
- UNAVCO, plate-motion calculator. URL <https://www.unavco.org/software/geodetic-utilities/plate-motion-calculator/plate-motion-calculator.html/models>. Accessed 5th March, 2020. **488**
- USGS, GTOPO30. URL <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30?qt->



[science_center_objects=0#qt-science_center_objects](#). Accessed 9th April, 2019. 252

V

Yrjö Väisälä. Maan toinen kuu. *Tähtitaivas*, 6, 1946. 462

Peter Vaníček and Edward Krakiwsky. *Geodesy – The Concepts*. Elsevier Science Publishers, Amsterdam, second edition, 1986. ISBN 978-0-4448-6149-8. i, 418

Vasamuseet, Deformation Monitoring. URL <https://www.vasamuseet.se/en/research/support-vasa/deformation-monitoring>. Accessed 22nd September, 2021. 114

Väylä, Radionavigaatiopalvelut. URL <https://vayla.fi/palveluntuottajat/ammattimerenkulku/liikkuminen-vesivaylilla/radionavigaatiopalvelut>. Accessed 22nd September, 2021. 352

Martin Vermeer. A new SEASAT altimetric geoid for the Baltic. Report 83:4, Finnish Geodetic Institute, Helsinki, 1983. 511

David Vine. *Island of Shame: The Secret History of the U.S. Military Base on Diego Garcia*. Princeton University Press, revised edition, 2011. ISBN 978-0-6911-4983-7. URL <https://press.princeton.edu/books/paperback/9780691149837/island-of-shame>. Accessed 22nd September, 2021. 300, 302

W

Paul Wessel, W. H. F. Smith, Remco Scharroo, J. Luis, and F. Wobbe. Generic Mapping Tools: Improved Version Released. *EOS Trans. AGU*, 94(45): 409–410, 2013. URL <http://dx.doi.org/10.1002/2013EO450001>. iii

Wikimedia Commons, A calcite crystal. A calcite crystal displays the double refractive properties while sitting on a sheet of graph paper. URL https://commons.wikimedia.org/wiki/File:Crystal_on_graph_paper.jpg. © 2011 User:APNMJM (CC BY-SA 3.0). Accessed 14th April, 2019. 192

Wikimedia Commons, Corkscrew. URL <https://commons.wikimedia.org/wiki/File:Korkenzieher.jpg>. © 2004 Horst Frank (GFDL). Accessed 14th April, 2019. 267

Wikimedia Commons, Decca Navigator Mk 12. URL https://commons.wikimedia.org/wiki/File:Decca_Navigator_Mk_12.jpg. © 2005 User:Stahlkocher (GFDL). Accessed 14th April, 2019. 285



- Wikimedia Commons, Exploration theodolite. URL https://commons.wikimedia.org/wiki/File:Exploration_theodolite_img_1660.jpg. © 2005 User:Rama (CeCILL). Accessed 14th April, 2019. 120
- Wikimedia Commons, Plane table with alidade. Cut-away view of plane table with alidade. URL https://commons.wikimedia.org/wiki/File:Plane_table_cgs00426.jpg. © NOAA. Accessed 13th April, 2019. 43
- Wikimedia Commons, The prime meridian (Greenwich). URL https://commons.wikimedia.org/wiki/File:Prime_meridian.jpg. © 2005 User:Prioryman (GFDL). Accessed 14th April, 2019. 48
- Wikimedia Commons, Standard metre in Paris. One of the historical (18th century), standard-meters (mètre-étalon) by Chalgrin located at 36, rue de Vaugirard in Paris. URL https://commons.wikimedia.org/wiki/File:M%C3%A8tre-%C3%A9talon_Paris.JPG. © 2010 User:LPLT (CC BY-SA 3.0). Accessed 13th April, 2019. 24
- Wikipedia, 1462 Zamenhof. URL https://en.wikipedia.org/wiki/1462_Zamenhof. Accessed 26th August, 2019. 457
- Wikipedia, 3D rotation group, Topology. URL https://en.wikipedia.org/wiki/3D_rotation_group#Topology. Accessed 8th April, 2019. 136
- Wikipedia, Alfvén's theorem. URL https://en.wikipedia.org/wiki/Alfv%C3%A9n's_theorem. Accessed 1st September, 2021. 523
- Wikipedia, The AN/UYK-1 (TRW-130) Computer. URL [https://en.wikipedia.org/wiki/Transit_\(satellite\)#The_AN/UYK-1_\(TRW-130\)_Computer](https://en.wikipedia.org/wiki/Transit_(satellite)#The_AN/UYK-1_(TRW-130)_Computer). Accessed 1st July, 2019. 287
- Wikipedia, Bonferroni correction. URL https://en.wikipedia.org/wiki/Bonferroni_correction. Accessed 10th April, 2019. 412
- Wikipedia, Chemical polarity. URL https://en.wikipedia.org/wiki/Chemical_polarity. Accessed 9th April, 2019. 203
- Wikipedia, Dandelin spheres. URL https://en.wikipedia.org/wiki/Dandelin_spheres. Accessed 12th April, 2019. 474
- Wikipedia, Discrete wavelet transform. URL https://en.wikipedia.org/wiki/Discrete_wavelet_transform. Accessed 9th April, 2019. 255



- Wikipedia, Earth's internal heat budget. URL https://en.wikipedia.org/wiki/Earth's_internal_heat_budget. Accessed 12th April, 2019. 487
- Wikipedia, Euler spiral. URL https://en.wikipedia.org/wiki/Euler_spiral. Accessed 9th April, 2019. 243
- Wikipedia, Gold code. URL https://en.wikipedia.org/wiki/Gold_code. Accessed 9th April, 2019. 293
- Wikipedia, Groundwater-related subsidence. URL https://en.wikipedia.org/wiki/Groundwater-related_subsidence. Accessed 12th April, 2019. 492
- Wikipedia, Kelvin–Stokes theorem. URL https://en.wikipedia.org/wiki/Stokes%27_theorem. Accessed 12th April, 2019. 522
- Wikipedia, Kilobyte. URL <https://en.wikipedia.org/wiki/Kilobyte>. Accessed 8th April, 2019. 25
- Wikipedia, The LAGEOS time capsule. URL https://en.wikipedia.org/wiki/LAGEOS#Time_capsule. Accessed 12th April, 2019. 481
- Wikipedia, Leap second. URL https://en.wikipedia.org/wiki/Leap_second. Accessed 10th April, 2019. 468
- Wikipedia, Magnetosphere. URL <https://en.wikipedia.org/wiki/Magnetosphere>. Accessed 12th April, 2019. 473
- Wikipedia, Non-SI units mentioned in the SI. URL https://en.wikipedia.org/wiki/Non-SI_units_mentioned_in_the_SI. Accessed 8th April, 2019. 25
- Wikipedia, Photon polarization. URL https://en.wikipedia.org/wiki/Photon_polarization. Accessed 8th April, 2019. 191
- Wikipedia, Planimeter. URL <https://en.wikipedia.org/wiki/Planimeter>. Accessed 9th April, 2019. 259
- Wikipedia, Pulsar. URL <https://en.wikipedia.org/wiki/Pulsar>. Accessed 9th April, 2019. 523
- Wikipedia, Rhumb line. URL https://en.wikipedia.org/wiki/Rhumb_line. Accessed 8th April, 2019. 54
- Wikipedia, Schmidt camera. URL https://en.wikipedia.org/wiki/Schmidt_camera. Accessed 10th April, 2019. 463



- Wikipedia, Shoelace formula. URL https://en.wikipedia.org/wiki/Shoelace_formula. Accessed 9th April, 2019. 258
- Wikipedia, Sophie Germain. URL https://en.wikipedia.org/wiki/Sophie_Germain#Correspondence_with_Gauss. Accessed 24th November, 2019. 357
- Wikipedia, Struve Geodetic Arc. URL https://en.wikipedia.org/wiki/Struve_Geodetic_Arc. Accessed 7th April, 2019. 8
- Wikipedia, Synoptic scale in meteorology. URL https://en.wikipedia.org/wiki/Synoptic_scale_meteorology. Accessed 9th April, 2019. 336
- Wikipedia, Theorema Egregium. URL https://en.wikipedia.org/wiki/Theorema_Egregium. Accessed 8th April, 2019. 54
- Wikipedia, Transit 5B-5 — oldest active satellite. URL [https://en.wikipedia.org/wiki/Talk:Transit_\(satellite\)#Transit_5B-5_-_oldest_active_satellite](https://en.wikipedia.org/wiki/Talk:Transit_(satellite)#Transit_5B-5_-_oldest_active_satellite). Accessed 30th October, 2019. 283
- Wikipedia, Trunnion. URL <https://en.wikipedia.org/wiki/Trunnion>. Accessed 8th April, 2019. 144
- Wikipedia, Zodiac, Name. URL <https://en.wikipedia.org/wiki/Zodiac#Name>. Accessed 14th April, 2019. 458
- Lydia E. Winkel and Hans de Vries. *De ondergrondse pers 1940-1945*. Veen, 6th edition, 1989. URL http://publications.niod.knaw.nl/publications/WinkelDeVries_OndergrondsePers_1940-1945.pdf#page=224. entry 748, © 2014 NIOD (CC BY-SA 3.0). Accessed 5th November, 2019. 425

ABCDEFGHIJKLMNOPQRSTUVWXYZ

A

- Aalto University, 15
- Aboa (Dronning Maud Land, Antarctica), 506
- Academy of Sciences of France, 8, 9
- acceleration, from GPS measurements, 494
- accuracy vs. precision, 28
- active spare (GPS), 291
- adjustment
 - as a projection, 399
 - average, 366
 - correction, 398
 - linear regression, 367
 - results, 28
 - weight and correction, 359, 360
 - weight coefficient, 359
 - weighting, 359
- adjustment calculus, 28
 - gross errors, 400
- adjustment method
 - by conditions, 396
 - parametric, 363, 396
- aerial mapping, in detail survey, 228
- aerial work platform, 260
- aerotriangulation, 43
- AGA/Minilir, 248
- Air Force, USA, 290
- air pressure
 - at GNSS station, 500
 - variation, 493, 507
- airborne laser scanning (ALS), 252, 254
- Akaike information criterion, 395
- Åland Sea (Finland, Sweden), 95
- Alexandria (Egypt), 3
- Alfvén, Hannes, 523
- Alfvén's theorem, 521
- alidade
 - description, 124
 - and plane table, 43
- alidade level, 127
 - adjustment, 128
- alien (LAGEOS), 481
- Alkmaar (The Netherlands), 3, 5, 42
- almanac, GPS, 291, 293, 328
- almucantar*, 449
- alternative hypothesis, 400, 411
- ambiguity
 - difference, 338
 - distance measurement, 196
 - GPS, 305, 342
 - spinning encoding circle, 151
- ambiguity problem, 286, 304
 - distance measurement, 196
- ambiguity resolution
 - distance measurement, 197
 - GPS, 343
 - one-dimensional, 342
- The American Practical Navigator*, 157, 170
- amplitude modulation, 291
- Amsterdam (The Netherlands), 78, 93
- Amsterdam Ordnance Datum, *see* NAP
- Andes (mountain range), 9
- angle measurement, 119
 - digital, 137
- angle transformation, 164
 - of direction unknown, 165
- angular distance, geocentric, 437, 438
 - calculation equation, 438
- ANNA (satellite), 464
- Antarctica, 506
- antenna (GNSS)
 - calibration, 299
 - choke-ring model, 298, 299
 - electric centre, 298, 300
 - forced centring, 133
 - phase-centre variation, 298, 299

- antenna reference point (ARP), 298
 anti-aircraft shells, 462
 anticorrelation, 37
 apogee, 525
 apogee height, 476
 Apollo project, 480
 approach map, 126
 approximate co-ordinates, 82
 approximate value (linearisation), 371
 approximate values, vector of, 372
 APPS (software), 355
 aquifer, 492
 Aquila (Eagle, constellation), 458
 arctan (function), 70
 Arctic summer, 505
 area levelling
 explanation, 112, 113
 check, 114
 earthwork, 260
 result, 113
 argument of latitude, 328
 Aries (Ram, constellation), 459
 Aries, First Point of, 459
 “as-built” mapping, 232, 245
 ascending node, right ascension of, 525
 Ashtech Z-12 (GPS receiver), 298
 asphalt laying, 247
 asphalt, hot, refraction, 179
 astatization of spring gravimeter, 427
 asteroid, orbit determination, 362
 asthenosphere, 489, 491
 astigmatism, 101
 astrolabe, 449, 450
 Danjon, 450
 Aswan (Egypt), 3
 atan2 (function), 70
 atmosphere, 502
 angular momentum, 466
 propagation delay, 309
 temperature, 500
 atmospheric inversion, 180
 atomic clock, relativity, 95
 atomic time (TAI), 468
 attraction (Newton), 6
 attribute data, 233, 234
 aurora, 473
 AUSPOS (software), 355
 automatic target recognition (ATR), 150,
 154
 automation of measurement, 107
 auxiliary closing direction, 175
 auxiliary point, 225, 229
 auxiliary starting direction, 175
 average, 359
 adjustment, 366
 mean error, 361
 avskärning, 158
 azimuth, 65
 origin of word, 66
 astronomical, 276
 geodetic, 277
 measurement, 385
B
 Baarda, Willem, 403
 Baeyer, Johann Jacob, 10
 Baker, James, 465
 Baker–Nunn camera, 464
 Baltic Sea
 mareographs, 511
 sea-surface topography, 508
 bar-code staff, 107
 barometric height determination, 95
 barycentric co-ordinates, 63
 base extension network, 4
 base network
 first-order, 137
 lower-order, 218
 measurement technology, 212
 purpose of use, 212
 base point, 246
 base station
 DGPS, 345
 virtual, 351
 baseline
 in Lapland, 42
 in Nummela, 185, 194
 of Snellius, 3
 base-network measurement, 214
 methods, 212, 214
 task, 211
 base-station network
 DGPS, 350
 RTK, 350
 battery of GPS satellite, 289
 beehive pattern, 255
 BeiDou (positioning system), 283, 344
 bell curve, *see* normal distribution
 benchmark (image), 131
 benchmark set, 211
 quality, 214
 requirements, 214
 Bergen op Zoom (The Netherlands), 3, 5,
 42
 Bergensbanen (Norway), 235
 Bergstrand, Erik Ö., 185
 BGI, International Gravimetric Bureau,
 448
 BIFROST project, 491
 birefringence, 189
 Birmingham (UK), 202



- black body, thermal radiation spectrum, 189
- black hole, 480
- GPS Block I, II, IIA, IIR, IIR-M, IIF, 289
- GPS Block I, 291
- GPS Block IIIA, 289
- block signal, 151
- blue film, 496, 497
- Boller and Chivens, 465
- Bomford, Guy, 279
- Bonferroni correction, 412
- Bonsdorff, Ilmari, 211
- Bouguer anomaly, 447
 - calculation, 447
 - Southern Finland, 448
- Bouguer plate, 447
- Bouguer, Pierre, 9, 448
- Bowditch method, 170, 175, 178
- Bowditch, Nathaniel, 157, 170
- “*The Bowditch*”, 157, 170
- Box, George E. P., 55
- break line in terrain, 254
- broadcast ephemeris, 291, 327
 - quality, 341
 - upload, 290
 - use, 328
- Bromarv guest harbour, 51
- Broom Bridge (Dublin), 265
- Bruns equation, 432, 433
- Bruns, Ernst Heinrich, 432
- building location, setting out, 245
- building permit, 245
- built environment, 17, 251
- Buys Ballot, Christoph, 508
- Buys Ballot’s law, 508
- C**
- C/A code, 292, 293, 295
 - measurement accuracy, 295
 - modulation state, 295
- CAD, computer-aided design, 53, 255
- cadastral system, 16
- caesium clock, 26
 - GPS receiver, 290
 - GPS satellite, 287
- calcite, 189
- calibration (metrology), 247
- calibration baseline, 200
- calving of icebergs, 505
- camwheel (modulator), 194
- Canada
 - land uplift, 85, 491
 - magnetic pole, 485
- Cancer (Crab, constellation), 459
- Capricorn (constellation), 459
- Caracas (Venezuela), 492
- Carloforte (Italy), 467
- carrier wavelength, 303
- Carrington event, 478
- cartography, 46, 53
- Cassini, Jean Dominique, 7
- catenary, 187
- cause-and-effect relationship, 36, 40
- Cayenne (French Guyana), 13
- CDMA, *see* code division multiple access
- celestial mechanics, 5
- celestial pole
 - elevation angle, 5
 - motion, 470
 - projection of direction onto horizontal, 277
- celestial sphere, 458, 460
- centigon, 27
- centring of theodolite, 126, 132
 - why*, 125
 - crude, 127
 - precise, 128
- certainty, statistical, 33, 34
- CHAMP (satellite), 494, 502
 - accelerometer, 494
 - GPS receiver, 494
- Chandler wobble, 466
- Chandler, Seth Carlo, 466
- change plate (levelling), 111
- characteristic polynomial, 380
- Charjui (Turkestan, Russia), 467
- check using redundancy, 414
- chemical polarity of water molecule, 203
- chip rate, 297
 - C/A code, 295
 - P code, 297
- chirality of life, 189
- χ^2 distribution, 402, 403
 - degrees of freedom, 402
 - MATLAB, 403
 - non-central, 403
 - test, 402–404
- χ^2_3 distribution, 407
- choke-ring antenna, 298, 299
- Cholesky decomposition, 398
- Cincinnati (Ohio, USA), 467
- Cinemascope (film format), 464
- circle singularity (GPS), 322
- circular arc, 240
 - setting out, 240
- civil day, 457
- civil engineering, 19
- civil time, 461, 467
- Clairaut, Alexis, 13
- Clairaut’s theorem, 13
- Clapeyron, Emile, 500
- classical mechanics, 5



- Clausius, Rudolf, 500
 Clausius–Clapeyron equation, 500
 climate model, 502
 climatic global warming, 502, 505, 510
 climatology, 308
 clock offset (GPS), 337
 clock unknown (GPS), 303
 closing error
 check, 398
 co-ordinates, 173, 177
 direction, 176
 triangle, 358, 359
 clothoid, 243, 244
 principle, 243
 equation, 243
 parameter, 244
 coaxial solution, 195
 code division multiple access (CDMA), 293
 GLONASS-K, 344
 codomain, 31
 cofactor, 519
 coherence length
 VLBI, 482
 white light, 191
 coin flip, 30
 coin, fair, 32
 coincidence level, 102
 coincidence microscope, 124
 collimation axis, *see* sight axis
 collimation error, 142, 143, 154
 definition, 122
 correction, 144
 determination, 143
 field check, 144
 from observation notebook, 143, 144
 Colorado Springs (USA), 289
 column vector, 514
 saving paper, 516
 comet, orbit determination, 362
 “common-mode” error assumption, 335, 336
 comparator (length calibration), 186
 compass direction, 54
 complete set, of horizontal directions, 163
 complete sets, method of, 137, 161, 162
 error reduction, 161
 focusing, 136
 complete-set average, 163, 168
 components of position vector, 309
 compound curve, 241
 setting out, 242
 comprehensive plan, 235
 computer (human being), 362
 computer, submarine, 287
 conic section, 474
 connecting levelling of tide gauge, 510
 conservative field, 440
 construction surveying, 245
 construction, benchmarks needed, 214
 continental ice sheet
 Fennoscandian, 489
 is a *glacier*, 505
 mass change, 505
 thickness, 505
 continuity condition of transition curve, 244
 continuous distribution, 32
 control measure, 246
 control segment (GPS), 289, 327
 convection
 in the Earth’s mantle, 486, 487
 in the Earth’s outer core, 485
 in the Sun, 470, 471
 Conventional International Origin (CIO), 468
 co-ordinate conversion, 49, 84
 co-ordinate correction (Helmert), 272
 co-ordinate frame
 geocentric, 269
 right-handed, 266, 267
 topocentric, 269
 co-ordinate reference frame, 55, 56, 77
 national, 211
 co-ordinate reference system, 55, 77
 geocentric, 47, 265
 inertial or celestial, 266
 plane, 47
 realisation, 77
 terrestrial (ECEF), 266
 three-dimensional, 46
 two-dimensional, 46
 co-ordinate reference systems, guidance, 213
 co-ordinate transformation, 55, 56, 213
 co-ordinates
 2D+1D, 83, 277
 Cartesian, 268
 ellipsoidal, 267
 geocentric, 266
 notation, 265
 geodetic, 47, 267
 figure, 49
 construction, 48
 term, 49
 why use, 50
 geographical, 47, 267
 in geodesy, 46
 KKJ, 279
 local, 66
 connection to national frame, 67
 map projection, 46, 47
 (N, E, U), 273, 274



polar, 65, 66
 project specific, 67
 rectangular, 47, 84, 268
 figure, 49
 why use, 50
 spherical, 267
 temporary, 67
 three-dimensional, 46, 47
 rectangular, 48
 topocentric, 47, 50, 268
 notation, 269
 rectangular, 268
 Coriolis force, 507
 Coriolis, Gaspard-Gustave de, 507
 corkscrew, 266, 267
 corner-cube prism, 197, 198
 correction dissemination
 alternatives, 350
 communications satellite, 351
 mobile Internet, 351
 correlation, 42
 definition, 36
 and causation, 40
 examples, 37
 of aerial images, 254
 of tree rings, 294
 of VLBI signals, 481
 perfect, 36
 correlation method (GPS), 294, 295
 correlator, optical (Väisälä), 192
 cosine rule on sphere, 86
 half-angle version, 87
 cost, of an error, 35, 411
 covariance, 36
 critical inclination, 478
 Croll, James, 503
 cross-dipole antenna, 299
 crosshairs
 adjustment screw, 105
 better, 137
 damage, 125
 history, 99
 cross-measure, 224, 246
 cross-section measurement, 115, 116
 crosstalk, 293
 cruise missile, 256
 crustal motion
 anthropogenic, 492
 monitoring, 493
 culmination of a star, 460
 cultural data, 233, 234
 curl (operator), 522
 curvature correction
 Earth's surface, 204
 measurement ray, 204

cut-off elevation angle, for GNSS
 observations, 291, 298
 cycle slip, 338, 340
 Cygnus (Swan, constellation), 458
D
 Dandelin spheres, 474
 dangerous circle
 GPS, 322
 resection, 160
 Danish Sound bridge (Denmark), 247
 data layer, 234
 data snooping, 405
 datum, 77, 79
 European, 280
 geodetic, 77
 datum difference (vertical datum), 81
 datum point
 concept, 77
 choice, 81, 82
 proximity, 78
 datum transformation, 78
 between two ellipsoids, 272, 275
 height, 81
 in the plane, 82
 small parameter values, 82, 83
 Decca
 system, 284, 285
 carrier wave, 284
 hyperbola, 285, 286
 lane, 286
 receiver, 285
 station, 284
 decimilligon, 27
 declination, 460
 deep-sea trench, 435, 487
 deformation
 detection, 494
 measurement, 114, 246
 monitoring, 246
 deformation analysis, 395
 datum point, 418, 419
 hypotheses, 417
 time, 417
 deformation monitoring, 493
 degree (angle unit), 26
 degree Celsius, 26
 Delaunay triangulation, 62, 255
 Delaunay, Boris, 62
 Delaunay, Vadim, 62
 dendrochronology, 193, 294
 Department of Defense, USA, 290
 dependence, statistical, 40
 Descartes, René, 268
 design matrix
 definition, 375



- azimuth, 385
- Helmert transformation, 389, 390
- linear regression, 367
- pseudo-range, 313
- slant range, 383, 384
- station adjustment, 163
- design speed, 244
- detail survey
 - definition*, 222
 - carrying out, 228
 - data measured, 228
 - equipment and software, 229
 - purpose of use, 229
 - RTK, 222
 - workflow, 229, 230
- detailed shore plan area, 236
- determinant, 518
 - antisymmetry, 519
 - computing, 518
 - interpretation, 374
 - vanishing, 518
- deuterium, 504
- dew point, 501
- DGPS, *see* differential GPS
- die, 30
 - cast, 30, 31
 - fair, 31
- Diego Garcia (island, Indian Ocean), 300, 302
- difference observation, 336, 337, 340
 - double, 336, 339–342
 - effect on errors, 338
 - single, 336, 340
 - triple, 336, 339, 340
- difference of ambiguities, 338
- difference transformation (Helmert), 75
- differential GPS (DGPS), 345–347
 - difference observation, 346
- Digges, Leonard, 119
- digital elevation model (DEM), 256
 - term*, 251
- digital height model (DHM), 251
- digital terrain model (DTM), 256
 - term*, 251
 - earthwork, 260
 - visualisation, 255
 - volume calculation, 263
- dilution of precision (DOP), 313, 318, 333
 - description*, 312
 - azimuth symmetric, 321
 - calculation, 315
 - calculation script, 332
 - example, 322, 323
 - variants, 314
- dimension, 23
- direction cosine (linearisation), 312
- direction vector
 - to quasar, 481
 - to satellite, 308, 310
- discrete wavelet transform (DWT), 255
- dispersion in ionosphere, 305
- distance
 - geometric, 302, 309
 - intercontinental, 45
- distance measurement
 - curvature corrections, 204
 - electronic, 44, 195
 - electro-optical, 44
 - height reduction, 206
 - infrared light, 44
 - instrumental correction, 201
 - laser light, 44, 196
 - LED, light-emitting diode, 44, 196
 - mechanical, 44, 185
 - microwaves, 44, 196
 - modulation, 196
 - random error, 201
 - reduction to reference level, 206, 207
 - refraction correction, 203
 - second velocity correction, 204, 205
 - steel tape, 185
 - terrain correction, 205
 - visible light, 44
 - white light, 196
- distance measurement device, *see* range-finder
- Distomat (range-finder), 195
- distribution, *see* probability density distribution
- disturbing potential, 430, 432
- div (operator), 522
- DNA (molecule), 189
- DOP ellipsoid, 316, 317
 - figure*, 317
 - principal axes, 316, 317
- Doppler positioning system, *see* Navy Navigation Satellite System
- DORIS
 - description*, 483
 - beacon, 483
- double difference (RTK), 347
- double pentagon prism, 223
- downbanding, 293
- drift
 - gravimeter, 426
 - quartz oscillator, 338
 - range-finder, 199
- drinking-water extraction, 114, 492
- Dronning Maud Land (Antarctica), 506
- dry air, partial pressure, 499
- dynamo, natural, 471, 485, 521



E**Earth**

- average density, 7
- centre of mass, 47, 57, 265
- core density, 7
- core heat, 487
- crustal density, 7
- gravitational field, 6
 - determination, 496, 498
- internal structure, 485
- magnetic field, 473, 484, 485
- magnetic poles, 485
- mantle density, 7
- mantle viscosity, 489
- mass distribution model, 6
- mean radius, 86
- pear shape, 11
- radius, 3
- radius of curvature, 5, 180
 - transversal, 267, 268
- rotation rate, 13
- thermal energy, 487
- Earth ellipsoid parameters, 7
- Earth flattening, 6–8, 13
- Earth orbit, 457, 458
 - eccentricity, 505
 - major axis, 504
- Earth orientation parameters (EOP), 85, 331, 468
- Earth rotation, 457, 479
 - centrifugal force, 429, 453
 - Coriolis force, 507
 - monitoring, 461, 482
 - period, 291
 - rate, 467
 - variation, 85, 465, 466, 468
- Earth rotation axis
 - direction, 4, 47, 57, 265
 - momentaneous, 465, 468
 - motion, 469, 470
 - tilt, 504
 - VLBI, 482
- Earth, figure of the, 15, 476
 - mathematical, 10, 15, 435
 - physical, 6, 15
 - true, 9
- earthquake
 - Izmit (Turkey), 495
 - plate boundary, 493
 - Sendai (Japan), 493
- earthquake-prone area, 493
- Earth's surface forms, 251
- earthwork calculation, 116, 245, 260
- Easter Scheldt storm-surge barrier (The Netherlands), 247, 248
- easting, 59, 65

- eccentric anomaly, 526
- eccentricity error, 142
- eccentricity in height direction, 183
- eclipse, 458
- ecliptic, 458
- ED50, 11, 277
 - precision, 280
- EGM2008, 118, 439
- EGNOS, 351
- Eiffel Tower, 72 names, 194, 330, 362, 453, 500, 508
- eigenvalue problem, 380, 381
- Einstein, Albert, 189, 426
- elasticity, 489
- electrical resistivity tomography, 248
- electromagnetic field, 189
- electromagnetic field theory, 188
- electromagnetic radiation, 188
 - in air, 201
 - polarisation, 192
 - quantisation, 189
 - spectrum, 190
- electromagnetic wave, 188
- elevator experiment, Einstein's, 426
- elevator speech, 282
- elimination and back-substitution, 368
- ellipse
 - definition, 474
 - focal point, 474
- ellipsoid of revolution, 7, 49
 - as figure of the Earth, 13
 - geodesic, 12
- eLoran (navigation system), 284
- encoding circle, 150
 - absolute, 149
 - incremental, 150
- energetic correctness, 446
- engineering geodesy, 114, 246
- engineering surveying, 246
- entrance pupil (aperture), 99
- eötvös (unit), 453
- Eötvös tensor, 452, 453
 - measurement, 452
 - on Earth's surface, 453
- Eötvös, Loránd, 426, 452
- ephemeris time (ET), 468
- equinox, vernal, 266, 459
- equipotential surface
 - description, 90
 - curvatures, 452
 - of the geopotential, 432, 433
 - of the normal potential, 430, 432, 433
 - radii of curvature, 452
 - spacing, 442
- equi-value curve, 428
- era, common or Christian, 2



- Eratosthenes, 1–3
Erdmessung, 15
 error
 examples, 30
 gross, 29, 402, 411
 simulated, 408
 of the first kind, 410
 of the second kind, 411
 random, 29
 systematic, 29
 elimination, 136
 error ellipse, 39, 40, 380
 and variance matrix, 41
 axes, 40
 axes directions, 381
 centre, 40
 semi-axes, 381
 error ellipsoid, 40, 314, 316, 317
 azimuth symmetric, 320
 error propagation, 37
 forward geodetic problem, 378
eteenpäin leikkaus, 158
 ETOPO1 (terrain model), 252
 ETOPO2 (terrain model), 253
 ETRS89, 57, 279
 ETRS-GK*n*, 62
 ETRS-TM35FIN, 61
 Euler free nutation, 466
 Euler notation, 428, 451
 Euler spiral, 243
 Euler, Leonhard, 243, 466
 Eurasian plate
 motion, 281
 sea floor, 486
 EUREF data centre, 86
 EUREF, Reference Frame Subcommittee
 for Europe, 57
 EUREF89, 57, 280
 EUREF-FIN, 279, 280
 creation, 57
 densification, 217
 determination, 217
 map projection, 61
 Everest, George, 10, 448
 Everest, Mount, 10, 220
 excavation work, damage, 249
 exit pupil, 99
 expectancy
 coin flip, 32
 continuous quantity, 34
 die cast, 32
 discrete quantity, 31
 expectancy (operator), 35, 360
 eyeglasses
 cylindrical, 101
 of observer, 100, 101
 prismatic, 101
F
 face left/right, 136
 error elimination, 161
 Falkland Islands (South Atlantic), 497
 false easting, 60
 farsightedness, 101
 fault, geological, 492, 495
 Fennoscandia
 crustal motions, 491
 land uplift, 85, 489
 cross-section, 490
 pattern, 490
 FGI, Finnish Geospatial Research
 Institute (National Land Survey)
 FinnRef, 217
 GNSS service, 353
 field equations
 Einstein's, 426
 Maxwell's, 521
 field gravimeter, 426
 field sketch, 224
 FIN2000 (geoid model), 94
 FIN2005N00 (geoid model), 279
 Finnish Transport Infrastructure Agency's
 DGPS service, 352
 FinnRef, 216, 217
 fix solution (RTK), 349
 Fizeau, Hippolyte, 194
 Fizeau's method, 195
 flat Earth approximation, 14, 89
 flight simulator, 256
 float solution (RTK), 349
 floating mark (photogrammetry), 43
 flood safety, 246
 fluid surface as a reference, 94
 fluxion, 325
 focusing, 99, 100
 footprint of satellite altimeter, 506
 footscrew, 124
 forced centring
 application, 134
 principle, 133
 forced-centring device, 131
 attachment to tripod, 124
 image, 125
 parallel shift, 130
 forward geodetic problem
 definition, 68
 error propagation, 378
 generally, 68
 in the plane, 68
 foundation of a building, 246
 four-parameter transformation, *see*
 Helmert transformation in the



- plane
 - free fall, acceleration in, 425
 - free nutation (Euler), 466
 - free-air anomaly, 434
 - Southern Finland, 448
 - free-stationing survey, 227, 228
 - figure, 228
 - setting out, 239
 - with a tacheometer, 227
 - frequency calibration, 201
 - frequency division multiple access (FDMA), 344
 - frequency measurement, 45
 - frequency modulation, 291
 - Fresnel, Augustin-Jean, 189
 - Frisius, Gemma, 3
 - frost heaving, 126, 128
 - “frozen-in” magnetic field, 472, 521, 522
 - Fuglenes (Norway), 10
 - functional model, 397
 - model error, 403, 404
 - slant range, 385
 - fundamental equation of physical geodesy, 435, 437
- G**
- Gaithersburg (Maryland, USA), 467
 - Galilei, Galileo, 194, 425
 - Galileo (positioning system), 283, 344
 - Gascoigne, William, 99
 - Gauss, Carl Friedrich, 357, 359, 362, 435
 - Gaussian distribution, *see* normal distribution
 - Gauss–Krüger (map projection), 59, 61
 - conformality, 62
 - scale distortion, 63
 - zone division, 59
 - GDGPS, Global Differential GPS, 353
 - general circulation model (GCM), 256
 - general plan, 235
 - generalisation, of a map, 18
 - Geneva (Switzerland), 114
 - GeoBasic (Leica), 155
 - geocentricity, 47, 56, 57
 - GeoCOM (Leica), 154
 - geodesic, 12
 - measurement line, 13
 - geodesy, definition, 14, 15
 - geodetic network
 - European, 11
 - GPS, 342
 - orientation, 276
 - reliability, 395
 - terrestrial, 219
 - three-dimensional, 218, 219
 - Geodimeter (range-finder), 44, 185
 - geodynamics, 15, 84, 331, 479
 - exotic techniques, 480
 - local, 492
 - geographic information technology, 233
 - geoid
 - definition, 14, 89
 - determination, 256, 435
 - European, 11
 - global, 438
 - term, 10, 15
 - geoid height, 433, 439
 - computation, 437
 - near zone, 438
 - variation, 436
 - geoid map
 - global, 494, 498
 - GOCE, 497
 - geoid model, 84
 - Bomford, 277
 - FIN2000, 94
 - FIN2005N00, 279
 - geoid undulation, 433
 - geological clock, 486
 - geomensuration, 15
 - geophysical research, 493, 505
 - geopotential, 90
 - as a co-ordinate, 46, 90
 - difference, 45, 443
 - energy, 439
 - fine structure, 494
 - gradient, 429
 - measurement unit, 441
 - variation, 436
 - visualisation, 429
 - geopotential model, 95
 - global, 438, 494
 - geopotential number, 443, 446
 - definition, 441
 - geopotential table, 429
 - geospatial information, 18
 - shared use, 233
 - geostrophic equations, 508
 - geostrophic flow, 508
 - Geotrim Oy, 352
 - Germain, Marie-Sophie, 357
 - GIS, geographic information system, 21, 233
 - glacial isostatic adjustment (GIA), 85, 488, 490, 509
 - glacial loading, 418
 - Global Navigation Satellite Systems, *see* GNSS
 - Global Positioning System, *see* GPS
 - GLOBE (terrain model), 252
 - GLONASS (positioning system), 283, 318
 - frequencies, 344



- RTK, 350
- SP3-c (precise ephemeris), 330
- GLONASS-K, 344
- GNSS, 45, 198
 - definition*, 284
 - base-network measurement, 221
 - co-location, 509
 - Earth rotation monitoring, 468
 - geocentricity, 47
 - geodynamics, 479
 - land-ice research, 506
 - tidal loading, 493
 - tropospheric sounding, 501
- GNSS buoy, 510
- GNSS limbsounding, 502, 503
- GNSS monitoring, 493
 - land uplift, 492
- GNSS network
 - continuously operating, 216, 217
 - global, 503
- GNSS radio occultation, 502, 503
- GNSS receiver
 - on satellite, 502, 507, 509
 - software-defined, 300, 344
- GOCE (satellite), 95, 497
 - name*, 498
 - accelerometer, 497
 - climate research, 508
 - gravitational gradiometer, 497, 498
 - sea-surface topography, 499
- gon (angle unit), 26, 27
- Google Earth, 56, 255
- Google Maps, 117, 118
- Google Street View, 223
- GPS, 45, 283
 - development, 289
 - modernisation, 291
 - original system, 284
- GPS constellation, original, 291, 292
- GPS measurement
 - geometry, 308, 311
 - goodness, 309, 318
 - repeat, 291
 - observable, 295, 300, 340
 - pseudo-range, 295, 309
 - relative, 319, 342
 - static, 319
 - vector, 340, 342
 - water-vapour influence, 203
- GPS positioning, 284
 - geodetic, 303, 354
 - geometry, 308, 318
 - real-timeness, 344
 - task, 308
- GPS post-processing, 344, 353
- GPS receiver
 - clock correction, 302
 - clock offset, 302, 303, 309, 338
 - dual-frequency, 297, 303
 - geodetic, 300
 - received signal, 293
- GPS satellite, 287
 - atomic clock, 287
 - clock correction, 289, 290, 327
 - clock offset, 303, 309, 339
 - communication channel, 289
 - components, 287
 - Doppler shift, 328
 - fingerprint, 293, 294
 - frequencies, 288
 - health status, 289, 327
 - lifetime, planned, 289
 - orbit, 324, 326
 - orbit data, 289
 - orbit perturbation, 289
 - orbital elements, 325
 - orbital inclination, 291, 324
 - orbital period, 291
 - orbital plane, 291, 325
 - radio transmitter, 288
 - solar panels, 289
 - three-axes stabilisation, 289
 - thruster, 289
 - transmission power, 288
- GPS signal, 291, 292
 - bandwidth, 293
 - carrier frequencies, 304
 - carrier phase, 297, 303, 305, 340
 - carrier phase, raw, 304
 - correlation, 295
 - lock-on, 293, 327
 - modulation, 292
 - replica, 294, 295
 - wavelength, effective, 296
- GPS time, 469
- GPS/MET (satellite), 502
- GPU, geopotential unit, 441
- GRACE (satellite pair), 496, 497
 - accelerometer, 496
 - land-ice research, 505
 - microwave link, 496
- GRACE Follow-On (GRACE-FO), 496
- grad (angle unit), *see* gon
- grad (operator), 450
- grade measurement
 - arc length, 2
 - in Lapland, 8, 9, 42, 185
 - in Peru, 8
 - of Eratosthenes, 2
 - of Snellius, 3–5, 42
 - of Struve, 8



- plumb-line directions, difference, 2, 4, 5, 8
 - gradient, 428
 - gradient vector, 428
 - graduation circle
 - eccentricity error, 122
 - manufacture, 137
 - reading, 140, 141
 - graduation error, 142
 - graduation machine, 137
 - graduation of levelling staff, 109
 - bar code, 109
 - chessboard, 108
 - “E”, 108, 110
 - line, 108
 - gravimeter, 426, 427
 - drift, 426
 - gravimetric point height, 434
 - gravimetry, 426
 - monitoring, 493
 - gravitation, 425
 - universal law of, 6
 - gravitational constant, universal, 6
 - gravitational gradient, 451
 - gravitational gradient tensor, 451
 - gravitational gradiometer, 497, 498
 - gravity
 - definition*, 426
 - along plumb line, 443, 444
 - centrifugal force, 429
 - change, land uplift, 492
 - monitoring, 492
 - on Earth’s surface, 426
 - plumb-line mean, 443, 444
 - strength (scalar), 430
 - variation, 436
 - gravity anomaly
 - definition*, 434
 - and disturbing potential, 435
 - and height, 434
 - calculation from measurements, 434
 - variation, 435
 - gravity disturbance, 433
 - gravity field, 430
 - fine structure, 494
 - is conservative, 440
 - temporal changes, 496, 497
 - gravity flattening, 13
 - gravity vector, 429, 434, 439
 - definition*, 430, 440
 - as gradient of geopotential, 450
 - direction, 448
 - properties, 440
 - gravity-gradient tensor, 451
 - Gray code, 149
 - Gray, Frank, 149
 - Great Belt bridge (Denmark), 247
 - Greenland, 484
 - Greenwich Apparent Sidereal Time (GAST), 459
 - Greenwich Mean Time (GMT), 48
 - Greenwich meridian
 - image*, 48
 - direction, 266
 - meridian treaty, 48
 - plane, 48
 - Greenwich Observatory, Royal, 47
 - meridian circle, 467
 - Greenwich paradox, 282
 - Greenwich sidereal time, 459, 461
 - Groningen (The Netherlands), 492
 - ground control point (GCP), 254
 - ground-penetrating radar, 248
 - group index of refraction
 - dry air, 202
 - information content, 202
 - ionosphere, 306
 - group velocity, 305, 306
 - GRS80 reference ellipsoid, 49
 - GSI (Leica), 154, 155
 - GTOPO30 (terrain model), 252
 - guidance of geodetic works, 213
 - gyroscope, 469
 - gyroscope phenomenon, 4
- ## H
- half-angle formula (arc tangent), 70
 - half-angle formula (cosine rule on sphere), 87
 - Hamilton, Sir William Rowan, 265
 - hand-held GNSS, 126
 - Hannover (Germany), 362
 - harbour logistics, 247
 - hard-drive pack (VLBI), 481
 - Härmälä, Seppo, 16
 - harmonisation of significance levels, 413
 - harp, Phoenician, 428
 - hat notation, 362
 - Hatanaka compression, 354
 - Hayford ellipsoid, 59
 - Heaviside, Oliver, 521
 - height
 - definition*, 89
 - above sea level, 91, 92, 95
 - concepts, 93
 - dynamic, 445
 - from levelling, 14
 - from the reference ellipsoid, 14, 49, 91, 267
 - metric, 90, 441, 446
 - normal, 445
 - definition*, 445



- orthometric, 89, 92, 93, 442–444
 - definition, 444
 - approximate value, 444
 - height anomaly, 432
 - height base point, 246
 - height contour, 428
 - calculation, 256
 - earthwork, 260
 - on topographic map, 96
 - height deformation
 - analysis, 418
 - displacement, 419
 - monitoring, 419
 - shifting variate (\mathcal{E}), 419
 - height determination, 90
 - barometric, 95
 - height location, of a building, 246
 - height system, 14, 96, 97
 - orthometric, 92
 - height type, 91, 446
 - properties, 446
 - Heiskanen, Veikko A., 16
 - helium-neon laser, 26, 202
 - Hellenes, 1
 - Helmert transformation, 71
 - centre-of-mass co-ordinates, 391
 - equations, 75
 - free stationing, 228
 - in the plane, 72–74, 76, 389
 - open traverse, 173
 - residuals, 391, 393
 - rotation angle, 73, 74
 - scale distortion, 76, 271
 - scale ratio, 73
 - scaling, 74
 - “small”, 272
 - steps, 73
 - three-dimensional, 270
 - translation parameters, 73, 75
 - translation vector, 74, 75
 - two-dimensional, 270
 - variances of parameters, 391
 - Helmert, Friedrich Robert, 72
 - Helmholtz brightness illusion, 108
 - Helmholtz, Hermann von, 108
 - Helsinki (Finland), 79, 336
 - Helsinki astronomical observatory, 78, 467
 - Helsinki mean sea level, 78, 93
 - Hermite, Charles, 420
 - Hermitian conjugate, 420
 - Hermitian variance matrix, 420
 - heterodyning, 293
 - heterophoria, 101
 - Hexagon Oy, 353
 - highway levelling, 112
 - Himalayas, 10
 - Hipparchus, 469
 - histogram, 32
 - Holmes, Arthur, 486
 - horizon plot, 318
 - horizon, plane of the, 119
 - horizontal angle, 120
 - definition, 120
 - figure, 121
 - measurement, 157, 158, 386
 - horizontal axis (theodolite), 121
 - horizontal circle
 - electronic readout, 151
 - graduation-error minimisation, 161
 - locking screw, 161
 - horizontal datum, 78, 81, 82
 - horizontal deformation
 - analysis, 419
 - co-ordinate displacement, 422
 - displacement vector, 420
 - mean error of unit weight, 423
 - monitoring, 421
 - shifting variate (\mathcal{E}), 420, 423
 - weight-coefficient matrix, 423
 - horizontal direction, 65, 227
 - hot spot, 488
 - hour angle, 460
 - humidity, relative, 500, 501
 - Huygens, Christiaan, 189
 - Earth flattening, 7, 13
 - nature of light, 188
 - polarisation, 189
 - hydrogen-maser clock, 287
 - hygrometer, 501
 - hyperbola (Decca), 285, 286
 - hyperbolic positioning system, 45, 284
 - hyperboloid of revolution, 347
 - hypothesis-freeness, 444, 446
- I**
- IAG
 - EUREF subcommission, 281
 - history, 11
 - International Earth Rotation and Reference Systems Service, 466
 - International GNSS Service, 331
 - International Gravimetric Bureau (BGI), 448
 - iceberg, calving, 505
 - Iceland spar, 189, 191, 192
 - IJsselmeer (The Netherlands), 95
 - illumination system (theodolite), 141
 - image plate, photographic, 463
 - image processing, 255
 - impossibility, statistical, 33
 - incremental measurement, 286
 - independence, statistical, 41



- independent, identically distributed
(*i.i.d.*), 364
- index error, 146, 147
definition, 145
field check, 148
removal, 147
- index of refraction, 201
air, 203
air, microwaves, 203, 307, 499
air, visible light, 202
dry air, 202
glass, 139
representativeness, 205
- industrial measurement, 246
- inertia, 425
- inertial measurement unit, 223
- infrastructure, 236
construction, 16, 236, 247
planning, 16, 17, 116, 236
- in-phase (modulation), 297
- InSAR, 494, 495
- inskärning*, 158
- insolation
on edge of ice sheet, 503
summer, 504
- INSPIRE (EU directive), 252
- installation measurement, 114, 245, 246
- instrument co-ordinates, 50, 119, 268, 269
- instrument height, 183
- instrument of appeal (zoning), 17
- instrument, terrestrial measurement, 45
- integrity monitoring (GPS), 351
- interference fringe (SAR), 494, 495
- interference measurement (Väisälä), 194
- interference of light waves, 188
- International Association of Geodesy, *see*
IAG
- international collaboration, 438
- International Earth Rotation and
Reference Systems Service
(IERS), 466, 468
- International Ellipsoid of 1924, 59
- International GNSS Service (IGS)
description, 331
Central Bureau, 331
ionosphere modelling, 503
precise ephemeris, 330
tracking stations, 330
- International Gravimetric Bureau (BGI),
448
- International Latitude Service, 466
- International Polar Motion Service, 467
- International Space Station, 478
- intersection, 158, 159
- intrinsic angular momentum (spin), 191
- invar (alloy), 44, 185
- invar staff, 109
- invar tape, 110
- invariant of point variance matrix, 382
- inverse geodetic problem, 69
- inverse mapping, 373
- inverse matrix, 516–518
- inverse problem, 160
- inversion, mathematical, 498
- inverted barometer (IB), 507
- ion, 521
- ionic engine (GOCE), 497
- ionosphere
dispersion, 305
electron density, 306, 503
propagation delay, 302–304, 306
research, 308
sounding, 503
total electron content (TEC), 503
- Ionosphere Map Exchange Format
(IONEX), 503
- ionosphere model, 503
- ionosphere-free linear combination, 307
- iron-ore particle, 485
- irrigation-water extraction, 114, 418, 492
- isostatic adjustment, 489
- isostatic compensation, 449
- isothermal surface, 206
- isotope
deuterium, 504
radioactive, 486
- ITRS, International Terrestrial Reference
System, 272
- Izmit (Turkey), 495
- ## J
- dynamic flattening (J_2), 476
- Jacobi, Carl Gustav Jacob, 373
- Jacobi, matrix of, 373
- Jason (satellite), 509
- Java (Dutch Indies, Indonesia), 435
- Jeffreys, Harold, 395
- jet fighter, 256
- Jet Propulsion Laboratory (JPL)
APPS, 355
GDGPS, 353
International GNSS Service, 331
- JHS recommendation
benchmark measurement, 213
EUREF-FIN, 213
municipalityGML, 213
N2000, 213
zoning base map, 213
- journalist, 282
- JPEG 2000, 255
- JUHTA, 213
- Jupiter's flattening, 13



Jyväskylä (Finland), 78

K

K XVIII (submarine), 425
 Kakkuri, Juhani, 205
 Karttakomitea, 211
 kelvin (temperature scale), 26
 Kelvin, Lord (Sir William Thomson), 522
 Kelvin–Stokes theorem, 522
 Kepler orbital elements, 326
 description, 525
 figure, 475
 GPS, 327
 Kepler, Johannes, 473
 Kepler's law of areas, 474
 Kepler's laws, 473, 474
 Kepler's third law, 476
 Kern (instrument manufacturer), 44, 130
 Kevo (Finland), 78
 kilocalorie, 25
 kinematic measurement, 286
 Kitab (Uzbekistan, Soviet Union), 467
 KKJ, 56, 57
 basic co-ordinate system, 59
 YKJ (Uniform Co-ordinate System), 61
 zone geometry, 60
 KM2 (terrain model), 252
 KM10 (terrain model), 252
 Kozai, Yoshihide, 328
 Kukkamäki method, 103
 Kukkamäki, T. J., 103
 Kvarken (Baltic Sea), 490

L

L_1, L_2 , GPS frequencies, 288, 291, 293, 304, 307
 L_5 , GPS frequency, 291
 LAGEOS (satellite), 481
 Lagrange interpolation, 330
 Lagrange, Joseph-Louis, 330
 land subsidence
 anthropogenic, 114
 mineral extraction, 418, 492
 water extraction, 492
 land uplift, post-glacial, 114, 466, 479
 horizontal motions, 491
 measurement, 491
 physical character, 489
 Land Use and Building Act 1999/132, 235
 land, purpose of use, 16
 landscape visualisation, 256
 land-use planning, 17
 lane (Decca), 286
 Langmuir, Irving, 523
 Laplace azimuth equation, 277

Laplace azimuth phenomenon, 278
 Laplace field equation, 437, 453
 Laplace, Pierre-Simon, 453
 Lapland, 8
 working from the large to the small, 214
 Large Hadron Collider (LHC), 114
 Larmor precession, 469
 laser level
 principle of operation, 115, 116
 earthwork, 260
 vertical plane, 116
 laser plummet, 124, 134, 154
 laser scanning
 airborne, 252, 254
 earthwork, 260
 terrestrial, 223
 laser-station network, global, 480
 latency, 344
 lateral refraction, 29
 latitude
 astronomical, 5, 449
 geocentric, 267, 268
 geodetic, 49, 267, 268
 variation, 466, 467

Lattamiehentie

Joensuu, Finland, 89
 Vantaa, Finland, 89

Laurentide land uplift area, 491
 law of large numbers, 31, 35
 leap second, 468
 least-squares method, 395
 invention, 362
 name, 360
 least-squares solution, 364
 precision, 365
 weighted, 396
 legend of map, 231
 Legendre, Adrien-Marie, 362
 Leica (instrument manufacturer), 152, 155
 Leiden (The Netherlands), 3
 length (quantity), 26, 185
 length of day (LoD), 85, 466
 monitoring, 466, 483
 levelling, 45, 95
 geometric, 97
 geometry, 96
 hydrostatic, 94
 safety, 112
 staff distance, 107, 112
 trigonometric, 182
 levelling instrument, 97, 98
 accuracy, 97
 automatic, 105
 classification, 98
 construction, 97
 digital, 107, 110



- field check, 102, 103
 - horizon adjustment, 103, 104
 - purpose of use, 97
 - self-levelling, 105, 106
 - traditional, 97
 - levelling network, 45, 96
 - loop, 45
 - planning, 112
 - levelling of theodolite, 126, 132
 - crude, 127
 - precise, 127, 128
 - levelling rod, *see* levelling staff
 - levelling spike, 110, 111
 - levelling staff
 - description*, 108
 - bull's-eye level, 110
 - choice, 98
 - classification, 110
 - for base network, 110
 - purpose of use, 110
 - reading through foliage, 108
 - self-calculating, 113
 - levitation, magnetic, 523
 - life, chemistry of, 189
 - lifting screw, 97, 103, 105
 - light, nature of, 188
 - line microscope, 138, 139
 - line of force
 - of gravity, 429, 432
 - of normal gravity, 430, 432
 - of solar magnetic field, 473
 - linear regression, 200, 361
 - adjustment, 367
 - definitions, 362
 - design matrix, 367
 - example, 369
 - in terms of averages, 362
 - observation equations, 367
 - shifting variate (\mathcal{E}), 407, 409
 - simulated error, 409, 410
 - testing, 406, 408
 - linearisation, 370
 - one-dimensional, 372
 - pseudo-range, 312, 313
 - vector mapping, 373
 - linearised quantity, 371, 375
 - link between significance levels, 412
 - Lippershey, Hans, 119
 - liquid compensator (height index), 148
 - Listing, Johann Benedict, 10
 - lithosphere, 489
 - living space, human, 219
 - loading
 - atmospheric, 493
 - glacial, 418
 - ocean, 493
 - ocean tidal, 493
 - Local Apparent Sidereal Time (LAST), 459
 - local detailed plan, 235
 - local master plan, 235
 - location data, 233, 234
 - location review, 246
 - logarithmic scale, 26
 - longitude
 - astronomical, 461
 - geocentric, 267
 - geodetic, 49, 267
 - longitude determination, 461
 - Loran-C (navigation system), 284
 - lorry, articulated, 243
 - lower mantle, 489
 - loxodrome (rhumb line), 54
 - lunar eclipse, 2
 - lunar laser-ranging (LLR), 480
 - Lunokhod (lunar rover), 480
 - Lyra (Lyre, constellation), 458
- M**
- “*Maan toinen kuu*” (article), 462
 - machine guidance, 247
 - magma, 485, 487
 - magnetic bottle (radiation belts), 473
 - magnetic flux, 523
 - magnetisation stripe (sea floor), 484
 - magnetism of pendulum compensator, 106
 - magnetohydrodynamics (MHD), 472, 523
 - magnetosphere, 473
 - magnification of telescope, 99
 - malfunction of device, 29
 - map
 - legend, 20
 - purpose of use, 18
 - map algebra, 234
 - map co-ordinates, 55
 - map north, 179
 - map projection, 53
 - application, 221
 - choice, 221
 - conformal, angle-preserving, 54, 221
 - distortion, 55
 - equidistant, 54
 - equivalent, equal-area, 54
 - gnomonic, 55
 - purpose of use, 54, 55
 - scale reduction, 207
 - map projection co-ordinates, 46, 47
 - map projection plane as computation
 - surface, 221
 - map projection science, 54
 - map projection zone, 58, 62
 - KKJ, 58
 - municipality, 62



- map scale, 18, 61
 - nominal, 62
 - on central meridian, 62
- map scale number, 61
- map sheet (x, y), 46
- map-sheet division, 61, 213
- marble, glass, 429
- mareograph, 14, 492
 - description*, 509
 - co-location, 509
 - spatial sampling, 509
- mareograph pillar, 510
- Mariana Trench, 220
- Mars, Olympus Mons, 219
- Martikainen, Martti, 18
- Masala, Kirkkonummi (Finland), 216
- mass
 - heavy, 425, 426
 - inertial, 425, 426
- mass flow, under Earth crust, 492
- Master Control Station (GPS), 289
- MATLAB (software), 36, 517
 - statistical testing, 403
- matrix
 - definition*, 513
 - addition and subtraction, 513
 - column, 515, 516
 - inversion, 517
 - main diagonal, 516
 - multiplication, 513, 515
 - with a constant, 514
 - orthogonal, 274
 - row, 515, 516
 - singularity, 518
 - transpose, 516
- Mauerpertuis, Pierre L. M. de, 1, 8, 185
- Maxwell, James Clerk, 188, 189
- Maxwell–Faraday equation, 522
- mean anomaly, 526
- mean error, 28
- mean error of unit weight, 317, 376
 - pseudo-range, 315
 - station adjustment, 167
- mean height of satellite, 525
- mean sea level
 - geopotential, 441
 - global, 510
- mean sea surface, global, 438
- measure (mathematics), 34
- measured quantity, 23, 24
- measured value, 28
- measurement base of project, 245
- measurement class (zoning), 235, 236
- measurement error, 28
- measurement plan, 18, 21
- measurement planning, 401
- measurement unit, 23, 24
- measuring pole (GPS), 298
- measuring tape, 126
 - calibration, 186
 - of steel, 44
 - sag, 186, 187
 - temperature correction, 186
- measuring telescope, 98, 100
 - levelling instrument, 97
 - adjustment screw, 105
 - tasks, 99
 - theodolite, 122
- measuring wire, precision, 44
- mechanics, classical, 5
- Mekometer (range-finder), 44, 196
- Mercator (map projection), 54, 221
- meridian circle, 449, 461, 467
 - graduation circle, 461
 - refraction correction, 461
- meridian transit, 449, 461, 467
- metadata, 125
 - definition*, 20
 - concept, 231
 - of theodolite measurement, 149
 - recording, 231
- metal detector, 248
- meteor observation, 55
- meteorology, 308
- metre, 185
 - public, Paris, 24
 - SI definition, 26
- metric correctness, 446
- metrology, 23, 247
- Metsähovi radio telescope, 483
- Metsähovi research station, 78, 79, 86, 354
 - aerial image, 117
 - benchmark, 78
 - co-ordinates, 49, 118
 - EUREF-FIN co-ordinates, 48
 - VLBI, 483
- microgal (μGal), 426
- micrometer, optical, 124, 139
 - levelling, 98, 109
- microwaves and water vapour, 203
- Mid-Atlantic Ridge, 485, 487
- Mid-Indian Ocean Ridge, 487
- Mid-Pacific Ridge, 487
- Milanković cycle, 504
- Milanković hypothesis, 503
- Milanković, Milutin, 503
- milligal (mGal), 426
- mine surveying, 170, 212, 247
- minute (angle unit), 26
- mistake, human, 29
- Misuzawa (Japan), 467
- Mitteleuropäische Gradmessung*, 10



- mobile mapping system, 223
- Moby (musician), 235
- model error, 398
- moisture, effect on levelling staff, 109
- monitoring measurement, 246
- Mont Valérien (Suresnes, France), 195
- Montmartre (Paris), 195
- monument-type choice, 125, 126
- Moon
 - attraction, 469
 - phases, 188
- mortgage, 16
- motorway
 - asphalt laying, 247
 - planning, 244
- MRI, magnetic resonance imaging, 469
- MSAS, 351
- multipath, 298, 319
- municipal technology, 237
- municipality
 - local co-ordinates, 66
 - planning, 17
 - utility lines, 247
- Munk, Walter, 479
- Mylar (PAGEOS), 464
- N**
- N60, 93
 - datum point, 78
- N2000, 279
 - description*, 93
 - datum point, 78
 - fundamental benchmark, 79
- nabla, 428
- naïve world model, 1
- nanogal (nGal), 426
- NAP (height system), 78, 93
- NASA (USA), 465
- National Geodetic Survey (NGS, USA), 328
 - forward geodetic problem, 68
 - inverse geodetic problem, 86
- National Geospatial-Intelligence Agency (NGA, USA), 439
- National Map Grid Co-ordinate System, *see* KKJ
- National Oceanic and Atmospheric Administration (NOAA, USA), 252
- natural-gas extraction, 114, 418, 492
- nature, data on, 233, 234
- navigation, 54, 344
- navigation message, GPS, 291–293, 327, 469
- Navy Navigation Satellite System (NNSS, “Transit”), 45, 286, 287
 - description*, 284
 - Doppler shift, 286
 - submarine computer, 287
- nearsightedness, 101
- network adjustment, 359, 362
 - in map co-ordinates, 222
 - in map plane, 221, 222
 - joint European, 11
 - three-dimensional, 218
- network densification, 111, 214
- network hierarchy
 - concept*, 214, 215
 - base-network measurement, 211
 - GNSS, 212
 - levelling, 96, 111
 - significance, 215
- network RTK archive service, 353
- new degree, new minute, new second, 27
- Newton, Sir Isaac, 5
 - dot notation, 281
 - Earth flattening, 13
 - gravitation, 425
 - nature of light, 188
 - Principia*, 5
- Nivavaara, 1, 9
- NLS, National Land Survey of Finland
 - co-ordinate transformation service, 86, 87
 - levelling, 96
 - terrain models, 251, 252
- non-centrality parameter of χ^2
 - distribution, 404
- non-Euclidean geometry, 358
- nonius, 140
- Nordic gravity profiles, 492
- normal distribution, 34
 - expectancy, 34
 - inflection points, 35
 - mean error, 34, 35
 - multivariate, 396
 - probabilities, 35
 - properties, 34
 - statistical testing, 406, 410, 411
 - two-dimensional, 39
 - two-sided test, 405
- normal equation, 364
 - as system of equations, 365
 - name, 364
 - solution, 364
- normal gravity, 430
 - calculation, 430
 - GRS80, 431
 - plumb-line mean, 445
 - strength (scalar), 430
 - vertical gradient, 451
 - GRS80, 431



- normal gravity field, 430, 431
normal gravity vector, 430, 434
normal matrix, 365
 azimuth symmetric, 319
 calculation of average, 366
 Helmert transformation, 390
 linear regression, 368
 example, 370
 pseudo-range, 314
normal potential, 430
 GRS80, 431
normal section, 12
Norsk Rikskringkasting, 235
North Star, 470
northing, 59, 65
no-shadow orbit, 477
nuisance parameter, 163
null hypothesis, 400, 410
numerical weather prediction (NWP), 256, 466
Nummela (Finland), 194
Nummela baseline, 185, 194
Núñez, Pedro, 140
Nunn, Joseph, 465
nutation, 469, 470
NUVEL (plate-tectonics model), 487
- O**
- observation equation, 363, 398
 azimuth, 386
 calculation of average, 366
 carrier phase, 304, 338
 formation, 363
 GNSS, 500
 Helmert transformation, 389–391
 horizontal direction, 386
 linearisation, 374
 linearised, 375, 382
 non-linear, 375
 pseudo-range, 302, 309, 313
 redundancy, 363
 slant range, 384
 standard form, 366
 station adjustment, 163, 164
 tachometer measurement, 388
 VLBI, 482
 zenith angle, 388
observation errors, vector of, 396, 398
observation function, 375
observation-station catalogue, 461
Observatory Hill, Helsinki, 78
ocean current
 map, 498
 thermal energy transport, 508
 transversal tilt, 508
 velocity vector, 508, 509
octave (software), 393, 517
Ölander, Victor Rafael, 182
“*The Old Survey*”, song, 119
Ollikainen, Matti, 280
Olympus Mons (Mars), 219
Omega (navigation system), 284
one-minute theodolite, 138
one-second theodolite, 138
on-the-fly (RTK), 348, 349
optical activity, 189
optical lattice clock, 95
orbit determination, 362
orbital ellipse
 eccentricity, 525
 focal point, 474
 semi-major axis, 474, 476, 525
 semi-minor axis, 525
orbital inclination, 525
 critical, 478
 mapping area, 475, 476
orbital period of satellite, 476
orientation unknown, 163
Orion (constellation), 458
Ørsted (satellite), 502
orthogonality of pseudo-random codes, 294
orthonormal basis, 519
orthophoto map, 256
ortho-rectification, 256
oscillation period of Cs atom, 26
outlier detection, 395, 405, 406
overall validation, 402
over-determination by observations, 398
ozone layer, 502
- P**
- P code, 292, 293, 295
 encryption, 297
 measurement accuracy, 297
 wavelength, effective, 297
PAGEOS (satellite), 464, 465
palaeomagnetism, 486
Pantometria (book), 119
paper machine, 246
paper, printed maps, 47
parallax of measuring telescope, 100, 101
parallel shift of forced-centring device, 130
parallelepiped, 374
parallelogram, 374
parasol, 125
parcel boundary, 245
parcel division, binding, 236
partial derivative (linearisation), 312
Pasteur, Louis, 189
pathological function, 263
pendulum clock, 13
pendulum compensator, 105



- principle of operation*, 105, 107
- magnetism, 106
- pentagon prism, 116
- perigee, 525
- perigee height, 476
- perigee, argument of, 525
- periglacial bulge, 488, 509
- perturbance in receiver location, 310
- Peru, 8
- petroleum extraction, 114, 418, 492
- phase error of signal, 136, 137
- phase index of refraction
 - dry air, 202
 - ionosphere, 306
- phase measurement, electronic, 196, 303
- phase modulation, 291, 293
- phase quadrature, 297
- phase velocity, 305, 306
- photodiode, 149
- photogrammetry
 - description*, 43
 - earthwork, 260
 - terrain model, 254
- photon, 189
- photosphere, 470, 471
- physical geodesy, 15, 46
 - definition*, 90
- pillar, for instrument, 128
- Pisa, tower of, 114
- Pisces (Fishes, constellation), 459
- pitch, viscosity, 489
- Pittacus of Mytilene, 23
- pixel graphics, 255
- plan (map), 53, 83
- Planck, Max, 189
- Planck's constant, 189
- plane angle, 24
- plane co-ordinates, 47, 66
 - as complex numbers, 419
 - in geodesy, 65
 - rectangular, 65
- plane surveying, 15
- plane table, 42, 43
- planimeter equation, 259
- planimeter, polar
 - principle of operation*, 259
 - image, 259
- planning
 - of GPS measurements, 312, 314, 333
 - of infrastructure, 17
 - of land use, 17
 - of measurement networks, 414
 - of roads and railways, 243
 - of the built environment, 17
 - technical, 17
- planning and measurement process, 401
- plasma, 521
 - name, 523
- plasma frequency, 306
- plasticity, 489
- plate boundary, strike-slip, 492
- plate co-ordinates, 463
- plate tectonics, 85, 483, 487
 - mechanisms, 488
- Plesetsk (Russia), 494, 497
- plumb line
 - definition*, 120
 - curvature, 121
- plumb-line deviation
 - definition*, 276
 - effect on azimuth, 278
 - equations, 275
 - figure, 276
- plumb-line direction
 - astronomical determination, 10, 276, 448
 - variation, 448
- plummet
 - optical, 124, 130, 131, 134, 135
 - adjustment, 134
 - checking, 134, 135
 - index, 130
 - separate, 133
 - precision, 128
 - rod, 129, 130
 - string, 124, 128, 129
- Poder, Knud, 57
- Poincaré-Prey reduction, 444
- point cloud, 254
- point data base, 21
- point description, 126
- point mean error, 316, 382
- point number, 126
- point precision, 382
- point variance, 382
- point-grid presentation, 254, 255
- polar amplification, 505
- polar drift, 466
- polar motion, 465–467
 - annual, 465
 - Chandler wobble, 466
 - monitoring, 466, 483
- polarimeter, 189
- Polaris (star), 5, 470
- polarisation, 189, 191
 - figure*, 192
 - and life, 189
 - circular, 299
 - elliptical, 191
- Porvoo (Finland), 392
- position determination, 462
 - astronomical, 4, 448, 449



- position vector
 of observation station, 308, 309
 of receiver, 310
 of satellite, 308–310, 326, 328
- positioning error, 340, 341
- potassium–argon, 486
- potential energy, 90, 439
- Pothenot, Laurent, 160
- power of a test, 411
- PP2000 (benchmark), 78
image, 79
- PPP, precise point positioning, 353
- pre-amplifier of GPS antenna, 293, 297
- precession, 469
 apsidal, 505
 astronomical, 504
 climatological, 505
 Larmor, 469
 of the equinoxes, 459, 469, 470
 satellite nodal, 477
 satellite orbital, 11
- precise ephemeris, 328, 329, 331
 quality, 341
- precise levelling
definition, 98
 Finnish national, 78
 of Finland, first, 111
 of Finland, second, 111
 repeated, 491
- precise-levelling network, Finnish
 national, 96, 111, 510
- precise-levelling staff
 invar staff, 109
 spring, 109
 staff unit, 109
- precision behaviour of a network, 81
- precision farming, 247
- precision theodolite, 138
- precision vs. accuracy, 28
- primordial heat, 487
- Principia*, 5
 Earth flattening, 6
- probability
 of a die cast, 31
 of a value interval, 33
- probability density distribution, 32
 as limit of histograms, 33
 three-dimensional, 40
 two-dimensional, 39
- problem situation, field-work, 132
- Procyon (star), 458
- profile measurement, 114, 115
- projected co-ordinate reference frame, 58
- projection
 onto map plane, 221
 onto reference ellipsoid, 221
- projection measurement (Väisälä), 193, 194
- propagation
 general, of variances, 377
 of errors, 37, 38
 of expectancies, 376
 of variances, 38, 41, 376
- propagation delay modelling, atmospheric, 351
- property
 monetary value, 16, 17
 owner rights, 17
- property formation, 236
- proxy, 294
- pseudo-observation, 170
- pseudo-random code, 291, 293, 294
- pseudo-range, 295, 302, 309, 340
 equivalent, 304, 305
 linearisation, 312, 313
 observation, 295, 303
- pseudo-range offset, 345
- public debate, 256
- puff pastry (metaphor), 521
- pulsar, 523
- pulse shape, 507
- Q**
- Q_{ll} metric, 398, 402, 405
- quadrant of the plane, 65, 66, 70
- quadrature
 alternatives, 262
 Simpson's rule, 260
- quadrature (modulation), 297
- quality
 of benchmark set, 214
 of broadcast ephemeris, 341
 of detail survey, 228
 of measurement results, 28, 148
 of precise ephemeris, 341
- quantum theory, 189
- quartz gauge, 192
- quartz oscillator drift, 338
- quasar, 480, 482
- quasi-geoid, 445
- quasi-two-dimensional network geometry, 218
- quaternion, 265
- R**
- R (software), 36
- radial survey, 225–227
figure, 226
 as a numerical mapping method, 226
 horizontal-angle measurement, 158
 metadata, 227
 setting out, 237, 238



- slope reduction, 226
- radian, 26, 27
- radiation belt, 289, 473
- radio navigation, 284
- radioactivity, 486, 487
- radiometric dating, 486
- radio-traffic disruption, 473
- rail shoe, 111
- railway
 - centrifugal force, 244
 - design speed, 244
- railway levelling, 112
- random-point method, 253
- range-finder, 44, 194
 - calibration, 199
 - constant correction, 201
 - constant error, 199, 200
 - construction, 195
 - damage, 125
 - drift, 199
 - frequency calibration, 199
 - frequency correction, 201
 - frequency error, 199, 200
 - determination by frequency standard, 200
 - hand-held, 197
 - instrumental errors, 198
 - scale error, 199
 - zero-point error, 199
- ranging rod, 223
- rapid orbits, 330
- rapid prototyping language, 333
- reading device, graduation circle, 123, 138
- reading microscope, 138, 141
 - eyepiece, 123
- real estate, monetary value, 16, 17
- real number 1, 516
- real numbers, in computers, 32, 302
- real-time kinematic measurement, *see* RTK
- real-time positioning, 344
- real-timeness, 344
- reconnaissance, 212
- reconnection (magnetism), 473
- rectangle rule (quadrature), 262
- red lead, 126
- reduction of observations, 29
- reduction to reference surface, 13, 219
- redundancy, 357, 398
 - meaning, 417
- redundancy matrix, 415
- redundancy, degree of, 417
- reference ellipsoid, 48, 84
 - as computation surface, 219
 - as equipotential surface, 430
 - centre shift, 273, 275
 - choice, 276
 - goodness of approximation, 220
 - surface normal, 49
- reference sequence (tree rings), 294
- reference sphere, goodness of approximation, 219
- reference surface, 93, 218, 220
 - computational, 183
- reflective panel on Moon, 480
- reflector, 197
 - zero-point error, 200
- refraction coefficient k , 179, 204
- refraction modelling, atmospheric, 182
- regional plan, 235
- rejection bound, 410
 - choice, 409
 - normal distribution, 411
 - two-sided, 406
- rejection of a hypothesis, 410
- relativity
 - general, 95, 426
 - special, 307
- reliability, 395, 398, 412
 - example, 413, 414
 - exterior, 404, 416
 - interior, 404, 416
- reliability and precision, 414, 416
- resection, 158–160
 - three-dimensional, 322
- reservoir dam (seismics), 492
- residual \neq error, 360
- residuals, 397
 - final, 168
 - first, 168
 - in quality control, 398
 - properties, 398
 - size, 398
 - sub-space of, 364
 - weighted, 405
- reticule, 99
- reversed curve, 242
- reversion staff, 110
- rhumb line (loxodrome), 54
- Richer, Jean, 13
- ridge, mid-ocean, 484, 487, 493
- right angle, 26, 27
- right ascension, 460
 - determination, 461
- right-angle survey, 223, 224
 - check, 224
 - setting out, 239
 - tools, 223
- right-hand side vector
 - calculation of average, 367
 - linear regression, 368



- RINEX, Receiver-Independent Exchange Format, 300, 302, 353
- road
- centre line, 243
 - centrifugal force, 244
 - curvature, 243
 - radius of curvature, 244
 - structures, 245
 - transversal tilt, 243, 244
- road plan, 244
- roadbuilding, 244
- rod plummet, 129, 130
- roede (length unit), 3, 4
- root, mountain, 449
- rounding of corners, 241
- circular arc, 240
 - compound curve, 241
- rounding residue, 143
- rover location (RTK), 348
- row vector, 514
- RTK, 347
- principle of operation*, 348
 - base-network measurement, 218
 - detail survey, 228
 - essence, 349
- rubidium clock, 287
- rugby ball (figure of the Earth), 7
- rust-protection paint, 126
- S**
- S waves, 484
- saddle point, 428
- Safety of Life (SoL), 291, 351
- sag correction, measuring tape, 186, 187
- Sagan, Carl, 481
- Sagittarius (Archer, constellation), 459
- Sagittarius A*, 480
- sample, 35
- sample mean, 35
- sample mean error, 360
- sample variance, 35
- San Andreas fault (California, USA), 492
- satellite geodesy, 265
- co-ordinate reference system, 265
 - geopotential model, 438
 - images (NASA), 465
 - orbit perturbations, 475
 - positioning, 45
- satellite laser-ranging (SLR), 45, 480
- satellite orbit choice, 475
- satellite orbital motion, 473
- satellite orbital period, 474
- satellite positioning
- base-network measurement, 212, 214
 - geodynamics, 479
 - of sea surface, 510
 - widespread use, 57
- satellite radar altimetry, 498, 506
- technique*, 506
 - ice sheet, 505
 - measurement geometry, 506
 - on-board GNSS, 509
- satellite-based augmentation systems (SBAS), 351, 352
- satellite-to-satellite tracking (SST), 497
- saturation (water vapour), 500
- saturation partial pressure, 500, 501
- SBAS differential correction, 351
- scalar product, 519
- scale height, 502
- scale microscope, 124, 138
- scale model, of network, 3
- scale reduction
- Gauss–Krüger (map projection), 207
 - UTM (map projection), 208
- scattered settlement area, 236
- Schmidt camera, 464
- Schmidt, Bernhard, 463
- Schmidt–Väisälä telescope, 463
- scilab (software), 517
- screw thread, geodetic standard, 124, 298
- S-curve, 242
- sea currents, geophysical modelling, 95
- sea level
- monitoring, 491
 - variation, 493
- sea-floor mapping, 14
- sea-floor pressure, 496, 497
- sea-floor spreading, 486
- sea-level rise, 509
- Seasat (satellite), 511
- sea-surface topography, 498, 508
- definition*, 14
 - causes, 507
 - GOCE, 499
 - magnitude, 508
 - observed, 509
 - theory, 507
- second (angle unit), 26
- second (time unit), 26
- SI definition, 26
- selective availability (SA), 345
- self-calculating staff, 110
- Sendai (Japan), 493
- sensitivity analysis, 310
- sensitivity of observations, 309, 311
- sensitivity of spirit level, 102
- setting out, 236
- building location, 245
 - process description, 237
 - radial survey, 237, 238
 - site reference point, 246



- zoning plan, 236
- setting-out measure
 - pre-calculated, 236
 - radial survey, 238
 - straight line, 239
 - surface-area calculation, 256
- sewer, functioning, 114, 246
- SFS 3161, utility-line mapping standard, 247
- shifting variate (\mathcal{E})
 - term*, 403
 - expectancy
 - alternative hypothesis, 404
 - null hypothesis, 403
- shimmer (levelling), 112
- shipyard, 246
- shoobox model, 83, 89
- shoelace formulas, 258
- shore area, 236
- Shuttle Radar Topography Mission (SRTM), 252
- SI system, 24
 - accepted unit, 25
 - base unit, 24
 - derived unit, 24
 - prefix, measurement-unit, 24, 25
 - unit symbols, 23
- SI unit
 - of distance, 185
 - of geopotential, 441
- sidereal clock, 458
- sidereal day, 457
- sidereal time, 457, 460
 - Greenwich, 266
- sight axis
 - adjustment, 142
 - levelling instrument, 97, 101
 - realisation, 142
 - theodolite, 121
- sighting azimuth, 276
- sighting direction projected onto the horizontal, 277
- signal
 - good, 136
 - problem, 136
 - with prism, 198, 199
- signal height, 183
- significance level, 411
 - link, harmonisation, 412, 413
 - normal distribution, 411
 - two-sided, 406
- silicone oil in compensator, 148
- similarity transformation, *see* Helmert transformation
- Simpson, Thomas, 260
- Simpson's rule, 260, 262
- Kepler, 251, 260
 - proof, 260
 - volume calculation, 261
- singularity
 - of a mapping, 374
 - of resection, 160
- Sirius (star), 458
- site plan, 246
- Sjökulla (Kirkkonummi, Finland), 483
- ski piste, 256
- sky plot, 313
- sky survey, 465
- slant-range measurement, 382–385
- slope percentage, 187
- slope reduction, 187, 188, 206
 - radial survey, 226
- SmartNet (Hexagon Oy), 353
- Snell van Rooyen, Willebrord (Snellius, Snell), 3, 160
- Sodankylä (Finland), 336
- Sodankylä Geophysical Observatory, 86, 354
- sodium, viscosity, 489
- solar eclipse, total, 471
- solar flare, 473
- solar maximum, 473
- solar wind, 471
- solid Earth
 - elasticity, 493
 - mass distribution, 475
 - motions, 331
 - tide, 85
- solid-angle element, 437
- solstice, 459
- solution space, 364
- SP3 (ephemeris format), 328, 329
- space geodesy, 85, 265
- space plasma, 521
- space segment (GPS), 289
- space weather, 473
- space-time
 - curvature tensor, 426
 - geometry, 426
- spatial correlation, 336
- spatial planning, 17
 - definition*, 17
 - figure, 19
 - topographic surveying, 17
- speed of light, 189, 201
 - definition*, 195
 - determination, 194
- spherical excess, 358
- spherical-harmonic coefficient, 476
- spider silk, 99
- spike (measurement), 112
- spin of a particle, 191



- spinning encoding circle, 151, 152
 spirit level in levelling instrument, 97
 spring balance, 426
 Sputnik 1 (satellite), 11
stadium (length unit), 3
 staff (levelling), *see* levelling staff
 staff calibration, 108
 staff distance, 107, 112
 staff support, 110, 111
 staff unit, 110
 standard
 of construction surveying, 245
 of geodetic works, 213
 standard deviation, 28
 standard ellipse, *see* error ellipse
 standard uncertainty, 28
 standing axis (theodolite), 121
 star catalogue, 461
 star programme, 449
 static base-network measurement, 218
 station adjustment, 162
 calculation, 164
 template, 169
 defect, 169
 degrees of freedom, 163, 170
 design matrix, 163
 error model, 167
 in observation notebook, 162
 observation equations, 163, 165
 redundancy, 163
 residuals, 166
 shifting variate (\mathcal{E}), 167
 summary, 167
 steel grid mast, 118
 stellar triangulation, 43, 462
 equipment, 462
 flashtube, 462, 464
 high signal, 463
 measurement geometry, 463, 464
 network adjustment, 464
 plate reduction, 464
 using satellites, 464
 weather balloon, 462
 steradian, 26
 stereo model restitution, 43, 44
 sticker, reflective, 198
 stochastic quantity, 30
 linear combination, 37
 multi-dimensional, 39, 40
 realisation, 30, 31
 real-valued, 32
 two-dimensional, 33
 Stokes equation, 436, 437
 computation point, 438
 integration point, 438
 Stokes function, 437
 Stokes kernel, 437
 Stokes loop integral theorem, 522
 Stokes, George Gabriel, 436
 straight angle, 26
 straight line
 fit, 361, 367
 setting out, 239
 straight setting-out method, 239
 stratification, atmospheric, 318
 stratosphere, 502
 strength of network, 412
 stress-energy tensor, 426
 string plummet, 124, 128, 129
 Struve chain, 8, 10
 Struve, Georg Wilhelm von, 8
 subduction, 435, 487
 sum of squares
 of plumb-line deviations, 276
 of residuals, 398
 Sun
 convection, 470
 convection layer, 470, 471
 core thermal energy, 470
 corona, 471
 heating, 471
 magnetic field, 471, 472
 lines of force, 473
 topology, 473
 nuclear power, 470
 plasma, 471
 X-rays, 470
 Sunda deep-sea trench, 435
 sunspot, 472
 Sun-synchronous orbit, 477
 superconductivity, 471, 521
 superconductor, 522
 surface area
 calculation, 256
 with co-ordinates, 257, 258
 with setting-out measures, 256, 257
 measurement by planimeter, 259
 Survey of India, 10
 surveying science, 15
 Syene (Egypt), 3
 symmetric GPS measurement geometry, 316
 synoptic scale, 336
 synthetic-aperture radar (SAR), 494, 505
 terrain model, 252
 system calibration, 107, 110
T
taaksepäin leikkaus, 158
 tablet (computer) in detail survey, 229
 tabloid press, 282



- tacheometer
term, 45
 electronic, 45
 observation equations, 388
- tape correction, 186
- tartaric acid, organic, 189
- Tasmania (Australia), magnetic pole, 485
- Taurus (Bull, constellation), 459
- Taylor series expansion (linearisation), 312, 371, 375
- TCA2003 (theodolite), 152, 153
 automatic target recognition (ATR), 154
 data exchange, 155
 distance reduction, 154
 laser plummet, 154
 level, electronic, 154
 monitoring measurement, 154
 self-calibration, 154
- tectonics, 418
 intra-plate, 493
- telescope
description, 99
 invention, 119
- Tellurometer (range-finder), 44
- terminology work, 213
- terrain attraction, 447
- terrain correction (distance), 205
- terrain correction (gravity), 256
- terrain model
 applications, 255
 measurement geometry, 253
 measurement technology, 254
 presentation form, 255
- terrain presentation, 255
- testing strategy, 411
- testing, statistical, 28, 29, 400
 deformation, 419, 494
 gross errors, 400
 hypotheses, 400
 normal distribution, 410
 outlier detection, 395
- theodolite
image, 120
 axes, 121, 122, 127
 axis errors, 142
 base, 124
 calibration, 142
 carrying case, 125
 circles, 121, 122
 classification, 137
 construction, 123
 electronic, 148–152
 gyro-, 269, 386
 handling, 124
 in mapping, 42
- instrumental errors
 horizontal-angle measurement, 142
 zenith-angle measurement, 148
- invention, 119
- measurement, 119
- sighting and targeting, 135
- thermal expansion coefficient
 invar, 44, 109, 185
 steel, 186
- thermal radiation spectrum of black body, 189
- thermodynamics, arrow of time, 307
- three-axes stabilisation, 289
- three-sigma rule, 29, 36, 411
- tidal force field, 451, 453
- tidal loading, 493
- tide gauge, *see* mareograph
- tie measurement of a profile, 116
- tie-in survey, 224
 check, 225
- tiling of images, 255
- tilt meter, digital, 148
- time (co-ordinate), 46, 84
- time (quantity), 26, 468
- time determination, 461
- time difference measurement, 285
- time, precise, dissemination of, 286
- time-zone system, 48, 468
- Tobler, Waldo R., 335
- toise* (length unit), 8
- Tom and Jerry, *see* GRACE
- top (spinning), 469
- TOPEX/Poseidon (satellite), 479, 509, 510
 orbital inclination, 509
- topographic attraction, 447
- topographic data
 catalogue, 231
 classification, 233
 encoding, 231–233
- topographic information system, 230
 national, 231
- topographic surveying, 16, 18, 19, 21
 computation, 229
 data, 231
 end products, 20
 map, 18
 in spatial planning, 17
 metadata, 231
 skills, 18
 software, 229
 tasks, 17
 terrain model, 254
 work volume, 17
- Topographic Surveying and Mapping*
 (booklet), 16
- Torge, Wolfgang, 15



- Torne river (Sweden, Finland), 42, 185
valley, 8, 78
- torsion balance, 452
- total station, 45
automatic, 493
- traceability (metrology), 185, 186, 247
- traffic route planning, 90, 256
- trajectory (road)
calculation, 245
setting out, 245
- transformation formula, local (RTK), 350
- transformation parameters
EUREF89 to ED50, 280
Helmert transformation, 73, 74, 389
ITRF2005 to ETRF2005, 281
- Transit 5B-5, 283
- Transit positioning system, *see* Navy
Navigation Satellite System
- transit through meridian, 449, 461, 467
- transition curve, 243
- transpose, 516
of a vector, 516
saving paper, 42, 516
- transversal wave motion, 484
- trapezoid rule (quadrature), 262
- travel time (GPS signal), 294
- traverse
definition, 170
auxiliary points, 171
calculation in the plane, 179
closed, 170, 171, 174, 175
closing point, 171, 178
computation, 170, 172
computation template, 178
co-ordinate adjustment, 176
directions, adjustment of, 175
horizontal-angle measurement, 157
incompletely closed, 171
least-squares adjustment, 178
open, 170–172
starting direction, 172
starting point, 171, 178
straight, 178
weighting, 177
- traverse levelling, 95, 111, 112
- traversing, 170, 215
- tree line, 294
- tree ring, 293, 294
correlation, 294
- trend in time, 40
- triangle condition, 29, 358
- triangulated affine transformation, 62, 64
- triangulated irregular network (TIN), 255
- triangulated-network presentation, 254,
255
- triangulation, 3
principle, 3
Finnish national, 57
first order, 215
horizontal-angle measurement, 157
lower order, 215
method, 42
of Snellius, 3
- triangulation network, 359
adjustment, 42
continental, 11
intercontinental, 465
principal side, 185
triangle, 358
- trigonometric heighting, 180
fundamental equation, 181
zenith angle, 180, 181
- Trimble AB, 44
- Trimnet VRS, 352
- tripod head, central hole, 128
- tropic, 459
- tropopause, 502
- tropospheric propagation delay, 302, 303,
305, 307
- tropospheric research, 308
- true anomaly, 526
- true value, 28, 35
of an observed quantity, 34, 396
- truncated co-ordinates, 61, 67
- trunnion axis (theodolite), 121
- trunnion-axis tilt, 144, 145
definition, 121
determination, 144
- truth, formal (datum), 81, 82
- tubular level
figure, 102
adjustment screw, 101, 103
construction, 101
horizon, 97, 101
task, 101
- tunnel surveying, 158, 170, 212, 247
- tunnels, network of (metaphor), 92
- Tuorla, Wizard of, *see* Yrjö Väisälä
- Turku (Finland), 78, 79, 462
- Turku Cathedral, 79
- Tuusula (Finland), 57
- U**
- UDP/IP protocol (VLBI), 481
- Ukiah (California, USA), 467
- ultra-rapid orbits, 330
- ultraviolet radiation (PAGEOS), 464
- UNAVCO
geoid model, 118
plate motion calculator, 488
- uncertainty, 30
type A, 30



type B, 30
 uncertainty ellipse, *see* error ellipse
 unit element of a set of numbers, 514
 unit matrix, 514, 516
 Universal Time Co-ordinated (UTC), 468
 University of Helsinki, 15
 uranium, depleted, 480
 uranium–lead, 486
 urban canyon, 170, 223
 urban landscape, 158
 US Geological Survey (USGS, USA), 252
 “use point”, 245
 user segment (GPS), 290
 utility line, 247
 map, 247
 mapping methods, 248
 utility-lines marking service, 249
 UTM (map projection), 61
 description, 62
 conformality, 62
 scale distortion, 62, 63

V

Vaakitsijantie, Oulu, Finland, 89
 Väisälä interferometry, 191, 193
 Väisälä, Yrjö, 43, 191, 457
 value set, 31
 Vanguard 1 (satellite), 11
 Vantaa (Finland), 352
 variance, 35, 376
 variance matrix, 40, 376
 definition, 40
 of co-ordinate displacements, 423
 of co-ordinates, 315, 316
 of GPS solution, 315
 of observations, 365, 397
 of solution, 397
 variance of unit weight
 a posteriori, 402
 pseudo-range, 315
Vasa, the good ship, 114
 vector graphics, 255
 vector mapping, 373, 374
 vectorial product, 519
 Vega (star), 470
 velocity correction, second, 204, 205
 velocity vector of satellite, 326, 328
 Venezuela, 492
 Venice (Italy), 114
 Vening Meinesz, Felix A., 425, 435
 vernier, 140
 Vernier, Pierre, 140
 vertical angle, *see* zenith angle
 vertical axis (theodolite), 121
 vertical datum, 80
 vertical-circle index, 145

automatic, 148
 video game, 256
 video tape (VLBI), 481
 virtual reference station (RTK), 352
 viscosity, 489
 pitch, 489
 sodium, 489
 water, 489
 visibility
 free stationing, 227
 in the terrain, 256
 of GPS satellites, 284
 visibility conditions of benchmarks, 212
 visible light, nature of, 188
 visual ellipse of point variance matrix,
 380, 381
 visual ellipsoid of gravitational gradient
 tensor, 451
 visualisation of spatial data, 234
 VLBI
 description, 480
 Earth rotation monitoring, 468
 Mark II, 481
 measurement geometry, 482
 observable, 481
 observation campaign, 480
 principle of operation, 482
 vector solution, 482
 Vocabulary of Geoinformatics, 214
 volcanism, 488
 volume calculation
 Simpson’s rule, 260
 terrain model, 263
 VRS-RTK, 352
 VVJ (Helsinki system), 57
 map projection, 61

W

W code, 292, 297
 WAAS, 351
 wall measure, 245, 246
 Washington DC (USA), meridian treaty, 48
 water gauge in interior waters, 95
 water molecule, chemical polarity, 203
 water vapour
 content of air, 307
 integrated, 502
 total precipitable, 502
 in stratosphere, 502
 partial pressure, 308, 500, 501
 wave function, complex, 188
 wave motion
 frequency, 189
 on open sea, 509
 phase, 188, 189
 wavelength, 189



- wave packet, 306
 - wave-motion theory of light, 188
 - weather observations, representativeness, 204, 206
 - Wegener, Alfred, 483
 - Wegener's continental-drift theory, 484
 - weight matrix of unknowns, *see* normal matrix
 - weight-coefficient matrix, 376
 - of observations, 396
 - of pseudo-ranges, 315
 - of residuals, 400
 - of unknowns, 365
 - West Siberian land uplift area, 491
 - WGS84, 58
 - what if... (reliability), 414
 - white noise (VLBI), 482
 - wide-area differential GPS (WADGPS), 351
 - wide-laning, 343, 350
 - wind drag, 507
 - wine-barrel formula, 251
 - wood dating (tree rings), 193
 - work, 45
 - path integral, 440, 441
 - world model, naive, 1
 - writing error in observation notebook, 29
- Y**
- Y code, 292
 - year ring, *see* tree ring
 - Young, Thomas, 189
- Z**
- 1462 Zamenhof, 457
 - Zamenhof, L. L., 457
 - Zeiss (Trimble) DiNi12 (levelling instrument), 107
 - zenith angle, 120
 - definition*, 120
 - figure, 121
 - measurement, 145, 146, 387
 - refraction, 179, 180
 - zenith propagation delay, 500
 - zenith tube, 449, 467
 - zero element of set of numbers, 514
 - zero error of level, 127, 128
 - zodiac, 458
 - zoning, 16
 - Zoning Base Map Guide, 213
 - zoning plan, 16
 - boundaries, 237
 - calculation, 237, 238
 - Zoning Survey Guide, 213
 - new (2014), 235
 - setting out, 236
 - topographic data classification, 233
 - zoning-plan types, 235



