Physical Geodesy

Martin Vermeer

martin.vermeer@aalto.fi
Preface

This book aims to present an overview over the current state of the study of the Earth’s gravity field and those parts of geophysics closely related to it, such as especially geodynamics, the study of the changing Earth. It grew out of two decades of teaching in Helsinki’s two universities: Helsinki University of Technology – today absorbed into Aalto University – and the University of Helsinki. As such, it presents a somewhat Fennoscandian perspective on a very global subject. Also the author’s own research, on gravimetric geoid determination, helped shape the presentation. While there exist excellent textbooks on all parts of what is presented here, he may still hope that this text will find a niche to fill.

Helsinki, August 1, 2016,

Martin Vermeer
martin.vermeer@aalto.fi

Acknowledgements

Thanks are due to the many students and colleagues, both in academia and at the Finnish Geodetic Institute, giving useful comments and corrections over the course of many years of lecturing this course both at the University of Helsinki and at the Helsinki University of Technology, today Aalto University.

Special thanks are due to the foreign students at Aalto University, who
forced me during recent years to provide an English version of this text. The translation work prompted a basic revision also to the Finnish text, which was long overdue as parts were written in the 1990s before the author had had the benefit of pedagogical training. Thanks are thus also due to the Aalto University’s pedagogical training programme.
# Contents

Preface iii

Contents v

List of Tables xiii

List of Figures xv

## 1. Fundamentals of the theory of gravitation 1

1.1 General ......................................................... 1
1.2 Gravitation between two masses ......................... 2
1.3 The potential of a point mass .............................. 4
1.4 Potential of a spherical shell ............................ 5
1.5 Computing the attraction from the potential ............ 7
1.6 Potential of a solid body .................................. 8
1.6.1 Behaviour at infinity ................................... 9
1.7 Example: The potential of a line of mass ............... 10
1.8 Equations of Laplace and Poisson ....................... 12
1.9 Gauge invariance ............................................ 13
1.10 Single mass density layer ................................ 14
1.11 Double mass density layer ............................... 15
1.12 The Gauss integral theorem .............................. 17
1.12.1 Presentation ............................................. 17
1.12.2 Intuitive description ................................. 17
1.12.3 The potential version of the Gauss theorem ....... 19
1.12.4 Example 1: a little box ............................... 19
1.12.5 Example 2: The Poisson equation for a sphere .... 20
1.12.6 Example 3: a point mass in a eight-unit cube ....... 21
1.13 Green’s theorems ................................. 21
1.14 The Chasles theorem .............................. 24
1.15 Boundary-value problems ....................... 26
1.16 What the boundary-value problem cannot compute .......... 27

2. The Laplace equation and its solutions .................. 29
   2.1 General ........................................... 29
   2.2 The Laplace equation in rectangular co-ordinates .... 30
   2.3 Example: The Laplace equation in polar co-ordinates ... 33
   2.4 Spherical, geodetic, ellipsoidal co-ordinates .......... 34
   2.5 The Laplace equation in spherical co-ordinates ....... 37
   2.6 Dependence on height ............................ 38
   2.7 Legendre’s functions ............................. 38
   2.8 Symmetry properties of the spherical harmonic expansion 43
      2.8.1 Dependence on latitude \( \phi \) .................. 43
      2.8.2 Dependence on longitude \( \lambda \) ............... 44
   2.9 The orthogonality of the Legendre polynomials .......... 44
   2.10 Spectral representations of various quantities ...... 45
      2.10.1 The potential ............................... 45
      2.10.2 Gravitation ............................... 46
   2.11 Splitting a function into degree constituents .......... 47
   2.12 Low-degree spherical harmonics ................... 48
   2.13 Often used spherical harmonic expansions ........... 50
   2.14 Ellipsoidal harmonics ........................... 51
      2.14.1 Equivalence of the Rapp formula and the ellipsoidal
             formula ....................................... 52
      2.14.2 Advantages of using ellipsoidal harmonics: ....... 53
      2.14.3 Disadvantage of using ellipsoidal harmonics: ...... 53

3. The normal gravity field ........................................ 55
   3.1 The basic idea of a normal field .................. 55
   3.2 The centrifugal force and its potential ............. 56
   3.3 Level surfaces and plumblines ..................... 59
   3.4 Natural co-ordinates ............................. 61
   3.5 The normal potential in ellipsoidal co-ordinates [difficult] . 62
   3.6 Normal gravity .................................. 64
   3.7 Numerical values and formulas .................... 66
   3.8 Example ......................................... 67
   3.9 The normal potential as a spherical harmonic expansion . 68
   3.10 The disturbing potential .......................... 69
   3.11 Exercises ...................................... 71
      3.11.1 Somigliana-Pizetti formula .................. 71
4. Anomalous quantities of the gravity field
   4.1 Disturbing potential, geoid height, deflections of the plumbline
   4.2 Gravity disturbances
   4.3 Gravity anomalies
   4.4 The boundary-value problem of physical geodesy
   4.5 The telluroid mapping and the “quasi-geoid”
   4.6 Free-air anomalies
   4.7 Exercises
   4.7.1 The spectrum of gravity anomalies

5. Geophysical reductions
   5.1 General
   5.2 Bouguer anomalies
      5.2.1 Calculation
      5.2.2 Properties
   5.3 Terrain effect and terrain correction
      5.3.1 Example: applying the terrain correction in a special case
   5.4 Spherical Bouguer anomalies
   5.5 Helmert condensation
   5.6 Isostasy
      5.6.1 Classical hypotheses
      5.6.2 Computation formulas
      5.6.3 Example: Norway
      5.6.4 The current understanding of isostasy
      5.6.5 Example: Fennoscandian land uplift
   5.7 Isostatic reductions
   5.8 The “isostatic geoid”

6. Vertical reference systems
   6.1 Levelling, orthometric heights and the geoid
   6.2 Orthometric heights
   6.3 Normal heights
      6.3.1 Molodensky’s theory
      6.3.2 Molodensky’s proof
      6.3.3 Normal height and height anomaly
   6.4 Difference between geoid height and height anomaly
   6.5 Difference between normal heights and orthometric heights
6.6 Calculating orthometric heights precisely .................. 120
6.7 Calculating normal heights precisely ......................... 122
6.8 Calculation example for heights ............................... 123
6.9 Orthometric and normal corrections .......................... 124
6.10 A vision for the future: relativistic levelling ................ 126

7. The Stokes equation and other integral equations 129
   7.1 The Stokes equation and the Stokes integral kernel ........ 129
   7.2 Plumbline deflections and the Vening Meinesz equations .... 131
   7.3 The Poisson integral equation ............................... 132
   7.4 Gravity anomalies in the exterior space .................... 135
   7.5 The vertical gradient of the gravity anomaly ............... 136
   7.6 Gravity reductions in geoid determination .................. 139
     7.6.1 The Molodensky method in linear approximation ...... 140
     7.6.2 The evaluation point as the reference level .......... 141
   7.7 The Remove-Restore method ................................ 142
   7.8 Kernel modification in the Remove-Restore method ........ 143
   7.9 Advanced kernel modifications .............................. 145
   7.10 The effect of the local zone ............................... 147

8. Spectral techniques, FFT 151
   8.1 The Stokes equation as a convolution ....................... 151
   8.2 Integration by FFT ......................................... 153
   8.3 Solution in rectangular co-ordinates ....................... 156
     8.3.1 The Strang van Hees method ......................... 156
     8.3.2 “Spherical FFT”, multi-band model .................. 157
     8.3.3 “Spherical FFT”, Taylor expansion model ............ 158
     8.3.4 “1D-FFT” ............................................. 160
   8.4 Complications; bordering, tapering ........................ 161
   8.5 Computing a geoid model with FFT .......................... 162
     8.5.1 The GRAVSOFT software ............................... 162
     8.5.2 The Finnish FIN2000 geoid ............................ 163
   8.6 Use of FFT computation in other contexts .................. 164
     8.6.1 Altimetry .............................................. 164
     8.6.2 Satellite gravity missions; airborne gravimetry ....... 164
   8.7 Computing terrain corrections with FFT ..................... 164
9. Statistical methods

9.1 The role of uncertainty in geophysics ...................... 169
9.2 Linear functionals ........................................ 170
9.3 Statistics on the Earth surface ............................ 171
9.4 The covariance function of the gravity field ............... 173
9.5 Least-squares collocation ................................. 175
9.5.1 Stochastic processes in one dimension ................. 175
9.5.2 Signal and noise ...................................... 176
9.5.3 The estimator and its error variance ................. 176
9.5.4 Showing optimality .................................. 177
9.5.5 The covariance function of gravity anomalies .......... 178
9.5.6 Least-squares collocation for gravity anomalies ....... 180
9.5.7 Calculation example .................................. 181
9.5.8 Theory of least-squares collocation ................... 183
9.6 Prediction of gravity anomalies ............................ 184
9.7 Covariance function and degree variances ................. 185
9.7.1 The covariance function of the disturbing potential .... 185
9.7.2 Degree variances and spherical harmonic coefficients 186
9.8 Propagation of covariances ................................ 188
9.8.1 Example: upward continuation of the potential ..... 188
9.8.2 Example: the covariance function of gravity anomalies 189
9.9 Global covariance functions ............................... 190
9.10 Collocation and the spectral viewpoint .................... 191
9.11 Exercises .................................................. 194
9.11.1 Hirvonen’s covariance formula and prediction .......... 194
9.11.2 Propagation of covariances ........................... 194
9.11.3 Underground mass points ............................. 195

10. Gravimetric measurement devices .......................... 197

10.1 History .................................................... 197
10.2 The relative (spring) gravimeter .......................... 198
10.2.1 Astatization ........................................... 200
10.2.2 Period of oscillation ................................ 202
10.2.3 Practicalities of measurement ........................ 203
10.3 The absolute (ballistic) gravimeter ......................... 204
10.4 Network hierarchy in gravimetry .......................... 207
10.5 The superconducting gravimeter ........................... 208
10.6 Atmospheric influence on gravity measurement .......... 210
10.7 Airborne gravimetry and GNSS ........................... 212
10.8 Measuring the gravity gradient ........................... 214
11. The geoid, mean sea level, sea-surface topography 217
   11.1 Basic concepts ........................................... 217
   11.2 Geoids and national height datums ...................... 218
   11.3 The geoid and post-glacial land uplift ................... 220
   11.4 Methods for determining the sea-surface topography ... 223
   11.5 Global sea-surface topography and heat transport ...... 224
   11.6 The global behaviour of sea level ....................... 226
   11.7 The sea-level equation .................................... 227

12. Satellite altimetry and satellite gravity missions 231
   12.1 Satellite altimetry ......................................... 231
   12.2 Crossover adjustment ....................................... 235
       12.2.1 Example ................................................. 238
       12.2.2 Global crossover adjustment ......................... 240
   12.3 Choice of satellite orbit ................................... 241
   12.4 Example ..................................................... 244
   12.5 Retracking .................................................. 246
   12.6 Oceanographic research using satellite altimetry ...... 247
   12.7 Satellite gravity missions ................................. 248

13. Tides, atmosphere and Earth crustal movements 253
   13.1 Theoretical tide ............................................ 253
   13.2 Deformation caused by the tidal force .................... 256
   13.3 The permanent part of the tide ............................ 259
   13.4 Loading of the Earth’s crust by sea and atmosphere .... 260

14. Earth gravity field research 261
   14.1 Internationally ............................................... 261
   14.2 Europe ........................................................ 262
   14.3 The Nordic countries ....................................... 262
   14.4 Finland ........................................................ 263
   14.5 Textbooks .................................................... 264

Bibliography 265
A. Field theory and vector calculus – core knowledge

A.1 Vector calculus .............................................. 271
  A.1.1 Scalar product .......................................... 271
  A.1.2 The scalar product, formally ......................... 272
  A.1.3 Exterior or vectorial product ......................... 272
  A.1.4 The vectorial product, formally ..................... 273
  A.1.5 Kepler’s second law .................................. 273
A.2 Scalar and vector fields .................................. 274
  A.2.1 Definitions .............................................. 274
  A.2.2 A base in space ....................................... 275
  A.2.3 The nabla operator .................................... 276
  A.2.4 The gradient .......................................... 276
  A.2.5 The divergence ....................................... 276
  A.2.6 The curl ............................................... 278
  A.2.7 Conservative fields .................................. 278
  A.2.8 The Laplace operator ................................ 279
A.3 Integrals .................................................... 280
  A.3.1 The curve integral .................................... 280
  A.3.2 The surface integral .................................. 281
  A.3.3 The Stokes edge integral theorem.................. 282
  A.3.4 The Gauss integral theorem ......................... 282
A.4 The continuity of matter ................................ 283

B. Function spaces ............................................. 285
  B.1 An abstract vector space ............................... 285
  B.2 Fourier function space .................................. 286
  B.3 Sturm-Liouville differential equations ................ 287
    B.3.1 The eigenvalue problem ............................ 287
    B.3.2 A self-adjoint operator ............................ 288
    B.3.3 Self-adjoint differential equations ............... 290
  B.4 Legendre polynomials ................................... 291
  B.5 Spherical harmonics .................................... 291

C. Why does FFT work? ....................................... 295
D. Helmert condensation

D.1  The interior potential of the topography . . . . . . . . . . . . 297
D.2  The exterior potential of the topography . . . . . . . . . . . . 298
D.3  The exterior potential of the condensation layer . . . . . . . 299
D.4  The total potential of Helmert condensation . . . . . . . . . . 299
   D.4.1  The gravity effect of Helmert condensation . . . . . . . 300
   D.4.2  The interior potential of Helmert condensation . . . . . 300
D.5  The dipole method . . . . . . . . . . . . . . . . . . . . . . . . . 301

Index

304
## List of Tables

2.1 Legendre polynomials. .......................................................... 39
2.2 Associated Legendre functions. ................................. 40
2.3 Coefficients of the EGM96 spherical harmonic expansion. . 52
2.4 Legendre functions of the second kind. ....................... 53

3.1 Orders of magnitude of gravity variations ......................... 59
3.2 GRS80 normal potential spherical harmonic coefficients. . 69

12.1 Altimetry satellites through the ages. ......................... 231

13.1 The various periods in the theoretical tide .................. 256
List of Figures

1.1 Gravitation is universal. .............................................. 2
1.2 A thin spherical shell consists of rings. ......................... 5
1.3 Dependence of potential and attraction on distance. ......... 7
1.4 A double mass density layer. ...................................... 16
1.5 A graphical explanation of the Gauss integral theorem. .... 18
1.6 A little rectangular box. ............................................. 19
1.7 Green’s theorem for an external point. .......................... 23
1.8 Green’s theorem for an internal point. .......................... 24
1.9 The Green theorem for a body in external space. ............. 25

2.1 Attenuation of the gravity field with height. ..................... 33
2.2 Definition of spherical co-ordinates. .............................. 35
2.3 Definition of geodetic co-ordinates. ............................... 36
2.4 Some Legendre polynomials ........................................ 39
2.5 Associated Legendre functions. ................................... 40
2.6 The algebraic signs of spherical harmonics on the Earth surface. 41
2.7 Surface spherical harmonics as maps. ............................. 42
2.8 Code for drawing a surface spherical-harmonic map. .......... 45
2.9 Monopole, dipole, and quadrupole, and their effects on the geoid. .................................................. 50

3.1 The normal gravity field of the Earth. ............................. 55
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>Gravitation and centrifugal force.</td>
<td>57</td>
</tr>
<tr>
<td>3.3</td>
<td>The curvature of level surfaces.</td>
<td>60</td>
</tr>
<tr>
<td>3.4</td>
<td>Natural co-ordinates.</td>
<td>62</td>
</tr>
<tr>
<td>3.5</td>
<td>The geometry of the meridian ellipse and various types of latitude.</td>
<td>65</td>
</tr>
<tr>
<td>3.6</td>
<td>The normal field’s potential curve over the equator.</td>
<td>68</td>
</tr>
<tr>
<td>4.1</td>
<td>Geoid undulations and deflections of the plumpline</td>
<td>74</td>
</tr>
<tr>
<td>4.2</td>
<td>A Finnish geoid model</td>
<td>75</td>
</tr>
<tr>
<td>4.3</td>
<td>Equipotential surfaces of the gravity and normal gravity fields.</td>
<td>76</td>
</tr>
<tr>
<td>4.4</td>
<td>Various reference surfaces.</td>
<td>78</td>
</tr>
<tr>
<td>4.5</td>
<td>Free-air anomalies for Southern.</td>
<td>84</td>
</tr>
<tr>
<td>5.1</td>
<td>The attraction of a Bouguer plate.</td>
<td>87</td>
</tr>
<tr>
<td>5.2</td>
<td>The Bouguer plate as an approximation to the topography.</td>
<td>88</td>
</tr>
<tr>
<td>5.3</td>
<td>The behaviour of different anomaly types in mountainous terrain.</td>
<td>89</td>
</tr>
<tr>
<td>5.4</td>
<td>Terrain corrected Bouguer anomalies for Southern Finland.</td>
<td>90</td>
</tr>
<tr>
<td>5.5</td>
<td>Calculating the classical terrain correction by the prism method</td>
<td>91</td>
</tr>
<tr>
<td>5.6</td>
<td>The steps in computing the Bouguer anomaly.</td>
<td>91</td>
</tr>
<tr>
<td>5.7</td>
<td>A special terrain shape.</td>
<td>92</td>
</tr>
<tr>
<td>5.8</td>
<td>Helmert condensation and the.</td>
<td>95</td>
</tr>
<tr>
<td>5.9</td>
<td>Friedrich Robert Helmert</td>
<td>96</td>
</tr>
<tr>
<td>5.10</td>
<td>Isostasy and the bending of plumlines towards the mountain.</td>
<td>97</td>
</tr>
<tr>
<td>5.11</td>
<td>Pratt-Hayford isostatic hypothesis.</td>
<td>97</td>
</tr>
<tr>
<td>5.12</td>
<td>Airy-Heiskanen isostatic hypothesis.</td>
<td>98</td>
</tr>
<tr>
<td>5.13</td>
<td>Quantities in isostatic compensation.</td>
<td>99</td>
</tr>
<tr>
<td>5.14</td>
<td>The current understanding of isostasy.</td>
<td>101</td>
</tr>
<tr>
<td>5.15</td>
<td>Isostatic gravity anomalies for Southern Finland.</td>
<td>103</td>
</tr>
<tr>
<td>5.16</td>
<td>Isostatic reduction as a pair of surface density layers.</td>
<td>106</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.1</td>
<td>The principle of levelling.</td>
<td>109</td>
</tr>
<tr>
<td>6.2</td>
<td>Height reference pillar in the...</td>
<td>111</td>
</tr>
<tr>
<td>6.3</td>
<td>Levelled heights and geopotential numbers.</td>
<td>112</td>
</tr>
<tr>
<td>6.4</td>
<td>Lake Päijänne: water flows “uphill”</td>
<td>113</td>
</tr>
<tr>
<td>6.5</td>
<td>Mikhail Sergeevich Molodensky, from an obscure Russian document.</td>
<td>114</td>
</tr>
<tr>
<td>6.6</td>
<td>A graphic cartoon of Molodensky’s proof.</td>
<td>116</td>
</tr>
<tr>
<td>6.7</td>
<td>Geoid, quasigeoid, terruroid and topography...</td>
<td>117</td>
</tr>
<tr>
<td>6.8</td>
<td>An optical lattice clock</td>
<td>126</td>
</tr>
<tr>
<td>7.1</td>
<td>The principle of gravimetric geoid determination.</td>
<td>130</td>
</tr>
<tr>
<td>7.2</td>
<td>Integrating the Stokes equation geometrically.</td>
<td>131</td>
</tr>
<tr>
<td>7.3</td>
<td>The Stokes kernel function $S(\psi)$</td>
<td>132</td>
</tr>
<tr>
<td>7.4</td>
<td>The geometry of the generating function.</td>
<td>133</td>
</tr>
<tr>
<td>7.5</td>
<td>The Poisson kernel for gravity anomalies...</td>
<td>138</td>
</tr>
<tr>
<td>7.6</td>
<td>Residual terrain Modelling (RTM)</td>
<td>140</td>
</tr>
<tr>
<td>7.7</td>
<td>Modified Stokes kernel functions</td>
<td>145</td>
</tr>
<tr>
<td>7.8</td>
<td>Simpson integration nodal weights in two dimensions.</td>
<td>148</td>
</tr>
<tr>
<td>8.1</td>
<td>Map projection co-ordinates $x, y$ in the local tangent plane</td>
<td>152</td>
</tr>
<tr>
<td>8.2</td>
<td>“Tapering” 25%</td>
<td>161</td>
</tr>
<tr>
<td>8.3</td>
<td>Example images for FFT transform.</td>
<td>162</td>
</tr>
<tr>
<td>8.4</td>
<td>The Finnish FIN2000 geoid</td>
<td>163</td>
</tr>
<tr>
<td>8.5</td>
<td>Calculating the terrain correction with the FFT method.</td>
<td>165</td>
</tr>
<tr>
<td>9.1</td>
<td>Definition of geocentric angular distance and azimuth.</td>
<td>174</td>
</tr>
<tr>
<td>9.2</td>
<td>Hirvonen’s covariance function in two dimensions</td>
<td>179</td>
</tr>
<tr>
<td>9.3</td>
<td>An example of least-squares collocation</td>
<td>180</td>
</tr>
<tr>
<td>9.4</td>
<td>Global covariance functions as degree variances</td>
<td>192</td>
</tr>
<tr>
<td>9.5</td>
<td>Circular geometry.</td>
<td>193</td>
</tr>
<tr>
<td>10.1</td>
<td>Jean Richer’s report</td>
<td>198</td>
</tr>
</tbody>
</table>
10.2 Autograv CG5 spring gravimeter .......................... 199
10.3 Operating principle of spring gravimeter ................. 201
10.4 The idea of astatization. .................................. 203
10.5 Operating principle of a ballistic absolute gravimeter. .... 205
10.6 FG5 absolute gravimeter .................................. 206
10.7 Principle of operation of an atomic gravimeter. .......... 208
10.8 International intercomparison of absolute gravimeters .. 209
10.9 Principle of operation of a superconducting gravimeter .. 209

11.1 The two different hypotheses on the mechanism of post-glacial land uplift. ........................................... 222
11.2 The Fennoscandian 63°N parallel gravity line. ............. 223
11.3 Connection between sea-surface topography and ocean currents. 225
11.4 Sea-surface topography map produced by GOCE. ......... 226
11.5 The sea-level equation ..................................... 228
11.6 Sea-level rise after the last ice age ......................... 230

12.1 Results from the TOPEX/Poseidon ......................... 233
12.2 Satellite altimetry as a measurement method; concepts. .... 234
12.3 A simple crossover geometry. ............................. 236
12.4 Orbital geometry of satellite altimetry. ..................... 239
12.5 Kepler’s orbital elements. ................................. 241
12.6 The mechanism of a Sun-stationary orbit .................. 243
12.7 Geometry of a “no-shadow” orbit. Season names are boreal. 243
12.8 A satellite orbit example. ................................. 244
12.9 Analyzing the altimeter pulse............................... 246
12.10 Ice volume on the Arctic Sea .............................. 247
12.11 Determining the Earth’s gravity field from ................ 249
12.12 The principle of the GRACE satellites ................... 250
12.13 Determining the Earth’s gravity field with the gravity gradiometer ................. 252
13.1 Theoretical tide. $z$ is the local zenith angle of the Moon (or Sun) \[ \text{254} \]

13.2 The main components of the theoretical tide \[ \text{257} \]

A.1 Exterior or vectorial product \[ \text{273} \]

A.2 Kepler’s second law \[ \text{274} \]

A.3 The gradient \[ \text{277} \]

A.4 The divergence \[ \text{277} \]

A.5 The curl \[ \text{279} \]

A.6 The Stokes curl theorem \[ \text{281} \]

A.7 The Gauss divergence theorem \[ \text{283} \]
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGU</td>
<td>American Geophysical Union</td>
</tr>
<tr>
<td>BGI</td>
<td>Bureau Gravimétrique International, International Gravity Bureau</td>
</tr>
<tr>
<td>CHAMP</td>
<td>Challenging Minisatellite Payload for Geophysical Research and Applications</td>
</tr>
<tr>
<td>DMA</td>
<td>Defense Mapping Agency (U.S.A.)</td>
</tr>
<tr>
<td>DTM</td>
<td>digital terrain model</td>
</tr>
<tr>
<td>EGM96</td>
<td>Earth Gravity Model 1996</td>
</tr>
<tr>
<td>EGM2008</td>
<td>Earth Gravity Model 2008</td>
</tr>
<tr>
<td>EGU</td>
<td>European Geosciences Union</td>
</tr>
<tr>
<td>ENSO</td>
<td>El Niño Southern Oscillation</td>
</tr>
<tr>
<td>ESA</td>
<td>European Space Agency</td>
</tr>
<tr>
<td>FIN2000</td>
<td>geoid model (Finland)</td>
</tr>
<tr>
<td>FIN2005N00</td>
<td>geoid model (Finland)</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transform</td>
</tr>
<tr>
<td>FGI</td>
<td>Finnish Geospatial Research Institute, formerly Finnish Geodetic Institute</td>
</tr>
<tr>
<td>GDR</td>
<td>Geophysical Data Record</td>
</tr>
<tr>
<td>GFZ</td>
<td>Geoforschungszentrum (Potsdam, Germany)</td>
</tr>
<tr>
<td>GIA</td>
<td>glacial isostatic adjustment</td>
</tr>
</tbody>
</table>
GNSS  Global Navigation Satellite Systems
GOCE  Geopotential and Steady-state Ocean Circulation Explorer
GPS   Global Positioning System
GRS80 Geodetic Reference System 1980
GRACE Gravity Recovery And Climate Experiment
HUT   Helsinki University of Technology
IAG   International Association of Geodesy
ICET  International Center for Earth Tides
ICGEM International Center for Global Earth Models
IDEMS International Digital Elevation Model Service
IGeC  International Geoid Commission (obsolete)
IGFS  International Gravity Field Service
ISG   International Service for the Geoid
IUGG  International Union of Geodesy and Geophysics
JILA  Joint Institute for Laboratory Astrophysics (Boulder CO, U.S.A.)
KKJ   National Grid Co-ordinate System (Finland)
Lego™ "Leg Godt", en. “Play Well”, Danish toy brand
LLR   Lunar laser ranging
LSC   least-squares collocation
Mf    Moon, fortnightly tide
moho  Mohorovičić discontinuity
N60   height system (Finland)
N2000 height system (Finland)
NAP   Normaal Amsterdams Peil, height system (Western Europe)
NASA  National Aeronautics and Space Administration (USA)
NAVD88 North American Vertical Datum 1988
NC normal correction
NGA National Geospatial-Intelligence Agency (U.S.A., formerly NIMA)
NIMA National Imagery and Mapping Agency (U.S.A., formerly DMA)
NKG Nordiska Kommissionen för Geodesi, Nordic Geodetic Commission
NKG2004 geoid model (Nordic area)
NKG2015 geoid model (Nordic area)
NOAA National Oceanic and Atmospheric Administration (U.S.A.)
OSU Ohio State University
OC orthometric correction
ppm parts per million
ppb parts per billion
RTM Residual Terrain Modelling
SI Système International d’Unités
SLR satellite laser ranging
$S_{sa}$ Sun, semi-annual tide
SWH Significant Wave Height
TC terrain correction
WGS84 World Geodetic System 1984
1. Fundamentals of the theory of gravitation

1.1 General

In this chapter we discuss the foundations of Newton’s theory of gravitation. Intuitively, the theory of gravitation is easiest to understand as “action at a distance”, where the force between two masses is proportional to the masses themselves, and inversely proportional to the square of the distance between them. This is the form of Newton’s Law of Gravitation familiar to all.

There exists an alternative but equivalent presentation, field theory, which describes gravitation as a phenomenon propagating through space, a field. The propagation is described in the field equations. The field approach isn’t quite as intuitive, but is a powerful theoretical tool.

In this chapter we acquaint ourselves with the central concept of field theory, the gravitational potential. We also get to know the potential fields of the theoretically interesting single and double mass density layers. Of the practical and theoretical applications of these may be mentioned the Bouguer plate and so-called Helmert condensation. In the following we will discuss their properties in detail. Mass density layers are also used in deriving the theorems of Green. We will learn about important integral theorems like the theorems of Gauss and Green, with the aid of which we may infer the whole potential field in space from field values given only on a certain surface. Other similar examples are the Chasles theorem, the Stokes theorem and the solution to Dirichlet’s problem.

In chapter 2 we apply these fundamentals of potential theory to derive a spectral representation of the Earth’s gravitational field, a so-called spherical harmonic expansion.
Here in the beginning we derive a large set of mathematical equations, such as integral equations. This is an unfortunately necessary preliminary work. These equations, however, are no end in themselves and they are not worth committing to memory. Try rather to understand their logic, and how historically these various results have been arrived at. Try as well to acquire an intuition, a fingertip feeling, about the nature of this theory.

### 1.2 Gravitation between two masses

We start the investigation of the Earth’s gravity field suitably with Isaac Newton’s general law of gravitation:

$$F = G \frac{m_1 m_2}{\ell^2}.$$  \hspace{1cm} (1.1)

$F$ is the attraction between bodies 1 and 2; $m_1$ and $m_2$ are the masses of the bodies and $\ell$ is the distance between them. We assume the masses to be

Sir Isaac Newton PRS (1642 – 1727) was an English universal genius who derived the mathematical underpinning of astronomy, and much of geophysics, in his main work “Philosophiae Naturalis Principia Mathematica” (“Mathematical Foundations of Physics”).
1.2. Gravitation between two masses

points. The constant $G$ has the value

$$G = 6.6726 \cdot 10^{-11} \text{m}^3\text{kg}^{-1}\text{s}^{-2}.$$  

The value of $G$ was determined for the first time by Henry Cavendish\(^2\) using a sensitive torsion balance (Cavendish, 1798).

Let us call the small body, the test mass, e.g., a satellite, $m$, and the large mass, the planet or the Sun, $M$. Then, $m_1 = M$ may be called the attracting mass, and $m_2 = m$ the attracted mass, and we obtain

$$F = G \frac{mM}{\ell^2}.$$

According to Newton’s law of motion

$$F = ma,$$

where $a$ is the acceleration of body $m$. From this follows

$$a = G \frac{M}{\ell^2}.$$

From this equation, the quantity $m = m_2$ has vanished. This is the famous observation by Galileo, that all bodies fall equally fast\(^3\), irrespective of their mass. This is also known as Einstein’s principle of equivalence.

Both the force $F$ and the acceleration $a$ have the same direction as the line connecting the bodies. For this reason one often writes equation (1.1) as a vector equation, which is more expressive:

$$\mathbf{a} = -GM \frac{\mathbf{r} - \mathbf{R}}{\ell^3},$$  \hspace{1cm} (1.2)

where the three-dimensional vectors of place of both the attracting and attracted masses are defined as follows in rectangular co-ordinates:\(^4\):

$$\mathbf{r} = xi + yj + zk,$$

$$\mathbf{R} = Xi + Yj + Zk,$$

---

\(^2\)Henry Cavendish FRS (1731–1810) was a British chemist and physicist from a wealthy nobility background.

\(^3\)At least in vacuum. The Apollo astronauts showed impressively, how on the Moon a feather and a hammer fall equally fast! [https://www.youtube.com/watch?feature=player_embedded&v=KDp1tiUsZw8\#].

\(^4\)As vector notation, one may use either $\vec{v}$ (an arrow above) or $\mathbf{v}$ (bold). Here we use the bold notation, except for vectors designated by Greek letters, which cannot be bolded.
where the triad of unit vectors \( \{i, j, k\} \) is an orthonormal base in Euclidean space, \( \mathbb{R}^3 \) and
\[
\ell = \|r - R\| = \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2}
\]
is the distance between the masses computed by the Pythagoras theorem.

Note that the vector equation (1.2) contains a minus sign! This only tells us, that the direction of the force is opposite to that of the vector \( r - R \). This vector is the location of the attracted mass \( m \) reckoned from the location of the attracting mass \( M \). In other words, this tells us that we are dealing with an attraction and not a repulsion.

### 1.3 The potential of a point mass

The gravitational field is a special field: if it is stationary, i.e., not time-dependent, it is conservative. This means that a body moving inside the field along a closed path will, at the end of the journey, not have lost or gained energy. Because of this, one may attach to each point in the field a “label” onto which is written the amount of energy gained or lost by a unit or test mass, when travelling from an agreed-upon starting point to the point under discussion. The value written on the “label” is called the potential. (Note, that the choice of starting point is arbitrary! We will return to this still.)

The potential function defined in this way for a pointlike body \( M \) is:
\[
V = \frac{G M}{\ell} = \frac{GM}{\ell},
\]
where \( \ell \) is again, like above, the length of the vector \( r - R \), \( \ell = \|r - R\| \).

The constant \( GM \) has in case of the Earth (according to the GRS80 system, conventionally) the value:
\[
GM_{\oplus} = 3,986005 \cdot 10^{14} \text{ m}^3/\text{s}^2.
\]

The currently best available physical value again is:
\[
GM_{\oplus} = 3,98600440 \cdot 10^{14} \text{ m}^3/\text{s}^2.
\]
1.4 Potential of a spherical shell

We may write, based on equation (1.4), the potential of an extended body into the following form:

$$V = G \int \frac{dm}{\ell}. \quad (1.5)$$

This is an integral over mass elements $dm$, where every mass element $dm$ is located at place $R$. The potential $V$ is evaluated at location $r$, and the distance $\ell = ||r - R||$.

We now derive the formula for the potential of a thin spherical shell, see figure 1.2, where we have placed the centre of the sphere at the origin $O$.

Because the circumference of a narrow ring, width $b \cdot d\theta$, is $2\pi b \sin \theta$, its surface area is

$$(2\pi b \sin \theta) (b \cdot d\theta).$$

Let the thickness of the shell be $p$ (small) and its density $\rho$. We obtain for the total mass of the ring:

$$2\pi pp b^2 \sin \theta d\theta.$$

Because every point of the ring is at the same distance $\ell$ from point $P$, we may write for the potential at point $P$:

$$V_P = \frac{2\pi G pp b^2 \sin \theta d\theta}{\ell}.$$
With the cosine rule:
\[ \ell^2 = r^2 + b^2 - 2rb \cos \theta \quad (1.6) \]
we obtain, using equation (1.5), for the potential of the whole shell:
\[ V_P = 2\pi G \rho pb^2 \int \frac{\sin \theta d\theta}{\sqrt{r^2 + b^2 - 2rb \cos \theta}}. \]
In order to evaluate this integral, we must replace the integration variable \( \theta \) by \( \ell \). Differentiating equation (1.6) yields
\[ \ell d\ell = br \sin \theta d\theta, \]
and remembering that \( \ell = \sqrt{r^2 + b^2 - 2rb \cos \theta} \) we obtain:
\[ V_P = 2\pi G \rho pb^2 \int_{\ell_1}^{\ell_2} \frac{d\ell}{br}. \]
In the case that point \( P \) is outside the shell, the integration bounds of \( \ell \) are \( \ell_1 = r - b \) and \( \ell_2 = r + b \), and we obtain for the potential of point \( P \)
\[ V_P = 2\pi G \rho pb^2 \left[ \frac{\ell}{br} \right]_{\ell=r-b}^{\ell=r+b} = \frac{4\pi G \rho pb^2}{r}. \]
Because the mass of the whole shell is \( M_b = 4\pi b^2 \rho p \), it follows that the potential of the shell is the same as that of an equal sized mass in its centre \( O \):
\[ V_P = \frac{GM_b}{r}, \]
where \( r \) is now the distance of computation point \( P \) from the centre of the sphere \( O \). We see that this is the same as equation (1.4).

In the same way, the attraction (acceleration) caused by the spherical shell is
\[ a_P = (\nabla V)_P = -4\pi G \rho pb^2 \frac{r_P - r_O}{r^3} = -GM_b \frac{r_P - r_O}{r^3}, \]
again identical the acceleration caused by an equal sized point mass located in point \( O \), see equation (1.2).

In the case that point \( P \) is inside the shell, \( \ell_1 = b - r \) and \( \ell_2 = b + r \) and the above integral changes to the following:
\[ V_P = 2\pi G \rho pb^2 \left[ \frac{\ell}{br} \right]_{\ell=b-r}^{\ell=b+r} = 4\pi G \rho pb. \]
As we see, this is a constant and not dependent upon the location of point \( P \). Therefore \( \nabla V_P = 0 \) and the attraction, being the gradient of the potential, vanishes.
1.5 Computing the attraction from the potential

The end result is, that the attraction of a spherical shell is, outside the shell,

\[ a = \|a\| = \frac{GM}{r^2}, \]

where \( M \) is the total mass of the shell and \( r = \|r_p - r_0\| \) the distance of the observation point from the shell’s centre; and 0 inside the shell.

In figure 1.3 we have drawn the curves of potential and attraction (i.e., acceleration, attractive force per unit of mass). If a body consists of many concentric spherical shells (like rather precisely the Earth and many celestial bodies), then inside the body only those layers of mass that are internal to the observation point, cause attraction, and this attraction is the same as it would be if all the mass of these layers was concentrated in the centre of the body. This case, where the distribution of mass density inside a body only depends on the distance from its centre, is called an isotropic density distribution.

1.5 Computing the attraction from the potential

As we argued earlier, the potential is a so-called path integral. Conversely we can compute, from the potential, the components of the gravitational acceleration vector by differentiating with respect to place, i.e., by taking the gradient:

\[ \mathbf{a} = \nabla V = \nabla \mathbf{V} = \hat{i} \frac{\partial V}{\partial x} + \hat{j} \frac{\partial V}{\partial y} + \hat{k} \frac{\partial V}{\partial z}. \]

(1.7)
Chapter 1. Fundamentals of the theory of gravitation

Here, the symbol $\nabla$ (nabla), is a frequently used, so-called differential operator

$$\nabla = i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z}.$$ 

Here, $\{i, j, k\}$ is again a base of mutually orthogonal unit vectors in Euclidean space $\mathbb{R}^3$.

Let us try this differentiation in the case of the potential field of the point mass $M$. Substitute the above equations for $V$ (1.4) and $\ell$ (1.3)$^5$:

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial \ell} \frac{\partial \ell}{\partial x} = GM \cdot \frac{x - X}{\ell^2} = -GM \frac{x - X}{\ell^3}.$$ 

Similarly we compute the $y$ and $z$ components:

$$\frac{\partial V}{\partial y} = -GM \frac{y - Y}{\ell^3}; \quad \frac{\partial V}{\partial z} = -GM \frac{z - Z}{\ell^3}.$$ 

These are the components of gravitational acceleration when the source of the field is one point mass $M$. So, in this concrete case the vector equation given above applies:

$$a = \nabla V = \nabla V.$$ 

Remark: in physical geodesy – unlike in, e.g., physics – the potential is reckoned always positive if the attracting mass $M$ is positive (as it is known to always be). However, the potential energy of body $m$ inside the field of mass $M$ is negative! More precisely, the potential energy of body $m$ is:

$$E_{\text{pot}} = -Vm.$$ 

In practice one calls the vector of gravitational acceleration the “gravitational vector”.

1.6 Potential of a solid body

In the following we study a solid body, the mass of which is distributed throughout space and thus not concentrated in a single point. The Earth

$^5$From the equation

$$\ell = \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2} = \left[(x - X)^2 + (y - Y)^2 + (z - Z)^2\right]^{\frac{1}{2}}$$

it follows with the chain rule, that

$$\frac{\partial \ell}{\partial x} = \frac{\partial}{\partial \left[(x - X)^2 + (y - Y)^2 + (z - Z)^2\right]} \cdot \frac{\partial (x - X)^2}{\partial x} =$$

$$= \frac{1}{2} \left[(x - X)^2 + (y - Y)^2 + (z - Z)^2\right]^{-\frac{1}{2}} \cdot 2(x - X) = \frac{x - X}{\ell}.$$
serves as an example of this, as its mass distribution in space may be described by a density function $\rho$:

$$
\rho (x, y, z) = \frac{dm}{dV (x, y, z)},
$$

where $dm$ is a mass element and $dV$ is the corresponding element of spatial volume. The dimension of $\rho$ is density, its unit in the SI system, kg/m³.

Because the gravitational acceleration (1.7) is a linear expression in the potential $V$, and both force and acceleration vectors may be summed linearly, it follows that also the total potential of the body can be obtained by summing together the potentials of all its parts. E.g., the potential of a set of $n$ mass points is

$$
V = G \sum_{i=1}^{n} \frac{m_i}{\ell_i},
$$

from which we obtain the gravitational acceleration simply using the gradient theorem (1.7).

The potential of a solid body is obtained similarly by replacing the sum by an integral, in the following way. (Note that unfortunately almost the same symbols $V$ and $V$ are used here for the potential and for volume, respectively):

$$
V = G \iiint_{\text{body}} \frac{dm}{\ell} = G \iiint_{\text{body}} \frac{\rho}{\ell} dV.
$$

(1.8)

The symbol $\rho$ inside the integral designates the mass density at the location of $dm$; $\ell = \|r - R\| = \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2}$ is the distance between point of measurement and attracting mass element. More clearly:

$$
V (x, y, z) = G \iiint_{\text{body}} \frac{\rho (X, Y, Z)}{\sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2}} dXdYdZ.
$$

As we showed already above for mass points, also the first derivative with respect to place or gradient of the geopotential $V$ of a solid body,

$$
\overrightarrow{\text{grad}}V = \nabla V = a,
$$

(1.9)

is the acceleration vector caused by the attraction of the body. This applies generally.

### 1.6.1 Behaviour at infinity

If a body is of finite extent (i.e., it lies completely within a sphere of size $\epsilon$ around the origin) and also its density is bounded everywhere, it follows that

$$
\|r\| \to \infty \Rightarrow V (r) \to 0,
$$
because, according to the triangle inequality,
\[ \ell = \| \mathbf{r} - \mathbf{R} \| \geq \| \mathbf{r} \| - \| \mathbf{R} \| \geq \| \mathbf{r} \| - \epsilon \]
and thus
\[ \frac{1}{\ell} \to 0 \quad \text{when} \quad \| \mathbf{r} \| \to \infty. \]

For the acceleration of gravitation the same applies for all three components, and thus also for the length of this vectorial quantity:
\[ \| \mathbf{r} \| \to \infty \implies \| \nabla V \| \to 0. \]

This result can still be sharpened: if \( \| \mathbf{r} \| \to \infty \), then, again by the triangle inequality,
\[ \ell = \| \mathbf{r} - \mathbf{R} \| \leq \| \mathbf{r} \| + \| \mathbf{R} \| \leq \| \mathbf{r} \| + \epsilon, \]
and thus
\[ \frac{1}{\| \mathbf{r} \| + \epsilon} \leq \frac{1}{\ell} \leq \frac{1}{\| \mathbf{r} \| - \epsilon} \implies \frac{1}{\| \mathbf{r} \| + \epsilon} \leq \frac{1}{\ell} \leq \frac{1}{\| \mathbf{r} \| - \epsilon / \| \mathbf{r} \|}. \]

It is seen that (again with the notation \( r = \| \mathbf{r} \| )):
\[ r \to \infty \implies \frac{1}{\ell} \to \frac{1}{r}. \]

When we substitute this into the above integral (1.8), it follows that for large distances \( r \to \infty \):
\[ V = G \iiint_{\text{body}} \frac{\rho r}{\ell} dV \approx \frac{G}{r} \iiint_{\text{body}} \rho dV = \frac{GM}{r}, \]
where \( M \), the integral of density over the volume of the body, is precisely its total mass. From this we see, that at great distance, the field of a finite sized body \( M \) is nearly identical with the field generated by a point mass the total mass of which is equal to the total mass of the body. This important observation was already made by Newton. As a result of this phenomenon we can treat, in the celestial mechanics of the Solar system, the Sun and planets\(^6\) as mass points, although we know that they are not.

1.7 Example: The potential of a line of mass

The potential of a (unit-mass) vertical line of mass is
\[ V(x, y, z) = \int_0^H \frac{1}{\sqrt{(X - x)^2 + (Y - y)^2 + (Z - z)^2}} dZ, \quad (1.10) \]

\(^6\)The only important exception is formed by the forces between a planet and its moons, both due to the flattening of the planet and due to tidal effects.
where $(X, Y)$ is the location of the mass line, $(x, y, z)$ is the location of the point in which the potential is being evaluated, and the mass line extends from sea level $Z = 0$ to height $Z = H$.

Firstly we write
\[
\Delta x = X - x, \quad \Delta y = Y - y, \quad \Delta z = Z - z,
\]
and the potential becomes
\[
V(\Delta x, \Delta y, \Delta z) = \int_{-z}^{H-z} \frac{1}{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}} d\Delta z.
\]
The indefinite integral is
\[
\ln \left( \Delta z + \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \right)
\]
and substituting the integration bounds yields
\[
V = \ln \frac{H - z + \sqrt{\Delta x^2 + \Delta y^2 + (H - z)^2}}{-z + \sqrt{\Delta x^2 + \Delta y^2 + z^2}}.
\]
Now we can expand this into a Taylor series in $H$ around the point $H = 0$:

the first derivative of equation (1.10) is
\[
\frac{\partial V}{\partial H} = \frac{1}{\ell} \frac{1}{\sqrt{(X-x)^2 + (Y-y)^2 + (H-z)^2}} = \ell^{-1}
\]
where $\ell = \sqrt{(X-x)^2 + (Y-y)^2 + (H-z)^2}$; the second derivative is (chain rule)
\[
\frac{\partial^2 V}{\partial H^2} = \ell^{-1} = -\frac{1}{2} \ell^{-3} \cdot 2 (H - z) = -\frac{H - z}{\ell^3}.
\]
The third derivative, in the same way:
\[
\frac{\partial^3 V}{\partial H^3} = \frac{\partial}{\partial H} \left( -\frac{H - z}{\ell^3} \right) = \frac{3 (H - z)^2}{\ell^5} - \frac{\ell^2}{\ell^5} = \frac{3 (H - z)^2 - \ell^2}{\ell^5},
\]
and so on. The Taylor expansion is
\[
V = \underbrace{\int_{0}^{H=0} \frac{1}{\ell} dZ}_{0} + \frac{1}{\ell_0} H + \frac{1}{2} \frac{z}{\ell_0^2} H^2 + \frac{1}{6} \frac{3z^2 - \ell_0^2}{\ell_0^5} H^3 + \ldots \quad (1.11)
\]
where $\ell_0 = \sqrt{(X-x)^2 + (Y-y)^2 + z^2}$.

**Question:** how can we exploit this result for computing the potential of a complete, realistic topography?

**Answer:** in this expansion, the coefficients $1/\ell_0, 1/\ell_0^2, \ldots$ like $\ell_0$, depend only on the differences $\Delta x = X - x$ ja $\Delta y = Y - y$ between the co-ordinates of the location of the mass line $(X, Y)$ and those of the evaluation location
(x, y) – and of the height z of the evaluation point. If the topography is given in the form of a grid, we may evaluate the above expansion (1.11) for the given z value and for all possible value pairs (Δx, Δy). Then, if the grid is, e.g., N × N in size, we need only $N^2$ operations for calculating every coefficient. The brute-force evaluation of the Taylor expansion itself for the whole topography, i.e., for every point of the terrain grid and every point of the evaluation grid, requires after that $N^4$ operations, but those are much simpler: the coefficients themselves have been precomputed. And brute force isn’t even the best approach: as we shall see, the above so-called convolution can be computed much faster using the Fast Fourier transform.

We shall return to this subject more extensively in connection with terrain effect evaluation, sections 5.3 and 8.7.

1.8 Equations of Laplace and Poisson

The second derivative with respect to place of the geopotential, the first derivative with respect to place of the gravitational acceleration vector, i.e., its divergence, is also of geophysical interest. We may write:

$$\text{div} \mathbf{a} = \langle \hat{\nabla} \cdot \mathbf{a} \rangle = \langle \hat{\nabla} \cdot \left( \hat{\nabla} V \right) \rangle = \langle \hat{\nabla} \cdot \hat{\nabla} \rangle V = \Delta V = \frac{\partial^2}{\partial x^2} V + \frac{\partial^2}{\partial y^2} V + \frac{\partial^2}{\partial z^2} V,$$

(1.12)

where

$$\Delta \equiv \langle \hat{\nabla} \cdot \hat{\nabla} \rangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

is a well known symbol called the Laplace operator.

In equation (1.4) for the potential of a mass point we may show, by performing all partial derivations (1.12), that

$$\Delta V = 0,$$

(1.13)

the well known Laplace equation. This equation applies outside a point mass, and more generally everywhere in empty space: all masses can in the limit

7Pierre Simon, marquis de Laplace (1749–1827) was a French universal genius who contributed to mathematics and natural sciences. He is one of the 72 French scientists, engineers and mathematicians whose names were inscribed on the Eiffel Tower, https://en.wikipedia.org/wiki/List_of_the_72_names_on_the_Eiffel_Tower.
be considered to consist of point-like mass elements. Or, in equation (1.8) we may directly differentiate inside the triple integral sign, exploiting the circumstance that it is allowed to interchange integration and partial differentiation, if both are defined.

A potential field for which the Laplace equation (1.13) applies, is called a harmonic field.

In the case where the mass density doesn’t vanish everywhere, we have a different equation, with \( \rho \) the mass density:

\[
\Delta V = -4\pi G \rho.
\]  

(1.14)

This equation is called the Poisson\(^8\) equation.

The pair of equations

\[
\begin{align*}
\overrightarrow{\text{grad}} V &= a \\
\text{div} a &= -4\pi G \rho
\end{align*}
\]

is known as the field equations of the gravitational field. They play the same role as Maxwell’s\(^9\) field equations in electromagnetism. Unlike Maxwell’s equations however, in the above there is no time co-ordinate. Because of this, it is not possible to derive a formula for the propagation in space of gravitational waves, like the one for electromagnetic waves in Maxwell theory.

We know today that these “Newton field equations” are only approximate, and that a precise theory is Einstein’s general theory of relativity. Nevertheless in physical geodesy Newton’s theory is generally precise enough and we shall use it exclusively.

### 1.9 Gauge invariance

An important property of the potential is, that, if we add a constant \( C \) to it, nothing related to gravitation that can actually be measured, changes. This is called gauge invariance. Gravitation itself is obtained by differentiation,

\(^8\)Siméon Denis Poisson (1781–1840) was a French mathematician, physicist and geodesist, one of the 72 names inscribed on the Eiffel Tower.

\(^9\)James Clerk Maxwell FRS FRSE (1831–1879) was a Scottish physicist, the discoverer of the field equations of electromagnetism. He found a wave-like solution to the equations, and, based on propagation speed, identified light as such.
an operation that eliminates the constant term. Therefore the definition of potential is in a sense ambiguous: all potential fields obtained by different choices of C are equally valid.

Observations only give us potential differences, as spirit levellers know all too well. An often chosen definition of potential is based on requiring that, if \( \|r\| \to \infty \), then also \( V \to 0 \), which makes physical sense and yields simple equations. However, in work on the Earth’s surface, a more practical alternative may be \( V = 0 \) at the mean sea surface – although also that is not free of problems.

E.g., for the mass of the Earth \( M \) a sensible potential definition is, in spherical approximation,

\[
V = \frac{GM}{r},
\]

which vanishes at infinity \( r \to \infty \), when again a practically sensible definition would be

\[
V = \frac{GM}{r} - \frac{GM}{R},
\]

where \( R = \|R\| \) is the radius of the Earth. The latter potential vanishes where \( r = R \), on the Earth’s surface. In the limit \( r \to \infty \) its value is \(-\frac{GM}{R}\), not zero.

1.10 Single mass density layer

If we apply to the surface \( S \) of a body a “coating” of mass surface density, of mass density value

\[
\kappa = \frac{dm}{dS},
\]

we obtain for the potential an integral equation similar to equation (1.8), but a surface integral:

\[
V = G \int_S \frac{dm}{\ell} = G \int_S \kappa \frac{dS}{\ell}. \quad (1.15)
\]

Here again \( \ell \) is the distance between the point of consideration or test mass \( P \) and the moving mass element in integration \( dm \) (or surface element \( dS \)). Note that the dimension of the mass surface density \( \kappa \) is \( \text{kg/m}^2 \), different from the dimension of ordinary (volume) mass density.

This case is theoretically interesting, though physically unrealistic. The function \( V \) is everywhere continuous, also at the surface \( S \); however already its
first derivatives with place are discontinuous. The discontinuity appears in
the direction perpendicular to the surface, i.e., in the normal derivative.

Let us look at a simple case where a sphere, radius $R$, has been coated with
a layer of standard surface density $\kappa$. By computing the above integral (1.15)
we may prove (in a complicated way) that the exterior potential is the same
as it would be if all of the mass of the body were concentrated in the sphere’s centre. Earlier (section 1.4) we proved that the potential interior to the sphere
is constant.

Thus, the exterior attraction ($\ell > R$) is

$$a_e (\ell) = G \frac{M}{\ell^2} = G \frac{\kappa \cdot 4\pi R^2}{\ell^2} = 4\pi G\kappa \left( \frac{R}{\ell} \right)^2.$$ 

The interior attraction ($\ell < R$) is

$$a_i (\ell) = 0.$$ 

This means that on the surface of the sphere, the attraction is discontinuous:

$$a_e (R) - a_i (R) = 4\pi G\kappa.$$

In this symmetric case we see, that

$$a = \|a\| = \frac{\partial V}{\partial n}$$

(1.16)

where the differentiation variable $n$ represents the normal direction, i.e., the
direction perpendicular to the surface $S$. If the surface $S$ is an equipotential
surface of the potential $V$, equation (1.16) applies generally; then, the attraction
vector – more precisely, the acceleration vector – is perpendicular to the
surface $S$, and its magnitude is equal to that of the normal derivative.

### 1.11 Double mass density layer

A double mass density layer may be interpreted as a dipole density layer. The
dipoles are oriented in the direction of the surface’s normal.

If the dipole consists of two “charges” $m$ and $-m$ in locations $\mathbf{r}_1$ and $\mathbf{r}_2$, in
such a way, that the vectorial separation between them is $\Delta \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, then
the dipole moment is $\mathbf{d} = m \Delta \mathbf{r}$, a vectorial quantity. See figure 1.4.

Let the dipole layer density be

$$\mu = \frac{dM}{dS}.$$
where $dM$ is a “dipole layer element”; this layer may be seen as made up of two single layers. If we have a positive layer at density $\kappa$ and a negative layer at density $-\kappa$ and the distance between them is $\delta$, we get for small values of $\delta$ an approximate correspondence:

$$\mu \approx \delta \kappa.$$ 

The combined potential of the two single mass density layers computed as explained in the previous section is

$$V = G \int_\text{surface} \kappa \left( \frac{1}{\ell_1} - \frac{1}{\ell_2} \right) dS.$$ 

Between $\ell_1$, $\ell_2$ and $\delta$ exists the following relationship (Taylor expansion of function $1/\ell$):

$$\frac{1}{\ell_1} = \frac{1}{\ell_2} + \delta \cdot \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) + \cdots,$$

where $\frac{\partial}{\partial n}$ is again the derivative of the quantity in the normal direction of the surface.

Substitution into the equation yields

$$V = G \int_\text{surface} \kappa \delta \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) dS = G \int_\text{surface} \mu \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) dS.$$

If $\delta$ is small enough (and $\kappa$ correspondingly large), this holds exactly.

One can easily show that the above potential isn’t even continuous; the discontinuity happens at the surface $S$. Let us look again, for the sake of simplicity, at a sphere, radius $R$, coated with a layer of constant mass density $\mu$. 

---

**Figure 1.4.** A double mass density layer.
The exterior potential is
\[ V_e = -G\mu \int_\text{surface} \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) dS = 0, \quad (1.17) \]
because the integral vanishes. To prove this, we need the Gauss integral theorem, on which more below.

The interior potential is
\[ V_i = -G\mu \int_\text{surface} \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) dS = -4\pi R^2 G\mu \left( \frac{1}{\ell} \right) \bigg|_{\ell=R} = -4\pi G\mu, \]
by computing the surface integral using the sphere’s centre as the evaluation point, and using the earlier established circumstance that inside a sphere covered by a single mass density layer the potential is constant.

Now in the limit \( \ell \to R \) the result is different for the exterior and interior potentials. The difference is
\[ V_e (R) - V_i (R) = 4\pi G\mu. \]

### 1.12 The Gauss integral theorem

#### 1.12.1 Presentation

The Gauss\(^{10}\) integral theorem, famous from physics, looks in its vector form like this:
\[ \iiint_V \text{div} \mathbf{a} \, d\mathbf{V} = \iint_{\partial V} (\mathbf{a} \cdot \mathbf{n}) \, dS \quad (1.18) \]
where \( \mathbf{n} \) is the exterior normal of surface \( S \), now as a vector: the length of the vector is assumed \( \|\mathbf{n}\| = 1 \). \( \partial V \) is the surface of body \( V \).

This theorem applies to all differentiable vector fields \( \mathbf{a} \) and all “well behaved” bodies \( V \) on whose surface \( S \), everywhere a normal direction \( \mathbf{n} \) exists.

In other words, this is not a special property of the gravitational acceleration vector, though it applies also to this vector field.

#### 1.12.2 Intuitive description

Let us remark that
\[ \text{div} \mathbf{a} = \Delta V = -4\pi G\rho \]

\(^{10}\)Johann Carl Friedrich Gauss (1777–1855) was a German mathematician and universal genius. “Princeps mathematicorum”.
Chapter 1. Fundamentals of the theory of gravitation

Figure 1.5. A graphical explanation of the Gauss integral theorem.

is a source function. It describes the amount, in the part of space inside surface $S$, of positive and negative “sources” and “sinks” of the gravitational field. The situation is fully analogous with the flow pattern of a liquid: positive charges correspond to points where liquid is added to the flow, negative charges correspond to “sinks” through which liquid disappears. The vector $a$ is in this metaphor the velocity of flow; in the absence of “sources” and “sinks” it satisfies the condition $\text{div} a = 0$, which describes the conservation and incompressibility of matter.

On the other hand, the function

$$\langle a \cdot n \rangle = \frac{\partial V}{\partial n}$$

is often called the flux; in other words, how much field stuff “flows out” – just like a liquid flow – from the space inside the surface $S$ to the outside through $S$.

The Gauss integral equation states the two amounts to be equal: it is in a way a book-keeping statement demanding that everything which is produced inside a surface – $\text{div} a$ – has also to come out through the surface – $\langle a \cdot n \rangle$.

In figure 1.5 it is graphically explained, that the sum of “sources” over the inner space of the body, i.e., $\sum (+ + + \ldots)$, has to be the same as the sum of “flux” $\sum (↑↑↑ \ldots)$ over the whole boundary surface delimiting this inner space.
1.12. The Gauss integral theorem

![Figure 1.6. A little rectangular box.](image)

1.12.3 The potential version of the Gauss theorem

Let us write the Gauss equation a bit differently, using the potential instead of the gravitational vector:

\[
\iiint_V \nabla V \, dV = \iint_{\partial V} \frac{\partial V}{\partial n} \, dS,
\]

where we have done the above substitutions. We see also here the popular notation \( \partial V \) for the surface of the body. The presentational forms (1.19) and (1.18) are connected by the formulas (1.12) and (1.9), between \( V \) and \( a \).

1.12.4 Example 1: a little box

Let us look at a little rectangular box with sides \( \Delta x, \Delta y, \Delta z \); so little, that the field \( a(x, y, z) \) is inside it an almost linear function of place. Let us write \( a \) as the gradient of potential \( V \):

\[
a = \nabla V = \frac{\partial V}{\partial x} \mathbf{i} + \frac{\partial V}{\partial y} \mathbf{j} + \frac{\partial V}{\partial z} \mathbf{k} = a_x \mathbf{i} + a_y \mathbf{j} + a_z \mathbf{k}
\]

where

\[
a_x = \frac{\partial V}{\partial x}, \quad a_y = \frac{\partial V}{\partial y}, \quad a_z = \frac{\partial V}{\partial z}.
\]

Now the volume integral

\[
\iiint_V \text{div} \, a \, dV \approx \left( \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \right) \Delta x \Delta y \Delta z
\]

while the surface integral

\[
\int_{\partial V} (a \cdot n) \, dS \approx (a_x^+ - a_x^-) \Delta y \Delta z + (a_y^+ - a_y^-) \Delta x \Delta z + (a_z^+ - a_z^-) \Delta x \Delta y.
\]
Here $a^+_x$ is the value of component $a_x$ on the one surface in the $x$ direction and $a^-_x$ its value on the other surface, etc. E.g., $a^+_z$ is the value of $a_z$ on the box’s upper and $a^-_z$ on its lower surface. A box has of course six faces, in each of three co-ordinate directions both “up- and downstream”.

Then

$$a^+_x - a^-_x \approx \frac{\partial a_x}{\partial x} \Delta x,$$

$$a^+_y - a^-_y \approx \frac{\partial a_y}{\partial y} \Delta y,$$

$$a^+_z - a^-_z \approx \frac{\partial a_z}{\partial z} \Delta z,$$

and by substitution we see that

$$\int_{\partial V} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS \approx \frac{\partial a_x}{\partial x} \Delta x \cdot \Delta y \Delta z + \frac{\partial a_y}{\partial y} \Delta y \cdot \Delta x \Delta z + \frac{\partial a_z}{\partial z} \Delta z \cdot \Delta x \Delta y =$$

$$= \left( \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \right) \Delta x \Delta y \Delta z,$$

the same equation as (1.20). So, in this simple case the Gauss equation applies.

Obviously the equation works also, if we build out of these “Lego™ bricks” a larger body, because the faces of the bricks touching each other are oppositely oriented and cancel from the surface integral of the whole body. It is a bit harder to prove that the equation also applies to bodies having inclined surfaces.

1.12.5 **Example 2: The Poisson equation for a sphere**

According to the Poisson equation (1.14) we have

$$\Delta V = -4\pi G \rho.$$  

Assume a sphere, radius $R$, within which the mass density $\rho$ is constant. The volume integral throughout the sphere gives

$$\iiint_V \Delta V dV = -4\pi G \rho \iiint_V dV = -4\pi G \rho V = -4\pi GM, \quad (1.21)$$

where $M = \rho V$ is the total mass of the sphere.

On the surface of the sphere, the normal derivative is

$$\frac{\partial V}{\partial n} = \frac{\partial}{\partial r} \left. \frac{GM}{r} \right|_{r=R} = -\frac{GM}{R^2},$$

a constant, and its integral over the surface of the sphere is

$$\int_{\partial V} \frac{\partial V}{\partial n} \, dS = -\frac{GM}{R^2} \cdot S = -\frac{GM}{R^2} \cdot 4\pi R^2 = -4\pi GM. \quad (1.22)$$

The results (1.22) and (1.21) are identical, as the Gauss theorem (1.19) requires.
1.12.6 Example 3: a point mass in a eight-unit cube

Let us assume that we have a point mass of size \( GM \) in the centre of a cube bounded by the co-ordinate planes \( x, y, z \in \{-1, 1\} \). In that case the volume integral is

\[
\iiint_{V} \Delta V d\mathbf{V} = -4\pi GM \iiint_{V} \delta (\mathbf{r}) d\mathbf{V} = -4\pi GM,
\]

where \( \delta (\mathbf{r}) \) is the Dirac delta function in space, having an infinite spike at the origin, being zero elsewhere, and producing a value of 1 upon volume integration. The surface integral is six times that over the top face:

\[
-GM \int_{-1}^{+1} \int_{-1}^{+1} \frac{1}{(x^2 + y^2 + 1)^{3/2}} dx \, dy.
\]

Integrating with respect to \( x \) (expression in square brackets) yields

\[
\int_{-1}^{+1} \frac{1}{(x^2 + y^2 + 1)^{3/2}} dx = \frac{2}{(y^2 + 1) \sqrt{y^2 + 2}}.
\]

Integrating this with respect to \( y \) yields

\[
\int_{-1}^{+1} \frac{2}{(y^2 + 1) \sqrt{y^2 + 2}} dy = 2 \arctan \frac{y}{\sqrt{y^2 + 2}} \bigg|_{-1}^{+1} = 4 \arctan \frac{1}{\sqrt{3}} = 4 \cdot \frac{\pi}{6} = \frac{2}{3} \pi.
\]

Adding the six faces together yields

\[
-GM \int_{-1}^{+1} \left[ \int_{-1}^{+1} \frac{1}{(x^2 + y^2 + 1)^{3/2}} dx \right] dy = -4\pi GM,
\]

agreeing with the volume result above.

1.13 Green’s theorems

Apply the Gauss integral theorem to the vector field

\[
\mathbf{F} = \mathbf{U} \cdot \nabla \mathbf{V}.
\]

Here \( \mathbf{U} \) and \( \mathbf{V} \) are two different scalar fields. We obtain:

\[
\iiint_{V} \text{div} \mathbf{F} d\mathbf{V} =
\]

\[
= \iiint_{V} \left( \nabla \cdot \left( \mathbf{U} \cdot \nabla \mathbf{V} \right) \right) d\mathbf{V} =
\]

\[
= \iiint_{V} \mathbf{U} \Delta \mathbf{V} d\mathbf{V} + \iiint_{V} \left( \nabla \mathbf{U} \cdot \nabla \mathbf{V} \right) d\mathbf{V} =
\]

\[
= \iiint_{V} \mathbf{U} \Delta \mathbf{V} d\mathbf{V} + \iiint_{V} \left( \frac{\partial \mathbf{U}}{\partial x} \frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial y} \frac{\partial \mathbf{V}}{\partial y} + \frac{\partial \mathbf{U}}{\partial z} \frac{\partial \mathbf{V}}{\partial z} \right) d\mathbf{V}
\]
and
\[ \iiint_{\mathcal{V}} F \cdot n \, dS = \iiint_{\mathcal{V}} (U \nabla V \cdot n) \, dS = \int_{\partial \mathcal{V}} U \left( \nabla V \cdot n \right) \, dS = \int_{\partial \mathcal{V}} U \frac{\partial V}{\partial n} \, dS. \]

The end result is Green's 11 first theorem:
\[ \iiint_{\mathcal{V}} U \Delta V \, dV + \iiint_{\mathcal{V}} \left( \frac{\partial U}{\partial x} \frac{\partial V}{\partial x} + \frac{\partial U}{\partial y} \frac{\partial V}{\partial y} + \frac{\partial U}{\partial z} \frac{\partial V}{\partial z} \right) \, dV = \int_{\partial \mathcal{V}} U \frac{\partial V}{\partial n} \, dS. \]

This may be cleaned up, because the second term on the left is symmetric for the interchange of $U$ and $V$.

Let us therefore interchange $U$ and $V$, and subtract the equations obtained from each other. The result is Green’s second theorem:
\[ \iiint_{\mathcal{V}} (U \Delta V - V \Delta U) \, dV = \int_{\partial \mathcal{V}} \left( U \frac{\partial V}{\partial n} - V \frac{\partial U}{\partial n} \right) \, dS. \]

We assume in all operations, that the functions $U$ and $V$ are “well behaved”, i.e., all necessary derivatives etc. exist everywhere in body $\mathcal{V}$.

A useful special case arises by choosing for the function $U$:
\[ U = \frac{1}{\ell}, \]
where $\ell$ is the distance from the given point of evaluation $P$. This function $U$ is well behaved everywhere except precisely in point $P$, where it is not defined.

In the case where point $P$ is outside the surface $\partial \mathcal{V}$, the result, Greens third theorem, is obtained by simple substitution:
\[ \iiint_{\mathcal{V}} \frac{1}{\ell} \Delta V \, dV = \int_{\partial \mathcal{V}} \left( \frac{1}{\ell} \frac{\partial V}{\partial n} - V \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) \right) \, dS. \]

This case is depicted in figure 1.7.

In the case that point $P$ is inside surface $\partial \mathcal{V}$, the computation becomes a little more complicated. One ought to learn about the clever technique that – in this case as in others – comes to the rescue. This is why we describe it shortly.

---

11George Green (1793–1841) was a British mathematical physicist, an autodidact, working as a miller near Nottingham. He also invented the word 'potential'.
http://www-history.mcs.st-and.ac.uk/Biographies/Green.html,
http://www.greensmill.org.uk/,
1.13. Green’s theorems

We form a small sphere of radius $\epsilon$ called $V_2$ around point $P$; now we can formally define the body (containing a hole) $V = V_1 - V_2$, and also its surface $\partial V$ which consists of two parts, $\partial V = \partial V_1 - \partial V_2$.

Now we may write the volume integral into two parts:

$$
\iiint_{V} \frac{1}{\ell} \Delta V d\ell = \iiint_{V_1} \frac{1}{\ell} \Delta V d\ell - \iiint_{V_2} \frac{1}{\ell} \Delta V d\ell,
$$

where the second term can be integrated in spherical co-ordinates:

$$
\iiint_{V_2} \frac{1}{\ell} \Delta V d\ell \approx \Delta V_p \int_{0}^{\epsilon} 4\pi \ell^2 \frac{1}{\ell} d\ell = 2\pi \Delta V_p \epsilon^2,
$$

which will go to zero if we let $\epsilon \to 0$.

For the first surface integral we obtain using the Gauss integral theorem (1.19):

$$
\int_{\partial V_2} \frac{1}{\ell} \frac{\partial V}{\partial n} dS = \frac{1}{\epsilon} \int_{\partial V_2} \frac{\partial V}{\partial n} dS = \frac{1}{\epsilon} \iiint_{V_2} \Delta V d\ell \approx \frac{1}{\epsilon} \Delta V_p \cdot \frac{4}{3} \pi \epsilon^3,
$$

which also goes to zero for $\epsilon \to 0$.

The second surface integral (the normal is pointing away from $P$):

$$
-\int_{\partial V_2} V \left( \frac{1}{\ell} \right) \frac{\partial}{\partial n} dS = -\int_{\partial V_2} V \cdot \frac{1}{\ell^2} dS \approx 4\pi \epsilon^2 \cdot \epsilon^{-2} V_p.
$$

By combining all results with their correct algebraic signs we obtain – for the case where $P$ is inside surface $\partial V_1 \sim \partial V$ –:

$$
\iiint_{V} \frac{1}{\ell} \Delta V d\ell = -4\pi V_p + \int_{\partial V} \left( \frac{1}{\ell} \frac{\partial V}{\partial n} - V \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) \right) dS. \tag{1.23}
$$
After this it must be intuitively clear, and we present without any formal proof, that
\[
\begin{align*}
\int_V \frac{1}{\ell} \Delta V dV &= -2\pi \nu P + \int_{\partial V} \left( \frac{1}{\ell} \frac{\partial V}{\partial n} - \nu \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) \right) dS,
\end{align*}
\]
if point \( P \) is on the boundary surface of body \( V \), i.e., on \( \partial V \). This however presupposes that the normal derivative, and especially the normal direction, actually exists in precisely point \( P \)!

In geodesy, the typical situation is that where the body \( V \) over the volume of which one wants to evaluate the volume integral, is the whole space outside the Earth. In this case, conveniently \( \Delta V = 0 \) and the whole volume integral vanishes.

The result (1.23) may be generalized to this case, where \( V \) is the whole space outside surface \( S \). This generalization is done by now choosing as the surface \( S \) the three-part surface \( S = S_1 + S_2 + S_3 \), where \( S_3 \) is a sphere with a large radius around \( P \). Its radius is then allowed to grow in the limit to infinity, so that all integrals over both the surface \( S_3 \) and the part of space outside it vanish. Also the normal direction to the surface \( S_3 \), like that of the little ball used earlier, is inverted, i.e., aimed to the inside, towards the Earth.

The end result is:
\[
\begin{align*}
\int_V \frac{1}{\ell} \Delta V dV &= -4\pi \nu P - \int_{\partial V} \left( \frac{1}{\ell} \frac{\partial V}{\partial n} - \nu \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) \right) dS, \quad (1.24)
\end{align*}
\]
Because in this case, where \( V \) is the part of the space external to the Earth, the left-hand side integral vanishes, we may express the potential at point \( P \) suitably as a two-term surface integral over surface \( \partial V \). See below.

### 1.14 The Chasles theorem

We study the above mentioned case where the “body” is the space outside the surface \( \partial V \) (i.e., in practice: the space outside the Earth).
From the Green theorem (1.24) derived above, we may derive for a harmonic function $V$ (so, $\Delta V = 0$) in the exterior space:

$$V_P = -\frac{1}{4\pi} \int_{\partial V} \frac{1}{\ell} \frac{\partial V}{\partial n} \, dS + \frac{1}{4\pi} \int_{\partial V} V \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) \, dS. \quad (1.25)$$

**Interpretation:** The exterior, harmonic potential of an arbitrary surface can be represented as the sum of a single and a double mass density layer on that surface.

**Explanation:**

We obtain the density of a single mass layer by equation (1.15):

$$\kappa = -\frac{1}{4\pi G} \frac{\partial V}{\partial n};$$

the density of a double mass density layer is obtained by equation (1.17):

$$\mu = \frac{V}{4\pi G}.$$  

If we plug these into equation (1.25), we obtain:

$$V_P = G \int_{\partial V} \left[ \frac{\kappa}{\ell} + \mu \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) \right] \, dS.$$  

In case that the surface $\partial V$ is an equipotential surface of potential $V$, i.e., $V = V_0$, it follows that a single mass density layer suffices, because in that
The case
\[ \int_{\partial V} V \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) dS = V_0 \int_{\partial V} \frac{\partial}{\partial n} \left( \frac{1}{\ell} \right) dS = 0, \]
because the right-hand side integral vanishes based on the Gauss theorem
(the function $1/\ell$ is harmonic inside the outer space $V$). This is the Chasles theorem\(^{12}\), also called the Green equivalent layer theorem.

The theorem is used in Molodensky’s\(^{13}\) theory. Also the representation of the Earth gravity field by underground point-mass layers might be justified with this theorem.

The case where $\partial V$ is an equipotential surface is realized if the body is fluid and seeks by itself an external form equal to an equipotential surface. For planet Earth, this applies for the ocean surface. Also in electrostatic theory, for a conductor in which the electrons can move freely, the physical surface will become one equipotential surface. This is why it is often stated that the electric charges are on the surface of the conductor. This isn’t necessarily so, but from a practical viewpoint the result is the same.

Equation (1.25) simplifies in this case as follows:
\[ V_p = -\frac{1}{4\pi} \int_{\partial V} \frac{1}{\ell} \frac{\partial V}{\partial n} dS = G \int_{\partial V} \frac{\kappa}{\ell} dS. \] (1.26)

The equation tells us that we can compute the whole potential external to the Earth, if only on the surface of the Earth (the shape of which we also assume given in order to compute $1/\ell$) is given the normal gradient of the potential $V_n = \partial V/\partial n$. This gradient is precisely the gravitational acceleration, a quantity derivable from measurement. All of gravimetric geopotential determination (“geoid determination”), ever since G.G. Stokes, is based on this.

### 1.15 Boundary-value problems

The boundary-value problem is the problem of computing the potential $V$ throughout space (or throughout the body’s exterior or interior part of space) from given values relating to $V$ on the boundary surface, e.g., on the surface

---

\(^{12}\)Michel Chasles (1793–1880) was a French mathematician and geometrician, one of the 72 whose names are inscribed on the Eiffel Tower.

\(^{13}\)Mikhail Sergeevich Molodenskii (1909–1991) was an eminent Russian physical geodesist.
The simplest boundary-value problem is Dirichlet’s problem: on the boundary surface the potential \( V \) itself is given. More complicated boundary-value problems are based on linear functionals of the potential: on the boundary, some linear expression in \( V \) is given, e.g., a derivative or a linear combination of derivatives, generally

\[
L \{V\},
\]

where \( L \{\cdot\} \) is a linear operator.

**Stokes’s theorem:** If on a surface \( S \) the value of a potential function \( V(\varphi, \lambda) \) is known, there will be at most one harmonic function \( V(x, y, z) \) in the whole exterior space that satisfies this boundary condition. (*Note:* the theorem does not guarantee the existence of a harmonic \( V! \)).

**Dirichlet’s principle:** The above mentioned harmonic function exists, i.e., the Dirichlet boundary-value problem can be solved.

The Dirichlet boundary-value problem in the form popular in geodesy is: determine the potential field \( V \) if its values are given on a closed surface \( S \), and furthermore is given that \( V \) is harmonic (\( \Delta V = 0 \)) outside surface \( S \). In the vacuum of space, the potential is always harmonic, as already earlier noted: the potential of a point mass \( m_P, \frac{Gm_P}{r} \) is harmonic everywhere except at point \( P \); and an extended body consists (in the limit) of many point masses or mass elements.

In the general case this is a theoretically challenging problem; the existence and uniqueness of the solution has been proven very generally, see Heiskanen and Moritz (1967) p. 18.

### 1.16 What the boundary-value problem cannot compute

Based on the values of the potential function \( V \) on the surface \( S \) we may thus compute the function \( V(x, y, z) \) throughout space outside the surface. The boundary-value problem is a powerful general method also applied in physical geodesy. One must however note, that from potential values given on the surface it is not possible to uniquely resolve the mass distribution inside the Earth, which generates this potential.

---

\(^{14}\)Peter Gustav Lejeune Dirichlet (1805–1859) was a German mathematician also known for his contributions to number theory.
This is clear already in the simple case of a constant potential on the surface
of a sphere. If additionally is given that the mass distribution is spherically
symmetric, then nevertheless the density profile along the radius remains
indeterminate. All mass may be concentrated in the centre, or it may be as
a thin layer just under the sphere’s surface, or any alternative between these
two extremes. Without additional information – e.g., from seismic studies
or geophysical density models – we cannot resolve this issue.

Also the Chasles theorem mentioned above, equation (1.25), and its special
case, equation (1.26), are examples of this: the theorem tells how one may
describe the external potential field as generated by a mass distribution on
the surface of a body, although we know that the field has been generated by
a mass distribution extending throughout the body!

This is a fundamental limitation of all methods that try to obtain information
on the situation inside the Earth based only on gravimetric measurements on
or outside the Earth.
2. The Laplace equation and its solutions

2.1 General

An equation central to the study of the Earth’s gravitational field is the Laplace equation,

$$\Delta V = \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) V = 0.$$ 

Here, we call the symbol $\Delta$ the Laplace operator.

If we study gravitation as a field, then the Laplace equation is more natural than Newton’s formalism. Newton’s formulas are used when the mass distribution is known; it gives directly the gravitational force caused by the masses.

The Laplace equation on the other hand is a partial differential equation; its solution gives the potential of the gravitational field throughout space or a part of space. From this potential one may then calculate the effect of the field on a body moving in space at the location where the body is. This is a two-phased process. The conceptual difference is, that a certain property, a field, is attributed to empty space; we no longer talk about action at a distance directly between two bodies.

Solving the Laplace equation in the general case may be difficult. The approach generally taken is, that we choose some co-ordinate system – a rectangular system (as above), spherical co-ordinates, cylindrical co-ordinates, toroidal co-ordinates, or whatever – which fits best with the geometry of the problem at hand; then, we transform the Laplace equation to those co-ordinates; we find special solutions of a certain form; and finally we compose a general (or not-so-general) solution as a linear combination of those special solutions, i.e., a series expansion.
Fortunately the theory of linear partial differential equations is well developed; similar theoretical problems are encountered in the theory of the electromagnetic field (Maxwell theory) and quantum mechanics (Schrödinger\textsuperscript{1} equation), not to mention fluid and heat flow.

An important observation is, that the Laplace equation is linear. This means that, if given are two solutions

\[ \Delta V_1 = 0 \quad \text{ja} \quad \Delta V_2 = 0, \]

then also their linear combinations

\[ V = \alpha V_1 + \beta V_2, \quad \alpha, \beta \in \mathbb{R} \]

are good solutions, i.e., $\Delta V = 0$. This linearity property makes it possible to seek solutions as linear combinations or series expansions of basic solutions.

A peculiarity that also distinguishes the Laplace equation from Newton’s equation is, that it is a local equation. It describes the behaviour of the potential field in one point and its small neighbourhood. However, the solution is sought for a whole area. The solution technique generally used is a so-called boundary-value problem. This means that the field values (“boundary values”) have to be given only on the boundary of a certain part of space; e.g., on the Earth surface. From this, one calculates the values of the field in outer space – the behaviour of the field inside the Earth remains outside the scope of our interest. From the perspective of the exterior gravitational field the precise mass distribution inside the Earth isn’t even necessary to know – and one cannot even determine it using only exterior measurement values, i.e., values obtained on and above the Earth surface!

### 2.2 The Laplace equation in rectangular co-ordinates

It is a learning experience to write and solve the Laplace equation in rectangular co-ordinates. The case is fully analogous to that of spherical co-ordinates but the math is much simpler.

Assume that the Earth surface is the level surface for $z$ co-ordinates $z = 0$.

\textsuperscript{1}Erwin Rudolf Josef Alexander Schrödinger (1887–1961) was a German physicist and quantum theorist, the inventor of the quantum matter wave equation named after him which earned him the 1933 physics Nobel, and of the eponymous unobserved cat, which finds itself in a superposition state of being both alive and dead.
Then
\[ \Delta V = \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) V = \Delta (X(x) \cdot Y(y) \cdot Z(z)), \]
where we have “experimentally” written
\[ V(x, y, z) = X(x) \cdot Y(y) \cdot Z(z). \]

In other words, we write experimentally \( V \) as the product of three factor functions, where each factor function depends only on one co-ordinate – “separation of variables”. A realistic potential function \( V \) will of course usually not be of this form. We may however hope that we might write it as a sum, or linear combination, of terms that are of the above form, thanks to the linearity of the Laplace equation.

By taking all derivatives we obtain
\[ YZ \frac{\partial^2}{\partial x^2} X + XZ \frac{\partial^2}{\partial y^2} Y + XY \frac{\partial^2}{\partial z^2} Z = 0. \]
Dividing by the expression \( XYZ \):
\[ \frac{\partial^2 X(x)}{\partial x^2} X + \frac{\partial^2 Y(y)}{\partial y^2} Y + \frac{\partial^2 Z(z)}{\partial z^2} Z = 0. \]

Because this has to be true for all values \( x, y, z \), it follows that every term must be a constant. If we take for the first and second term \( -k_1^2 \) and \( -k_2^2 \), we get in conclusion for the third constant \( k_1^2 + k_2^2 \). Now by writing this definition and result out and moving the denominator to the other side, we obtain
\[ \frac{\partial^2}{\partial x^2} X = -k_1^2 X, \]
(why the minus sign? We shall soon see...)
\[ \frac{\partial^2}{\partial y^2} Y = -k_2^2 Y, \]
and
\[ \frac{\partial^2}{\partial z^2} Z = (k_1^2 + k_2^2) Z. \]

Now, the solution is readily found at least to the first two equations: they are harmonic oscillators, and their base solutions are
\[ X(x) = \exp(\pm ik_1 x), \]
\[ Y(y) = \exp(\pm ik_2 y). \]

\(^2\)Alternative base solutions are: \( X(x) = \sin k_1 x, X(x) = \cos k_1 x \) etc. They are equivalent with those presented because \( \exp(ik_1 x) = \cos k_1 x + i \sin k_1 x, \exp(-ik_1 x) = \cos k_1 x - i \sin k_1 x. \)
The solution of the $Z$ equation on the other hand is exponential:

$$Z (z) = \exp \left( \pm z \sqrt{k_1^2 + k_2^2} \right).$$

In principle we can now form the solution in space:

$$V_{k_1 k_2} (x, y, z) = \exp \left( \pm i \left( k_1 x + k_2 y \right) \pm z \sqrt{k_1^2 + k_2^2} \right).$$

The general solution is obtained by summing the terms $V_{k_1 k_2}$ with varying coefficients, for different values of $k_1, k_2$.

Let us assume that both in the $x$ and in the $y$ direction the size of our world is $L$ ("shoebox world"). Let us make things a little simpler by assuming that, on the borders of our shoebox world, we have the boundary conditions

$$V (0, y) = V (L, y) = V (x, 0) = V (x, L) = 0.$$

It then follows that the only pairs $(k_1, k_2)$ yielding a solution that fits the box are

$$k_1 = \frac{\pi j}{L}; \quad k_2 = \frac{\pi k}{L}$$

($j, k$ integer numbers), and the only suitable functions are sine functions. Thus we obtain as the general solution:

$$V_{jk} (x, y, z) = \sin \left( \frac{\pi j x}{L} \right) \sin \left( \frac{\pi k y}{L} \right) \exp \left( \pm \pi \sqrt{(j^2 + k^2)} \frac{z}{L} \right).$$

This particular solution may now be generalized by multiplying it with suitable coefficients, and summing it over different values $j = 0, \pm 1, \pm 2, \ldots; k = 0, \pm 1, \pm 2, \ldots$. We may however remark, that the terms for which $j = 0$ or $k = 0$ will always vanish, and the terms that contain $j = +n$ and $j = -n$, or $k = +n$ and $k = -n$, $n \in \mathbb{N}$, are (apart from their algebraic signs) identical. Therefore in practice we sum over the values $j = 1, 2, \ldots; k = 1, 2, \ldots$.

Different boundary conditions will give slightly different general solutions. Their general form is however always similar.

The zero-level $z = 0$ expansion resulting from the general solution is the familiar Fourier sine expansion:

$$V (x, y, 0) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} V_{jk} (x, y, 0) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} \sin \left( \pi \left[ \frac{j x}{L} \right] \right) \sin \left( \pi \left[ \frac{k y}{L} \right] \right),$$

Joseph Fourier (1768–1830), French mathematician, physicist – and some would say, climatologist; one of the Eiffel Tower’s 72 names.
2.3 Example: The Laplace equation in polar co-ordinates

In polar co-ordinates (two-dimensionally) the Laplace equation is

\[ \Delta V = \frac{\partial^2 V}{\partial r^2} + \frac{1}{r} \frac{\partial V}{\partial r} + \frac{1}{r^2} \frac{\partial^2 V}{\partial \alpha^2} = 0. \]

Perform on this the same kind of derivation as in 2.2, i.e., write first

\[ V (r, \alpha) = R(r) A(\alpha) \]

and then split the above equation into two equations, one for the right-hand side function \( R(r) \) and one for the function \( A(\alpha) \).

What form does the \( A(\alpha) \) function of the general solution have?

Figure 2.1. The exponential attenuation of gravitational field Fourier waviness with height. Rectangular geometry, one dimension. Long waves (small wave numbers, red) attenuate slower with height than short waves (green), i.e., the height acts as a low-pass filter.

where \( v_{jk} \) are the Fourier coefficients, when again a complete (three-dimensional) expansion is

\[
V (x, y, z) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} V_{jk} (x, y, z) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} \sin \left( \pi \left[ \frac{jx}{L} \right] \right) \sin \left( \pi \left[ \frac{ky}{L} \right] \right) \exp \left( \pm \pi \sqrt{\frac{j^2}{L^2} + \frac{k^2}{L^2}} \frac{z}{L} \right).
\]

Note, that in the \( z \) expression there may be a positive as well as a negative algebraic sign! Of course the solution with a positive sign goes to \( \rightarrow \infty \) when \( z \rightarrow \infty \), which is not physically realistic in the exterior space.
Chapter 2. The Laplace equation and its solutions

Answer:

Substitution yields

$$\frac{A \frac{\partial^2 R}{\partial r^2}}{r^2} + \frac{A \frac{\partial R}{\partial r}}{r} + \frac{R \frac{\partial^2 A}{\partial \alpha^2}}{r^2} = 0.$$  

Multiply by the expression $r^2/AR$:

$$\left( \frac{r^2}{R(r)} \frac{\partial^2 R(r)}{\partial r^2} + \frac{r}{R(r)} \frac{\partial R}{\partial r} \right) + \frac{\frac{\partial^2 A(\alpha)}{\partial \alpha^2}}{A(\alpha)} = 0.$$  

Both terms must be constant:

$$r \left( \frac{r^2}{R(r)} \frac{\partial^2 R(r)}{\partial r^2} + \frac{r}{R(r)} \frac{\partial R(r)}{\partial r} \right) - k^2 R(r) = 0,$$

$$\frac{\partial^2 A(\alpha)}{\partial \alpha^2} + k^2 A(\alpha) = 0.$$  

Here the algebraic sign of $k^2$ has been chosen so, that $A(\alpha)$ gets a periodic solution. Such a general solution would be

$$A(\alpha) = a \cos(ka) + b \sin(ka),$$

where, because the angle $\alpha$ has a period of $2\pi$, $k$ has to be integer: $k = 0, 1, 2, 3, \ldots$ Negative $k$ do not give different solutions, because $a \cos(ka) = a \cos((-k)a)$ and $b \sin(ka) = (-b) \sin((-k)a)$.

The other equation in $R(r)$ is harder to solve; the solution is found in terms of Bessel\textsuperscript{4} functions.

2.4 Spherical, geodetic, ellipsoidal co-ordinates

In physical geodesy we use side by side geometrical and physical concepts. E.g., co-ordinates of place can be given in the form $(x, y, z)$, which are in principle geometric – except for the physical assumption that the origin of the co-ordinate system is in the centre of mass of the Earth.

As the Earth is not precisely a sphere but rather an oblate ellipsoid of revolution, one cannot use geographical co-ordinates as if they were spherical co-ordinates. Because the flattening of the Earth – some $0.3\%$ – cannot be

\textsuperscript{4}Friedrich Wilhelm Bessel (1784–1846) was a German astronomer and mathematician.
ignored, this difference is significant. The connection between spherical co-ordinates \((r, \phi, \lambda)\) and rectangular ones \((X, Y, Z)\) is the following:

\[
\begin{align*}
X &= r \cos \phi \cos \lambda, \\
Y &= r \cos \phi \sin \lambda, \\
Z &= r \sin \phi.
\end{align*}
\tag{2.2}
\]

Here \(\phi\) and \(\lambda\) are geocentric latitude and (ordinary, i.e., geodetic or geographic) longitude. \(r\) is the distance from the Earth’s centre. Generally the \(x\) axis points in the direction of the Greenwich meridian. See figure 2.2.

On the Earth’s surface these spherical co-ordinates are not very useful because of the Earth’s flattening, but in space, spherical co-ordinates are much used. On the Earth surface on the other hand, usually geodetic (or geographic) co-ordinates \(\varphi, \lambda, h\) are most often used:

\[
\begin{align*}
X &= (N + h) \cos \varphi \cos \lambda, \\
Y &= (N + h) \cos \varphi \sin \lambda, \\
Z &= (N + h - e^2 N) \sin \varphi,
\end{align*}
\tag{2.3}
\]

where

\[
N(\varphi) = \frac{a}{\sqrt{1 - e^2 \sin^2 \varphi}} = \frac{a^2}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}}.
\tag{2.4}
\]

The quantity \(N\) defined in equation (2.4) is the East-West direction, or transversal, radius of curvature of the reference ellipsoid; in the equation \(a\) is
Chapter 2. The Laplace equation and its solutions

the Earth’s equatorial radius, $e^2 = \frac{a^2 - b^2}{a^2}$ is the square of the so-called first eccentricity\(^5\), and in equation (2.3) $h$ is the height of the point above the reference ellipsoid, see figure 2.3.

Transforming rectangular co-ordinates into geodetic ones is easiest to do iteratively, although the literature also offers closed formulas.

Spherical co-ordinates and geodetic/geographical co-ordinates are considerably different. In latitude, the difference is at most 11 minutes of arc or almost 20 km. This maximum is attained for latitudes $\pm 45^\circ$.

In theoretical work one also uses ellipsoidal co-ordinates $u$ and $\beta$. The co-ordinate $\beta$ is called the reduced latitude. The relationship with rectangular co-ordinates is

\[
X = \sqrt{u^2 + E^2} \cos \beta \cos \lambda, \\
Y = \sqrt{u^2 + E^2} \cos \beta \sin \lambda, \\
Z = u \sin \beta.
\]

If the semi-major axis of the Earth ellipsoid is $a$ and its semi-minor axis $b$, it follows (exercise!) that $E^2 = a^2 - b^2$.

\(^5\)The parameter is connected to the flattening $f$ through the equation $e^2 = 2f - f^2$. 

---

**Figure 2.3.** Definition of geodetic co-ordinates.
2.5 The Laplace equation in spherical co-ordinates

The Laplace equation transformed to spherical co-ordinates reads:

\[ \Delta V = \frac{\partial^2 V}{\partial r^2} + \frac{2}{r} \frac{\partial V}{\partial r} + \frac{1}{r^2} \frac{\partial^2 V}{\partial \phi^2} + \frac{1}{r^2 \sin^2 \phi} \frac{\partial^2 V}{\partial \lambda^2} = 0, \]

where \( \phi \) is the (geocentric) latitude, \( \lambda \) is the longitude, and \( r \) is the distance from the origin or centre of the Earth.

We shall here not derive the solution of this equation, as it is pretty complicated and can be found in ready form in the literature. What is significant is, that the solution looks somewhat similar to the above solution in rectangular co-ordinates. The base solutions of the Laplace equation are

\[ V_{n,1}(\phi, \lambda, r) = r^n Y_n(\phi, \lambda), \quad V_{n,2}(\phi, \lambda, r) = \frac{Y_n(\phi, \lambda)}{r^{n+1}}, \]  

(2.6)

where the first is again nonphysical in outer space, because, unlike the true geopotential, these terms grow to infinity for \( r \to \infty \).

In the above equations, the functions \( Y_n \) are so-called surface spherical harmonics, whereas the functions \( V_n \) are solid spherical harmonics. The latter are harmonic everywhere in space except at the origin (eq. (2.6), second formula) or at infinity (first, physically unrealistic formula).

The functions \( Y_n \) are (the functions \( P_{nm} \) are so-called Legendre functions, on which more later on):

\[ Y_n(\phi, \lambda) = \sum_{m=0}^{n} P_{nm}(\sin \phi) (a_{nm} \cos m\lambda + b_{nm} \sin m\lambda). \]  

(2.7)

With the help of these we obtain, by using the second, physically realistic alternative from equation (2.6), the following solution or series expansion for the potential in space \( V \):

\[ V(\phi, \lambda, r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin \phi) (a_{nm} \cos m\lambda + b_{nm} \sin m\lambda). \]  

(2.8)

The subscripts \( n \) and \( m \) are called degree and order. The coefficients \( a_{nm} \) and \( b_{nm} \) are called the coefficients of the spherical harmonic expansion, or shortly spectral coefficients. Together they describe the function \( V \), in somewhat the same way that the Fourier coefficients \( v_{jk} \) do in rectangular co-ordinates in equation (2.1).

Often we will be using a somewhat freer notation for the functions \( Y_n \). E.g., if we expand the disturbing potential \( T \) into spherical harmonics, we shall use the notation \( T_n \) for its surface harmonics; similarly \( \Delta g_n \) are the surface harmonics of the gravity anomaly \( \Delta g \) for degree \( n \), and so on.
Chapter 2. The Laplace equation and its solutions

2.6 Dependence on height

From the above equation \((2.6)\) one sees that for different values of the degree \(n\) the function \(Y\) has a different dependence on the distance \(r\) from the Earth centre, or equivalently, on the height \(H = r - R\), if by \(R\) we denote the radius of the Earth. The dependence is

\[
V_n (\phi, \lambda, r) = \frac{Y_n (\phi, \lambda)}{r^{n+1}}.
\]

On the Earth surface we have

\[
V_n (\phi, \lambda, R) = \frac{Y_n (\phi, \lambda)}{R^{n+1}}.
\]

Therefore we may write

\[
V_n (\phi, \lambda, r) = \left( \frac{R}{r} \right)^{n+1} V_n (\phi, \lambda, R) = \left( \frac{R + H}{R} \right)^{-(n+1)} V_n (\phi, \lambda, R) = \left( 1 + \frac{H}{R} \right)^{-(n+1)} V_n (\phi, \lambda, R) \approx \exp \left( -\frac{H}{R} (n + 1) \right) V_n (\phi, \lambda, R).
\]

We see that the dependence of the potential on height is again exponential, and the degree number \(n\) appears in the exponent, as did also the wave number in rectangular geometry, see equation \((2.1)\) and figure \(2.1\). The analogy works well.

2.7 Legendre’s functions

In the above equations the functions \(P\) are so-called Legendre\(^6\) functions that pop up whenever we solve a Laplace-like equation in spherical co-ordinates. There exist various effective, so-called recursive algorithms, e.g., the following (only for ordinary Legendre polynomials \(P_n = P_n^0\)):

\[
nP_n (t) = - (n - 1) P_{n-2} (t) + (2n - 1) t P_{n-1} (t).
\]

Similar equations exist also for the functions \(P_{nm}, m > 0\); There are even alternatives to choose from, though the formulas are generally complicated. One should be careful that in their computation, the factorials don’t go overboard! Already 30! (factorial of 30) is a larger number than most computers

\(^6\)Adrien-Marie Legendre (1752–1833) was a French mathematician known for his work on number theory, statistics – he invented independently from Gauss the method of least squares – and on elliptical functions. His name is inscribed on the Eiffel Tower.
2.7. Legendre’s functions

Table 2.1. Legendre polynomials.

<table>
<thead>
<tr>
<th>Function of $t$</th>
<th>Function of $\sin (n\phi)$ and $\cos (n\phi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_0 (t) = 1$</td>
<td>$P_0 (\sin \phi) = 1$</td>
</tr>
<tr>
<td>$P_1 (t) = t$</td>
<td>$P_1 (\sin \phi) = \sin \phi$</td>
</tr>
<tr>
<td>$P_2 (t) = \frac{3}{2}t^2 - \frac{1}{2}$</td>
<td>$P_2 (\sin \phi) = -\frac{3}{4} \cos 2\phi + \frac{1}{4}$</td>
</tr>
<tr>
<td>$P_3 (t) = \frac{5}{2}t^3 - \frac{3}{2}t$</td>
<td>$P_3 (\sin \phi) = -\frac{5}{8} \sin 3\phi + \frac{3}{8} \sin \phi$</td>
</tr>
<tr>
<td>$P_4 (t) = \frac{1}{8} (35t^4 - 30t^2 + 3)$</td>
<td></td>
</tr>
<tr>
<td>$P_5 (t) = \frac{1}{8} (63t^5 - 70t^3 + 15t)$</td>
<td></td>
</tr>
<tr>
<td>$P_6 (t) = \frac{1}{16} (231t^6 - 315t^4 + 105t^2 - 5)$</td>
<td></td>
</tr>
</tbody>
</table>

can handle... not to mention 360!. Heiskanen and Moritz (1967), equation 1-62, unlike is stated there, is not suitable for computer use!

The first Legendre polynomials are listed in table 2.1.

Higher polynomials than this are rarely needed in manual computation. Note that the even polynomials are mirror symmetric around the origin, $P_n (-t) = P_n (t)$, and the odd ones are antisymmetric, $P_n (-t) = -P_n (t)$.

For comparison: also the Fourier base functions (like, in a somewhat more complicated way, also sines and cosines!)

$$F_j (x) = \exp \left( 2\pi ij \frac{x}{L} \right)$$

(where $i^2 = -1$) can be computed recursively:

$$F_{j+1} (x) = F_j (x) \cdot F_1 (x).$$

Figure 2.4. A number of Legendre polynomials ($P_0 (t) \ldots P_{25} (t)$) as functions of the argument $t = \sin \phi$. 
Table 2.2. Associated Legendre functions.

<table>
<thead>
<tr>
<th>Function of $t$</th>
<th>Trigonometric function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{11} (t) = \sqrt{1-t^2}$</td>
<td>$P_{11} (\sin \phi) = \cos \phi$</td>
</tr>
<tr>
<td>$P_{21} (t) = 3t\sqrt{1-t^2}$</td>
<td>$P_{21} (\sin \phi) = 3 \sin \phi \cos \phi$</td>
</tr>
<tr>
<td>$P_{22} (t) = 3 (1 - t^2)$</td>
<td>$P_{22} (\sin \phi) = 3 \cos^2 \phi$</td>
</tr>
<tr>
<td>$P_{31} (t) = \frac{3}{2} (5t^2 - 1) \sqrt{1-t^2}$</td>
<td>$P_{31} (\sin \phi) = \frac{3}{2} (5 \sin^2 \phi - 1) \cos \phi$</td>
</tr>
<tr>
<td>$P_{32} (t) = 15t (1 - t^2)$</td>
<td>$P_{32} (\sin \phi) = 15 \sin \phi \cos^2 \phi$</td>
</tr>
<tr>
<td>$P_{33} (t) = 15 (1 - t^2)^{3/2}$</td>
<td>$P_{33} (\sin \phi) = 15 \cos^3 \phi$</td>
</tr>
</tbody>
</table>

Of the associated Legendre functions $P_{nm}$ let us mention only those in table 2.2.

One defining equation for these is:

$$P_{nm} (t) = (1-t^2)^{m/2} \frac{d^m P_n (t)}{dt^m}. \tag{2.10}$$

Starting from equation (2.7) we may write

$$Y_n (\phi, \lambda) = \sum_{m=0}^{n} (a_{nm} P_{nm} (\sin \phi) \cos m\lambda + a_{n,-m} P_{n,m} (\sin \phi) \sin m\lambda) = \sum_{m=-n}^{n} a_{nm} Y_{nm} (\phi, \lambda),$$

where now $m$ runs from $-n$ to $+n$. Here

$$Y_{nm} (\phi, \lambda) = \begin{cases} P_{nm} (\phi, \lambda) \cos m\lambda & \text{if } m \geq 0, \\ P_{n||m||} (\phi, \lambda) \sin ||m|| \lambda & \text{if } m < 0. \end{cases}$$
2.7. Legendre’s functions

Figure 2.6. The algebraic signs of spherical harmonics on the Earth surface. Grey means positive, white negative. The functions “wave” in a sine or cosine function like fashion.

These are the surface spherical harmonics of degree \( n \) and order \( m \).

Such surface spherical harmonics come in three kinds:

- **Zonal functions:** \( m = 0 \); these functions depend only on latitude.
- **Sectorial functions:** \( m = n \); the algebraic signs of these functions depend only on longitude and not on latitude (the functions themselves however depend on both latitude and longitude).
- **Tesseral functions:** \( 0 < m < n \). These functions, the algebraic sign of which changes with both latitude and longitude, form a checkerboard pattern on the surface of the sphere, if the positive values are painted white and the negative ones black (Lat. *tessera* = a tile, as used in a mosaic).

Every function will, on the interval \( \sin \phi \in [-1, +1] \), go precisely \( n - m \) times through zero. Every function is either symmetric or antisymmetric about the origin as a function of \( \phi \) or \( t = \sin \phi \).

Spherical harmonics thus represent a wave phenomenon of sorts. The are however not wave functions (sines or cosines), the connection to those is complicated at least. It is nevertheless sensible to speak of their wavelength.

In figure 2.6 is depicted how the algebraic signs of the different spherical harmonics behave on the Earth surface (and above).

When looking in equation (2.7) at the expressions \( \cos m\lambda \) and \( \sin m\lambda \), we observe that around a full circle (at the equator) \( 0 \leq \lambda < 2\pi \) they go precisely \( 2m \) times through zero. The “semi-wavelength” is thus

\[
\frac{2\pi R}{2m} = \frac{\pi R}{m},
\]
where $R$ is again the radius of the Earth.

A similar formula applies also for functions $P_{nm}(\sin \phi)$: as the function passes through zero $n - m$ times on the interval $-\frac{\pi}{2} \leq \phi < \frac{\pi}{2}$, it follows that also here the semi-wavelength is

$$\frac{\pi R}{n - m}.$$

If we plug various values for $m$ and the expression $n - m$ into this, we obtain the following table:

<table>
<thead>
<tr>
<th>$m$/$n-m$</th>
<th>Semi-wavelength (km)</th>
<th>In degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2000</td>
<td>$18^\circ$</td>
</tr>
<tr>
<td>40</td>
<td>500</td>
<td>$4.5^\circ$</td>
</tr>
<tr>
<td>180</td>
<td>111</td>
<td>$1^\circ$</td>
</tr>
<tr>
<td>360</td>
<td>55</td>
<td>$0.5^\circ = 30'$</td>
</tr>
<tr>
<td>1800</td>
<td>11</td>
<td>$0.1^\circ = 6'$</td>
</tr>
<tr>
<td>10,800</td>
<td>1.85</td>
<td>$1'$</td>
</tr>
</tbody>
</table>

This table also gives the resolution that can be achieved with a spherical harmonic expansion, i.e., in how detailed a fashion the expansion can describe the gravity field of the Earth. The expansions available today, like the model EGM2008, go to degree $n = 2159$; the “sharpness” of a geopotential image based on them is thus 9 km. Models based on satellite orbit perturbations often extend only to degree 40; in this case, only details the size of continents.
– order 500 km – are visible. On the other hand, experimental spherical harmonic expansions of the topography go even up to degree 10,800 (Balmino et al., 2012).

2.8 Symmetry properties of the spherical harmonic expansion

We recapitulate the spherical harmonic expansion given at the beginning, equation (2.8):

\[ V(\phi, \lambda, r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi) \left( a_{nm} \cos m\lambda + b_{nm} \sin m\lambda \right). \]  

(2.8)

2.8.1 Dependence on latitude \( \phi \)

It is seen that the dependence on \( \phi \) only works through the associated Legendre function \( P_{nm}(\sin \phi) \). This function can be either symmetric in \( \phi \), or antisymmetric in \( \phi \), meaning that either (symmetric) \( P_{nm}(\sin (-\phi)) \), or (antisymmetric) \( P_{nm}(\sin \phi) = -P_{nm}(\sin (-\phi)) \). Equivalently it means that, with \( t = \sin \phi \), that either (symmetric) \( P_{nm}(t) = P_{nm}(-t) \), or (antisymmetric) \( P_{nm}(t) = -P_{nm}(-t) \).

Which case applies, depends on the values of both \( n \) and \( m \). To figure it out, one can look at, e.g., equation (2.10):

\[ P_{nm}(t) = \left(1 - t^2 \right)^{n/2} \frac{d^n P_n(t)}{dt^n}. \]

We need to answer two questions:

1. For which values \( n \) is the polynomial \( P_n \) symmetric, for which is it antisymmetric in \( t \)? For this, you need to look at the equation for recursive computation of the polynomials, eq. (2.9). We already know that \( P_0(t) = 1 \) is symmetric, and that \( P_1(t) = t \) is antisymmetric. The rule for other \( n \) values follows recursively (or you could cheat by looking at table 2.1).

2. What does differentiation \( \frac{d}{dt} \) do to the symmetry or antisymmetry of the function?

3. (Multiplication by \( \sqrt{1-t^2} = \cos \phi \) changes nothing, as this factor is symmetric in \( t \) or \( \phi \).)


So, in order to make expansion (2.8) mirror symmetric between Northern and Southern hemispheres, one has to set the coefficients $a_{nm}, b_{nm}$ for which the corresponding $P_{nm}$ is antisymmetric, to zero. The coefficients remaining are those for which the corresponding $P_{nm}$ is symmetric.

In figure 2.8 we give a code fragment to plot an arbitrary surface spherical harmonic, e.g., in order to visually judge its symmetry properties. Don’t believe; test.

### 2.8.2 Dependence on longitude $\lambda$

This dependence works though the “Fourier functions” $\cos m\lambda$ and $\sin m\lambda$. The interesting property here is rotational symmetry: does the spherical harmonic expansion (2.8) change when we change $\lambda$?

We see immediately that, for $m \neq 0$, there will be dependence on $\lambda$ if any coefficient $a_{nm}, b_{nm}$ is non-zero. So all coefficients $a_{nm}, b_{nm}$ for values $m > 0$ must be suppressed: $a_{11} = b_{11} = a_{21} = b_{21} = a_{22} = b_{22} = \cdots = 0$.

Of the remaining coefficients, we can say that for $m = 0$, $\sin m\lambda = 0$ identically, so the coefficients $b_{00}, b_{10}, b_{20}, \ldots$ simply don’t matter. They may be any value, including zero. The coefficients $a_{00}, a_{10}, a_{20}$ however do matter, as for $m = 0$, $\cos m\lambda = 1$ identically. So we obtain as the rotationally symmetric expansion

$$V(\phi, \lambda, r) = V(\phi, r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} a_{n0} P_n(\sin \phi),$$

where $P_n = P_{n0}$.

### 2.9 The orthogonality of the Legendre polynomials

Legendre’s polynomials are orthogonal: the integral

$$\int_{-1}^{+1} P_n(t) P_m(t) \, dt = \begin{cases} 2 & \text{jos } n = m \\ 0 & \text{jos } n \neq m \end{cases}.$$

(2.11)

This orthogonality is just one example of a more general way to look at functions and integrals over functions. There exists a useful analogy with vector spaces: see appendix B.

Alternatively we may write, on the surface of a unit sphere $\sigma$, where we used
2.10. Spectral representations of various quantities

2.10.1 The potential

Starting from equation (2.8) we write the following spectral expansion of the geopotential in space:

\[ V(\phi, \lambda, r) = \sum_{n=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} V_n(\phi, \lambda), \quad (2.12) \]

where the spectral components \( V_n \) are the old familiar \( Y_n \) from equation (2.7), just scaled a little differently:

\[ V_n(\phi, \lambda) = \sum_{m=-n}^{n} v_{nm} Y_{nm}(\phi, \lambda), \]

\[ = \sum_{m=-n}^{n} v_{nm} Y_{nm}(\phi, \lambda), \]
where
\[
Y_{nm}(\phi, \lambda) = \begin{cases} 
  P_{nm}(\sin \phi) \cos m\lambda & \text{jos } m \geq 0 \\
  P_{n|m|}(\sin \phi) \sin |m|\lambda & \text{jos } m < 0 
\end{cases}
\] (2.13)
and similarly
\[
v_{nm} = \begin{cases} 
  a_{nm}^V & \text{jos } m \geq 0 \\
  b_{n|m|}^V & \text{jos } m < 0 
\end{cases}
\] (2.14)
This is a sometimes used, compact notation.

On the Earth surface \((r = R)\) we obtain
\[
V(\phi, \lambda, R) = \sum_{n=0}^{\infty} V_n(\phi, \lambda) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} v_{nm} Y_{nm}(\phi, \lambda).
\] (2.10.2) Gravitation

In the Neumann boundary-value problem we solve a function \(V\) of which the normal derivative, \(\partial V/\partial n\), is given on the surface of a body. By differentiating equation (2.12) we obtain
\[
\frac{\partial V}{\partial n} = \frac{\partial V}{\partial r} = -\sum_{n=0}^{\infty} \frac{n + 1}{r} \left(\frac{R}{r}\right)^{n+1} V_n(\phi, \lambda).
\] (2.15)
On the Earth surface this is
\[
\left(\frac{\partial V}{\partial r}\right)_R = -\sum_{n=0}^{\infty} \frac{n + 1}{R} V_n(\phi, \lambda).
\] (2.16)
If we also write on the Earth surface (gravitation):
\[
g(\phi, \lambda, R) \doteq \left(\frac{\partial V}{\partial r}\right)_R \doteq \sum_{n=0}^{\infty} g_n(\phi, \lambda),
\]
it follows by analogy that
\[
g_n(\phi, \lambda) = -\frac{n + 1}{R} V_n(\phi, \lambda),
\]
and conversely, that
\[
V_n(\phi, \lambda) = -R g_n(\phi, \lambda) \frac{n + 1}{n+1}.
\] (2.17)
As a result of this we obtain the spectral expansion of the solution to a certain Neumann problem:
\[
V(\phi, \lambda, r) = -R \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{g_n(\phi, \lambda)}{n+1}.
\] (2.18)

\(^7\text{Carl Gottfried Neumann (1832–1925) was a German mathematician.}\)
Let us write in an analogue fashion

\[
g (\phi, \lambda, R) = \left( \frac{\partial V}{\partial n} \right)_{r=R} = \sum_{n=0}^{\infty} g_n (\phi, \lambda) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} g_{nm} Y_{nm} (\phi, \lambda),
\]
where the definition used is consistently

\[
g_{nm} = -\frac{n+1}{R} V_{nm},
\]
see equation (2.14). This is an interesting result worth thinking about. If we have available, over the whole surface area of the Earth, measurement values of gravity acceleration \( g \), we may derive from these the coefficients \( g_{nm} \) and the surface spherical harmonics \( g_n (\phi, \lambda) \) using the method explained earlier. In this way we can then obtain the solution by means of equation (2.18) for the whole exterior potential field! This is the basic idea of geopotential – or geoid – determination, from the spectral perspective.

### 2.11 Splitting a function into degree constituents

Finally we give still a useful computation formula (integral equation) for surface spherical harmonics, if the function itself \( f \) on the surface of the sphere has been given. The formula is (Heiskanen and Moritz, 1967, equation 1-71, but using our notation \( f_n \equiv Y_n \)):

\[
f_n (\phi, \lambda) = \frac{2n + 1}{4\pi} \int_{\sigma} f (\phi', \lambda') P_n (\cos \psi) d\sigma',
\]
where \( \psi \) is the angular distance between evaluation point \( (\phi, \lambda) \) and moving or integration point \( (\phi', \lambda') \). In this degree constituent equation (2.19) there is a certain similarity with the projection or coefficient computation formula (B.7). Nevertheless, here we don’t have a computation of spectral components, but of “spectral component functions” \( f_n \).

We bring to mind the core property of the functions \( f_n \)

\[
f (\phi, \lambda) = \sum_{n=0}^{\infty} f_n (\phi, \lambda)
\]
on the surface of the sphere.

For the proof, we choose as the “north pole” of the co-ordinate system the point \( (\phi, \lambda) \); then, \( \phi' = 90^\circ - \psi \). By writing (see equation (2.8)):

\[
f (\phi', \lambda') = \sum_{n=0}^{\infty} \sum_{m=0}^{n} P_{nm} (\sin \phi') (a_{nm} \cos m\lambda' + b_{nm} \sin m\lambda')
\]
and substituting this into the degree constituent equation (2.19), we obtain, by exploiting the orthogonality of the Legendre polynomials on the right-hand side of the equation:

\[
I_R = \frac{2n+1}{4\pi} \int_\sigma f(\phi', \lambda') P_n(\cos \psi) \, d\sigma' = \\
= \frac{2n+1}{4\pi} a_{n0} \int_\sigma P_n^2(\cos \psi) \, d\sigma' = \\
= \frac{2n+1}{4\pi} a_n \int_{-1}^{+1} P_n(t) \cdot \left( \sin \psi \int_{0}^{2\pi} d\lambda' \right) \left[ \frac{1}{\sin \psi} \cdot dt \right] = \\
= \frac{2n+1}{4\pi} \cdot 2\pi a_n \cdot \frac{2}{2n+1} = a_n,
\]

where we used the notation \(a_n = a_{n0}\). On the left-hand side of the equation we obtain similarly, because according to definition (2.7) \(\phi = 90^\circ\) and \(\sin \phi = 1\):

\[
I_L = f_n(\phi, \lambda) \equiv Y_n(90^\circ, \lambda) = \sum_{m=0}^{n} P_{nm}(1) (a_{nm} \cos m\lambda + b_{nm} \sin m\lambda) \\
= P_n(1) a_n = a_n,
\]

by using

\[
P_n(1) \equiv P_{n0}(1) = 1 \\
P_{nm}(1) = 0 \quad \text{if} \ m \neq 0.
\]

As this applies for every point \((\phi, \lambda)\) (and note that the value of \(a_n\) depends on this choice!) it follows that the degree constituent equation (2.19) is generally true.

### 2.12 Low-degree spherical harmonics

The potential field of a point mass is (equation (1.4)):

\[
V = \frac{GM}{r}.
\]

The corresponding term in the potential expansion (2.8) for degree \(n = 0\) is

\[
V_0 = \frac{1}{r} a_{00} P_0(\sin \phi) = \frac{a_{00}}{r},
\]

from which

\[
a_{00} = GM.
\]

i.e., \(a_{00}\) describes the force field of the centre of mass, the field of a point mass or spherically symmetric mass distribution. The higher spherical harmonic coefficients are “perturbations” on top of this.
The expansion for the degree one coefficients looks as follows:

\[ V_1(\phi, \lambda, r) = \frac{1}{r^2}(a_{11}\cos\phi \cos\lambda + b_{11}\cos\phi \sin\lambda + a_{10}\sin\phi). \]

Write this in vector form using the equation for the location vector

\[ r = (r \cos\phi \cos\lambda)i + (r \cos\phi \sin\lambda)j + (r \sin\phi)k \]

(where \( \{i, j, k\} \) is an orthonormal base of the space \( \mathbb{R}^3 \)) yielding

\[ V_1(r) = \frac{1}{r^3} \langle (a_{11}i + b_{11}j + a_{10}k) \cdot r \rangle. \]

Remember that the potential field of a dipole is

\[ V = \frac{G}{r^3} \langle d \cdot r \rangle, \]

where \( d \) is the dipole moment. Comparison yields

\[ a_{11}i + b_{11}j + a_{10}k = Gd, \]

i.e., the first degree \( n = 1 \) spherical harmonic coefficients represent the Earth’s gravitational field’s dipole moment.

Every mass element \( dm \) of our Earth may be taken to consist of

- a monopole at the origin of the co-ordinate system, size \( dm \), and
- a dipole, size \( r \cdot dm \), where \( r \) is the location vector of the mass element.

In that case we may compute the dipole moment of the whole Earth by integration:

\[ d_\oplus = \iiint_\mathcal{V} rdV = \iiint_\mathcal{V} \rho rdV = M \cdot r_{centre \ of \ mass}, \]

by definition, the location of the centre of mass of the Earth! From this follows that, if we choose our co-ordinate system so, that the origin is in the centre of mass of the Earth, the spherical harmonic coefficients \( a_{11}, b_{11}, a_{10} \) vanish. If the equations of motion of satellites are formulated in a certain co-ordinate system, like in the case of GPS satellites the WGS84 system, then the origin of the system is automatically in the centre of mass of the Earth, and the degree one spherical harmonic coefficients are really zero.

The same logic applies to higher degrees of spherical harmonics. The degree two coefficients describe the so-called quadrupole moment of the Earth – corresponding to her inertial tensor \( - \), etc.
2.13 Often used spherical harmonic expansions

Of the existing global spherical harmonic expansions must be mentioned the already somewhat dated EGM96. It was developed by researchers from Ohio State University using very extensive, mostly gravimetric, data collected by the American NIMA (National Imagery and Mapping Agency, the former Defense Mapping Agency, the current NGA, National Geospatial-Intelligence Agency). This expansion goes up to degree 360; its standard presentation is\(^8\)

\[
V = \frac{GM}{r} \left( 1 + \sum_{n=2}^{360} \left( \frac{a}{r} \right)^n \sum_{m=0}^{n} P_{nm}(\sin \phi) \left[ C_{nm} \cos m \lambda + S_{nm} \sin m \lambda \right] \right).
\]

This form of presentation – the algebraic sign in front of the expansion, which starts from degree number \(n = 2\), the number one inside the parentheses which represents the point mass in the origin equal in magnitude to the total mass of the Earth, and the “fully normalized” coefficients \(C\) and \(S\) – has been already for some time been an industry standard in the global research community in the field of computing spherical harmonic expansions of the Earth’s gravitational field. A pioneer has been Professor Richard H. Rapp at Ohio State University, which is why the models are often called OSU models.

Generally in these models the lower terms \(2 \leq n \leq 20\) are derived primarily from analysis of satellite orbit perturbations. Because of this, the models are in a co-ordinate system with the origin in the Earth’s centre of mass. This explains the absence of the degree one coefficients, as argued earlier.

\(^8\)Note that here is used \(a\), the equatorial radius of the Earth, not \(R\), and \(\phi\), i.e., the geocentric latitude. The co-ordinates \((r, \phi, \lambda)\) form a spherical co-ordinate system.
2.14. Ellipsoidal harmonics

The higher coefficients again – 20 < \( n \leq 360 \) – were before the year 2000 mostly the result of the analysis of gravimetric data (over land) and satellite radar altimetric data (over the ocean). After the launches of the gravimetric satellite missions CHAMP, GRACE and GOCE, and as a result of their measurements, nowadays also this degree number interval is the product of space geodesy. Only the still higher degree numbers – the new model EGM2008 (Pavlis et al., 2008, 2012) goes up to degree 2159 – are still due to terrestrial measurements.

In table 2.3 we give the first and last coefficients of the EGM96 model, the last and best model from the time just before the satellite gravity missions. The values tabulated are \( n, m, C_{nm}, S_{nm} \) and the mean errors (standard deviations) of both coefficients from their computation. Note that all \( S_{n0} \) vanish!

Sometimes also non-normalized coefficients are used, and we write

\[
V = \frac{GM}{r} \left[ 1 - \sum_{n=2}^{\infty} \left( \frac{d}{r} \right)^n \sum_{m=0}^{n} P_{nm} \left( \sin \phi \right) \left( J_{nm} \cos m\lambda + K_{nm} \sin m\lambda \right) \right].
\] (2.21)

Then we use the notation \( J_0 \equiv J_{00} \), and \( J_2 \) is the most important parameter of the Earth’s gravity field for describing the flattening of the Earth. The relationship with the parameters \( C, S \) is:

\[
\begin{align*}
J_{n0} &= -\sqrt{2n+1} C_{n0}, \quad K_{n0} = -\sqrt{2n+1} S_{n0}, \\
\begin{cases} J_{nm} \\ K_{nm} \end{cases} &= -\sqrt{2} \frac{(n-m)!}{(n+m)!} \begin{cases} C_{nm} \\ S_{nm} \end{cases}, \quad m \neq 0.
\end{align*}
\] (2.22)

2.14 Ellipsoidal harmonics

The Laplace equation (1.12) may be written, and solved, instead of in spherical co-ordinates, in ellipsoidal co-ordinates. The result is known as an ellipsoidal harmonic expansion. They are little used, because the math needed is more complicated. Also ellipsoidal co-ordinates are mostly only theoretically interesting and not in any broad use within geodesy.

The form of presentation is:

\[
V (u, \beta, \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{Q_{nm} (i \frac{u}{r})}{Q_{nm} (i \frac{\beta}{r})} P_{nm} (\sin \beta) \left( a_{nm} \cos m\lambda + b_{nm} \sin m\lambda \right),
\] (2.22)

where \( Q_{nm} (z) \) are the so-called Legendre functions of the second kind, sampled in table 2.4. Though the general argument \( z \) is complex, equation (2.22) gives a real result for real valued coefficients \( a_{nm}, b_{nm} \).
### Table 2.3. Coefficients of the EGM96 spherical harmonic expansion.

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>( \overline{C}_{nm} )</th>
<th>( \overline{S}_{nm} )</th>
<th>( \overline{C}_{nm} ) mean error</th>
<th>( \overline{S}_{nm} ) mean error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>-0.48165371736E-03</td>
<td>0.00000000000E+00</td>
<td>0.35610635E-10</td>
<td>0.00000000E+00</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-0.18698735955E-09</td>
<td>0.11952801291E-08</td>
<td>0.10000000E-29</td>
<td>0.10000000E-29</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.24391435239E-05</td>
<td>-0.14001668365E-05</td>
<td>0.53739154E-10</td>
<td>0.54353269E-10</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.95725417379E-06</td>
<td>0.00000000000E+00</td>
<td>0.18094237E-10</td>
<td>0.00000000E+00</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.90427768056E-06</td>
<td>0.24851315871E-08</td>
<td>0.10000000E-29</td>
<td>0.10000000E-29</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.90427768056E-06</td>
<td>-0.10625942456E-06</td>
<td>0.11182866E-09</td>
<td>0.11182866E-09</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.72102657075E-06</td>
<td>0.14135626958E-08</td>
<td>0.93285090E-10</td>
<td>0.93285090E-10</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.35973863780E-06</td>
<td>0.00000000000E+00</td>
<td>0.10423678E-09</td>
<td>0.10423678E-09</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-0.53632161971E-06</td>
<td>-0.47344026583E-06</td>
<td>0.82408489E-10</td>
<td>0.82408489E-10</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.35069410578E-06</td>
<td>-0.14001668365E-05</td>
<td>0.53739154E-10</td>
<td>0.54353269E-10</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>-0.53632161971E-06</td>
<td>0.24851315871E-08</td>
<td>0.10000000E-29</td>
<td>0.10000000E-29</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>-0.99077180392E-06</td>
<td>0.13965165E-09</td>
<td>0.13645882E-09</td>
<td>0.13645882E-09</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.68532437563E-07</td>
<td>0.00000000000E+00</td>
<td>0.87315399E-10</td>
<td>0.87315399E-10</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-0.62101212852E-07</td>
<td>-0.94226127525E-07</td>
<td>0.28082822E-09</td>
<td>0.28082822E-09</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.65243829761E-06</td>
<td>-0.32334961266E-06</td>
<td>0.24356998E-09</td>
<td>0.24356998E-09</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>-0.45195546067E-06</td>
<td>-0.21484719962E-06</td>
<td>0.16810647E-09</td>
<td>0.16810647E-09</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>-0.29530164765E-06</td>
<td>0.49665887679E-07</td>
<td>0.11849793E-09</td>
<td>0.11849793E-09</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.17491983203E-06</td>
<td>-0.66938427821E-06</td>
<td>0.11590031E-09</td>
<td>0.11590031E-09</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>-0.14995799416E-06</td>
<td>0.00000000000E+00</td>
<td>0.54383090E-10</td>
<td>0.54383090E-10</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-0.76808793849E-07</td>
<td>0.26289054550E-07</td>
<td>0.21957296E-09</td>
<td>0.21957296E-09</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0.48173244823E-07</td>
<td>-0.37372820134E-06</td>
<td>0.27693639E-09</td>
<td>0.27693639E-09</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.51739990516E-06</td>
<td>0.98269451716E-08</td>
<td>0.18682712E-09</td>
<td>0.18682712E-09</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>-0.86214260109E-07</td>
<td>0.47140815426E-07</td>
<td>0.15328004E-09</td>
<td>0.15328004E-09</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>-0.26713325490E-06</td>
<td>-0.53648834283E-06</td>
<td>0.87820965E-10</td>
<td>0.87820965E-10</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0.96761612189E-08</td>
<td>-0.23719200693E-06</td>
<td>0.11518036E-09</td>
<td>0.11518036E-09</td>
</tr>
</tbody>
</table>

Those interested in the derivation of the above equation can find it in Heiskanen and Moritz (1967) or other textbooks on potential theory. Heiskanen and Moritz give a slightly different form to the equation, the auxiliary equations needed for the normalization used here can be found on their pages 66–67.

### 2.14.1 Equivalence of the Rapp formula and the ellipsoidal formula

We can demonstrate the equivalence of equations (2.21), (2.22), if the flattening of the Earth \( \rightarrow 0 \), and thus also \( a, b \rightarrow a, \beta \rightarrow \phi \), and \( u \rightarrow r \). We assume that Heiskanen and Moritz (1967) equation (1-122),

\[
\lim_{E \rightarrow 0} \frac{Q_{nm}(i \frac{r}{a})}{Q_{nm}(i \frac{r}{E})} = \left( \frac{a}{r} \right)^{n+1}
\]
Table 2.4. Legendre functions of the second kind.

<table>
<thead>
<tr>
<th>( Q_0(z) )</th>
<th>( \frac{1}{2} \ln \frac{z+1}{z-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q_1(z) )</td>
<td>( \frac{3}{2} \ln \frac{z+1}{z-1} - 1 )</td>
</tr>
<tr>
<td>( Q_2(z) )</td>
<td>( \frac{3z^2-1}{4} \ln \frac{z+1}{z-1} - \frac{3z}{z^2} )</td>
</tr>
<tr>
<td>( Q_3(z) )</td>
<td>( \frac{5z^3-3z}{2} \ln \frac{z+1}{z-1} - \frac{5z^2}{2} + \frac{3z}{2} )</td>
</tr>
</tbody>
</table>

is valid. Substitution into eq. (2.22) yields

\[
V(u, \beta, \lambda) = V(r, \phi, \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \left( \frac{a_r}{r} \right)^{n+1} P_{nm}(\sin \phi) \left( a_{nm} \cos m\lambda + b_{nm} \sin m\lambda \right),
\]

which, despite different designations used for the coefficients, with the identifications \( a_{00} \hat{=} \frac{GM}{a}, a_{10} \hat{=} a_{11} \hat{=} b_{10} \hat{=} 0 \) corresponds to eq. (2.21) for spherical harmonics.

2.14.2 Advantages of using ellipsoidal harmonics:

1. The formula for the normal field is in this form of presentation simple, see Heiskanen and Moritz (1967) equation 2-56. A spherical harmonic expansion of the same field would instead require theoretically an infinite number of coefficients (in practice 3 to 4, i.e., an expansion up to \( J_6 \) or \( J_8 \), will suffice).

2. The convergence will be more rapid, as less terms are needed. This is because, due to the Earth’s flattening, the equator is some 23 km farther from the Earth centre than the poles. Therefore, especially high degree spherical harmonics have difficulty converging efficiently both at the poles and in the equatorial region. This problem is worst for very high degree expansions (e.g., Wenzel, 1998). Already for a degree number of 360, the semi-wavelength of a spherical harmonic will be only 55 km!

2.14.3 Disadvantage of using ellipsoidal harmonics:

Evaluation of ellipsoidal harmonics is clearly more laborious, i.e., expensive, than spherical harmonics, in terms of computer resources.
3. The normal gravity field

3.1 The basic idea of a normal field

Just like the figure of the Earth is in good approximation an ellipsoid of revolution, also the gravity field of the Earth is in just as good an approximation a field of which one equipotential surface is precisely this ellipsoid of revolution.

This brings a logical idea to mind: why not define intercompatibly a reference ellipsoid, a geopotential field or normal potential – one of the equipotential surfaces of which is the reference ellipsoid – and a gravity formula, computed

![Diagram of the normal gravity field of the Earth](image-url)

Figure 3.1. The normal gravity field of the Earth.
by taking the gradient of this normal potential?

After this we may define anomalous potential and gravity quantities, which then again will be intercompatible.

Let the normal potential be \( U(x, y, z) \) and the normal potential value at the reference ellipsoid \( U_0 \).

Then normal gravity will be

\[
\gamma(x, y, z) = -\frac{\partial U}{\partial n},
\]

where \( \frac{\partial}{\partial n} \) denotes differentiation in the direction of the external normal \( n \) to the ellipsoid. This normal will differ from the direction of the normal to the level surfaces, or plumbline, by precisely the plumbline deflection, typically a very small angle.

We shall see that the pseudo-force generated by the Earth rotation may, in a system rotating along with the Earth, be described by a rotational potential \( \Phi \). Also the normal potential \( U \) is defined in such a way, that the rotational potential \( \Phi \) is included in it: the normal potential is the reference potential of the gravity field, not the gravitational field. If we denote the normal gravitational potential by \( \Psi \) (a rarely used quantity in geodesy), then the normal gravity potential \( U \) is

\[
U = \Psi + \Phi,
\]

where \( \Phi \) is the centrifugal potential. In other words: \( \Psi \), like \( V \), is defined in a non-rotating (inertial) system, whereas \( U \), like \( W \), is defined in a system that co-rotates with the Earth and is non-inertial. Just like also the word gravity refers to a force acting in a co-rotating system, when in an inertial system we use the word gravitation.

### 3.2 The centrifugal force and its potential

The rotation of the Earth is important for the gravity field. In an inertial reference system one speaks of gravitation and gravitational potential \( V \); on the Earth surface however, in a non-inertial or co-rotating system, we talk of gravity and gravity potential \( W \). They are different things, and the rotational motion and its centrifugal force are the cause of the difference. See figure 3.2.
To derive the equation for centrifugal force, write first

\[ p = X \mathbf{i} + Y \mathbf{j}, \]

where the vectors \( \{\mathbf{i}, \mathbf{j}, \mathbf{k}\} \) form an orthonormal basis along the \((X, Y, Z)\) axes.

Then

\[ p = \|p\| = \sqrt{(p \cdot p)} = \sqrt{X^2 + Y^2}. \]

Now the centrifugal force (or rather, acceleration) is

\[ f = \omega^2 p = \omega^2 (X \mathbf{i} + Y \mathbf{j}), \]

with \( \omega \) the rotation rate in radians per second.

Here on Earth, gravity measurements are generally done with a device that is at rest with respect to the Earth surface: it follows the rotation of the Earth. If the device moves, one must, in addition to the centrifugal force, take into account another pseudo-force: the Coriolis force. Also fluids – water, air – on the Earth’s surface, if they are at rest, sense only the centrifugal force. Currents also sense the Coriolis force, which deflects them sideways and causes the well known eddy phenomena in the oceans and atmosphere.

If we forget for the moment about the Coriolis force, we may describe the centrifugal force as the gradient of a potential. If we write for this centrifugal potential

\[ \Phi = \frac{1}{2} \omega^2 (X^2 + Y^2), \]

\(^1\)Gaspard-Gustave Coriolis (1792 – 1843) was a French mathematician, physicist and mechanical engineer. His name is inscribed on the Eiffel Tower.
we may directly calculate that
\[ f = \nabla \Phi = i \frac{\partial \Phi}{\partial x} + j \frac{\partial \Phi}{\partial y} + k \frac{\partial \Phi}{\partial z} = \frac{1}{2} i \omega^2 \cdot 2x + \frac{1}{2} j \omega^2 \cdot 2y + 0 = \omega^2 (iX + jY), \]
which corresponds to the above centrifugal force equation.

If we start out with the gravitational potential \( V \) and add to it the centrifugal potential \( \Phi \), we obtain the gravity potential \( W \):
\[ W = V + \Phi. \]

We may also derive from the centrifugal potential \( \Phi \) the following equation by differentiating it twice:
\[ \Delta \Phi = \nabla^2 \Phi = \nabla f = \frac{\partial}{\partial x} \omega^2 x + \frac{\partial}{\partial y} \omega^2 y + 0 = 2\omega^2, \quad (3.1) \]
from which follows, with the Poisson equation (1.14),
\[ \Delta W = -4\pi G \rho + 2\omega^2, \quad (3.2) \]
the Poisson equation for the gravity potential.

The difference between gravitation and gravity is essential. The force (acceleration) of gravitation \( a = -\nabla V \) is just an attractive force; gravity (acceleration) \( g = -\nabla W \) is the resultant of gravitation and centrifugal force. Attraction and centrifugal force act in the same fashion; the force is proportional to the mass of the test object, in other words, the acceleration is always the same independently of the mass of the test object. This is the famous equivalence principle (Galileo, Einstein), which has been proven to hold to very great precision. Especially we may mention the clever tests by the Hungarian baron Loránd Eötvös\(^2\).

Water masses on the Earth surface, like also the atmosphere (and on a vastly longer time scale also the “solid” rock forming mountain ranges and ocean depths) react to gravity without distinguishing between attraction and centrifugal force. For this reason the sea surface coincides within a metre or so with an equipotential surface of the \( W \) function. Also on dry land we measure heights from this surface, the geoid (Gauss: “the mathematical figure of the Earth”).

The unit of measurement of gravity is \( \text{mGal} = 10^{-5} \text{m/s}^2 \). Also \( \mu \text{Gal} \) or \( 10^{-8} \text{m/s}^2 \) is used. In modern books are also used \( \text{m/s}^2 \) ja \( \text{nm/s}^2 \), which formally

\(^2\)Loránd baron Eötvös de Vásárosnamény (1848–1919) was a Hungarian physicist and student of gravitation.
belong to the SI system. Nevertheless, milligals and microgals are more familiar still, and correspond to 1 ppm (part per million) and 1 ppb (part per billion) of ambient gravity.

A popular unit for measuring gravity gradients is the Eötvös. In SI units it is $10^{-9} \text{s}^{-2}$, corresponding to $10^{-4} \text{mGal/m}$. In table 3.1 we give a few values in order to get an idea of the orders of magnitude of phenomena. On the Earth surface the vertical gradient $\partial g/\partial h$ is on average some $-0.3 \text{mGal/m}$.

### 3.3 Level surfaces and plumblines

Surfaces of the same gravity potential, equipotential surfaces or level surfaces are the following surfaces:

$$W(x,y,z) = W_0 = \text{const.}$$

Let \(\{i,j,k\}\) again be an orthonormal base along the \((x,y,z)\) axes. Then, in the direction of the unit vector

$$\mathbf{e} = e_1 \mathbf{i} + e_2 \mathbf{j} + e_3 \mathbf{k}$$

the potential changes as follows:

$$\frac{\partial W}{\partial \mathbf{e}} = e_1 \frac{\partial W}{\partial x} + e_2 \frac{\partial W}{\partial y} + e_3 \frac{\partial W}{\partial z},$$

which vanishes if and only if

$$\langle \mathbf{e} \cdot \mathbf{\nabla} W \rangle = 0,$$

in other words, the potential is stationary only in directions that are perpendicular to the Earth’s gravity vector

$$\mathbf{\nabla} W = \mathbf{g}.$$
Chapter 3. The normal gravity field

The normal gravity field

\[ P \in xW = W_P - \delta W \]

\[ W = W_P \]

\[ \rho_1 \]

\[ x_0 \]

\[ x \]

\[ \epsilon \]

Figure 3.3. The curvature of level surfaces.

So: level surfaces and gravity vectors, or plumblines, are always perpendicular to each other.

Curvature of level surfaces: Given in point \( P \) a plane that in \( P \) has the same direction as the level surface, i.e., a tangent plane. If the local curvature of the level surface in the \( x \) direction is \( \rho_1 \), and the \( x \) co-ordinate of point \( P \) is \( x_0 \), we may develop the distance between the surfaces in a Taylor series:

\[ \epsilon = \frac{1}{2\rho_1} (x - x_0)^2. \]

From this we obtain the difference in \( W \) values between the surfaces (\( g = \|g\| \)):

\[ \delta W = -\epsilon g = - (x - x_0)^2 \frac{g}{2\rho_1}. \]

By differentiating (note that \( W \) here is now the geopotential on the tangent plane) we obtain

\[ \frac{\partial^2}{\partial x^2} \delta W = \frac{\partial^2}{\partial x^2} W = W_{xx} = - \frac{g}{\rho_1} \]

from which (with \( W_{xx} = \frac{\partial^2}{\partial x^2} W \))

\[ \rho_1 = - \frac{g}{W_{xx}}, \]

where the \( W \) that is being differentiated with respect to the \( x \) co-ordinate is its restriction to the tangent plane.
By determining the curvature in the $x$ direction

$$K_1 = \frac{1}{\rho_1} = -\frac{W_{xx}}{g},$$

and similarly in the $y$ direction

$$K_2 = \frac{1}{\rho_2} = -\frac{W_{yy}}{g},$$

we obtain the mean or Germain\(^3\) curvature (a positive number):

$$J = \frac{1}{2}(K_1 + K_2) = -\frac{W_{xx} + W_{yy}}{2g},$$

and by using Poisson, equation (3.2),

$$\Delta W = W_{xx} + W_{yy} + W_{zz} = -4\pi G \rho + 2\omega^2,$$

we obtain

$$-2gJ + W_{zz} = -4\pi G \rho + 2\omega^2.$$  

By using ($H$ = height co-ordinate)

$$W_{zz} = -\frac{\partial g}{\partial z} = -\frac{\partial g}{\partial H}$$

we obtain (Heiskanen and Moritz, 1967, kaava 2-20):

$$\frac{\partial g}{\partial H} = -2gJ + 4\pi G \rho - 2\omega^2,$$

an equation found by Ernst Heinrich Bruns.

### 3.4 Natural co-ordinates

Before the satellite era it was impossible to directly measure the geocentric $X, Y$ and $Z$. Today this is possible, and we obtain at the same time $h$, a purely geometric quantity.

In earlier times one could measure only the direction of the plumbline as shown in figure 3.4, and the potential difference between an observation point and sea level. The direction of the plumbline $n$ was measured astronomically: astronomical latitude $\Phi$ (don’t confuse with the centrifugal potential) and astronomical longitude $\Lambda$. The third co-ordinate, the potential

---

\(^3\)Marie-Sophie Germain (1776–1831) was a brilliant French mathematician, number theorist and student of elasticity. She corresponded with Gauss, among others, on number theory, and did foundational work toward a proof of Fermat’s last theorem. Her name is missing from the Eiffel Tower.
Chapter 3. The normal gravity field

The normal gravity field $\Phi, \Lambda$

...co-ordinates $\Phi, \Lambda$.

Additionally, a natural height co-ordinate, e.g., the geopotential $W$ is needed.

These co-ordinates, $\Phi, \Lambda$, and $W$, are called natural co-ordinates.

Often, instead of the potential, orthometric height is used. Its definition is easy to understand if one writes

$$\frac{\partial W}{\partial n} = \frac{\partial W}{\partial H} = -g \implies dH = -\frac{1}{g}dW \implies H_P = -\int_{W_0}^{W_P} \frac{1}{g(W')} dW',$$

where the integral is taken along the plumbline of point $P$. $\frac{\partial}{\partial H} = \frac{\partial}{\partial n}$ is the derivative in the direction of the plumbline, i.e., the local normal to the level surfaces. $g$ is the acceleration of gravity along the plumbline as a function of place – or of geopotential level. This means, in this case of orthometric heights, the true gravity inside the rock, which is a non-linear function of place and will also depend on rock density. This is a problem specific to orthometric heights. We’ll return to this later on (Heiskanen and Moritz, 1967 chapter 4).

Also the co-ordinates $\Phi, \Lambda, H$ form a natural co-ordinate system.

3.5 The normal potential in ellipsoidal co-ordinates [difficult]

We already presented an equation (2.22) for the expansion of the geopotential into ellipsoidal harmonics. It is demanded of the normal potential...
3.5. The normal potential in ellipsoidal co-ordinates [difficult]

$U$, that it is a constant on the reference ellipsoid $u = b$. We expand the centrifugal force $\Phi$ into ellipsoidal harmonics. We have

$$\Phi(u, \beta) = \frac{1}{2} \omega^2 (x^2 + y^2) = \frac{1}{2} \omega^2 (u^2 + E^2) \cos^2 \beta =$$

$$= \frac{1}{2} \omega^2 (u^2 + E^2) (1 - \sin^2 \beta) =$$

$$= \frac{1}{2} \omega^2 (u^2 + E^2) \left( -\frac{2}{3} P_2 (\sin \beta) + \frac{2}{3} P_0 (\sin \beta) \right) =$$

$$= -\frac{1}{3} \omega^2 (u^2 + E^2) (P_2 (\sin \beta) - P_0 (\sin \beta)).$$

Additionally we have, based on equation (2.22), for a rotationally symmetric gravitational potential $\Psi$:

$$\Psi(u, \beta) = \sum_{n=0}^{\infty} \frac{Q_n (i \frac{u}{E})}{Q_n (i \frac{b}{E})} A_n P_n (\sin \beta),$$

as well as

$$U(u, \beta) = \Phi(u, \beta) + \Psi(u, \beta).$$

On the reference ellipsoid we have as a requirement $U(b, \beta) = U_0$, which is possible only if

$$A_0 + \frac{1}{3} \omega^2 (b^2 + E^2) = U_0,$$

$$A_2 - \frac{1}{3} \omega^2 (b^2 + E^2) = 0,$$

and all other $A_n = 0$.

The quantity $U_0$ can be computed uniquely, if the Earth’s mass $GM$ and the measures of the reference ellipsoid $a$, $b$ are known.

The result, given in Heiskanen and Moritz (1967) equation 2-61, is:

$$U_0 = \frac{GM}{E} \arctan \frac{E}{b} + \frac{1}{3} \omega^2 a^2.$$

From this follows, using $a^2 = b^2 + E^2$:

$$A_0 = U_0 - \frac{1}{3} \omega^2 a^2 = \frac{GM}{E} \arctan \frac{E}{b}.$$

Assume $A_0 = a_0$ (the zeroeth degree coefficients of normal and true fields are the same, i.e., the mass of the normal field is realistic), and $a_{10} = a_{11} = b_{11} = 0$ (the vanishing of the dipole moment!).

After this, we may “scale” equation (2.22) as follows – By using equation $Q_0 (i \frac{b}{E}) = -i \arctan \frac{E}{b}$ (Heiskanen and Moritz, 1967, p. 66) and moving
suitable constants into the new coefficients \( c_{nm}^e, s_{nm}^e \):

\[
V(u, \beta, \lambda) = \frac{GM}{E^2} \arctan \frac{E}{u} \left[ 1 - \sum_{n=2}^{\infty} \sum_{m=0}^{n} \frac{Q_{nm}(i_E^m)}{Q_{nm}(i_E^n)} \frac{P_{nm}((\sin \beta))}{(c_{nm}^e \cos m \lambda + s_{nm}^e \sin m \lambda)} \right],
\]

(3.5)

which agrees with the spherical harmonic expansion (2.20). This equation has however apparently not been used for any geopotential computation.

The gravity field normal potential \( U \) is obtained as follows (remember that \( a^2 = b^2 + E^2 \)):

\[
U(u, \beta) = \frac{GM}{E} \arctan \frac{E}{u} + \frac{1}{3} \omega^2 a^2 \frac{Q_2(i_E^m)}{Q_2(i_E^n)} \left( \frac{3}{2} \sin^2 \beta - \frac{1}{2} \right) + \frac{1}{2} \omega^2 (u^2 + E^2) \cos^2 \beta = C_1(u) + C_2(u) \sin^2 \beta + C_3(u) \cos^2 \beta,
\]

where \( C_1, C_2, C_3 \) are suitable functions of \( u \).

On the surface of the reference ellipsoid \( u = b \):

\[
U(b, \beta) = \left( \frac{GM}{E} \arctan \frac{E}{b} - \frac{1}{6} \omega^2 a^2 \right) + \frac{1}{2} \omega^2 a^2 \sin^2 \beta + \frac{1}{2} \omega^2 a^2 \cos^2 \beta = GM \arctan \frac{E}{b} + \frac{1}{3} \omega^2 a^2,
\]

a constant, as it better be!

### 3.6 Normal gravity

Without proof we mention that for normal gravity (the quantity \( \gamma = \partial U/\partial h \)) the following equation applies on the reference ellipsoid:

\[
\gamma = \frac{a \gamma_b \sin^2 \beta + b \gamma_a \cos^2 \beta}{\sqrt{a^2 \sin^2 \beta + b^2 \cos^2 \beta}}.
\]

By substitution we find immediately, that \( \gamma_a \) is normal gravity on the equator \( (\beta = 0) \) and \( \gamma_b \) normal gravity on the poles \( (\beta = \pm 90^\circ) \).

By using equations (2.3) and (2.5) we obtain

\[
\tan \beta = \frac{\sin \beta}{\cos \beta} = \frac{Z/b}{\sqrt{X^2 + Y^2} / a} = \frac{a}{b} \tan \phi
\]
3.6. Normal gravity

Figure 3.5. The geometry of the meridian ellipse and various types of latitude.

and

\[
\tan \varphi = \frac{\sin \varphi}{\cos \varphi} = \frac{Z/(1-e^2)N}{\sqrt{X^2+Y^2}/N} = \frac{Z}{\sqrt{X^2+Y^2}} \frac{1}{1-e^2} = \frac{a^2}{b^2} \tan \phi,
\]

where \( \phi \) is the geocentric latitude, see eq. (2.2). From this follows directly:

\[
\tan \beta = \frac{b}{a} \tan \varphi,
\]

where the latitude angle \( \varphi \) is the geodetic (or geographic) latitude. (\( \beta \) is still the so-called reduced latitude). Now it is easy to show (exercise!) that

\[
\gamma = \frac{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}}.
\]

This is the famous Somigliana-Pizzetti\(^4\) equation. These geodesists showed for the first time that an “ellipsoidal” normal field, which has the reference ellipsoid as one of its equipotential surfaces, exists exactly and that also in geographical co-ordinates the gravity formula is a closed expression in latitude.

\(^4\)Carlo Somigliana (1860–1955) was an Italian mathematician and physicist; Paolo Pizzetti (1860–1918) was an Italian geodesist.


Chapter 3. The normal gravity field

3.7 Numerical values and formulas

When the reference ellipsoid has been chosen, we may calculate the normal potential and normal gravity corresponding to it. The fundamental quantities are

\[ a \] the equatorial radius of the ellipsoid of revolution, or its semi-major axis;

\[ f \] the flattening, \( f = \frac{a-b}{a} \), where \( b \) is the polar radius or semi-minor axis;

\[ \omega \] the rotation rate;

\[ GM \] the total mass (including the atmosphere).

Alternatively one may choose also \( \gamma_a \), i.e., equatorial gravity.

Nowadays the most commonly used reference system is GRS80, the Geodetic Reference System 1980:

\[
\begin{align*}
 a &= 6378137 \text{ m}, \\
 \frac{1}{f} &= 298.257222101, \\
 \omega &= 7292115 \cdot 10^{-11} \text{ s}^{-1}, \\
 GM &= 3986005 \cdot 10^8 \text{ m}^3 \text{ s}^{-2}.
\end{align*}
\]

In reality \( f \) is not a defining constant of GRS80, but instead is used \( J_2 \), which is a defining quantity for the gravity field, see equation (2.21).

The system WGS84 (World Geodetic System 1984) used by the GPS system is almost identical to GRS80.

The normal potential is (Heikkinen, 1981), units [m] and [s]:

\[
U = 62636860.8500 + \\
+ \left[ -9.78032677 - 0.05163075 \sin^2 \varphi - \right. \\
- 0.00022761 \sin^4 \varphi - 0.00000123 \sin^6 \varphi \left. \right] h + \\
+ \left[ 0.01543899 \cdot 10^{-4} - 0.00002195 \cdot 10^{-4} \sin^2 \varphi \right. \\
- 0.00000010 \cdot 10^{-4} \sin^4 \varphi \left. \right] h^2 - \\
- \left[ -0.00002422 \cdot 10^{-8} + 0.00000007 \cdot 10^{-8} \sin^2 \varphi \right] h^3,
\]


and normal gravity (note the minus sign; $U$ is positive and diminishes going upward):

$$\gamma = - \frac{\partial U}{\partial h} =$$

$$= + 9.78032677 + 0.05163075 \sin^2 \varphi +$$

$$+ 0.00022761 \sin^4 \varphi + 0.00000123 \sin^6 \varphi -$$

$$- \begin{bmatrix}
0.03087798 \cdot 10^{-4} - 0.00004390 \cdot 10^{-4} \sin^2 \varphi \\
-0.00000020 \cdot 10^{-4} \sin^4 \varphi
\end{bmatrix} h -$$

$$- \left[-0.00007265 \cdot 10^{-8} + 0.00000021 \cdot 10^{-8} \sin^2 \varphi\right] h^2. \quad (3.7)$$

More precise equations can be found from Heikkinen (1981). In these equations, the coefficient $9.78032 \ldots$ is equatorial gravity, and $0.03087 \ldots$ is the (equatorial) vertical gradient of gravity. All units are in the SI system. $\varphi$ is (geodetic) latitude, $h$ is height above the reference ellipsoid.

Other gravity formulas (and reference ellipsoids) still in legacy use (and slowly vanishing) are Helmert’s formula (Krasowsky’s ellipsoid) in the countries of Eastern Europe, the International or Hayford ellipsoid (1924) and its gravity formula, and Geodetic Reference System 1967.

### 3.8 Example

According to the above equation, the normal potential over the equator is

$$U = 62636860.8500 - 9.78032677h + 0.01543899 \cdot 10^{-4} h^2 + 0.00002422 \cdot 10^{-8} h^3.$$

- Draw this function for values of $h$ in the range $0 - 7,000$ km.
- Draw for comparison the quadratic version, from which the last term is left off.

#### Questions

1. What are the minima of these functions?
2. How physically realistic is this?

#### Answers

1. See figure 3.6. The minima are at heights 3000 km and 2000 km, approximately.
Chapter 3. The normal gravity field

2. Not very physical: the stationary point for potential $U$ (the normal potential in a co-rotating frame) should be located at approx. 36,000 km height, at the geostationary orbit. This tells us that polynomial approximation cannot be extrapolated very far. In this case the interval of extrapolation is of the same order as the radius of the Earth, and that won’t work any more.

3.9 The normal potential as a spherical harmonic expansion

The spherical harmonic expansion of an ellipsoidal gravitational field contains, besides the second degree harmonic, also higher degree harmonics. If we write, as is customary, the potential outside the Earth in the following form (Heiskanen and Moritz, 1967 section 2-39, also equation (2.21)):

$$ V = \frac{GM}{r} \left\{ 1 - \sum_{n=2}^{\infty} \left( \frac{a}{r} \right)^n \sum_{m=0}^{n} P_{nm}(\sin \phi) \left[ j_{nm} \cos m\lambda + k_{nm} \sin m\lambda \right] \right\}, $$

we may also write the normal gravitational potential, $\Psi$, into the form

$$ \Psi = \frac{GM}{r} \left[ 1 - \sum_{n=1}^{\infty} f_{2n} \left( \frac{a}{r} \right)^{2n} P_{2n}(\sin \phi) \right], $$

Figure 3.6. The normal field’s potential curve over the equator.
Table 3.2. GRS80 normal potential spherical harmonic coefficients.

<table>
<thead>
<tr>
<th></th>
<th>Non-normalized</th>
<th>Fully normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_2 = J_{20}$</td>
<td>$1082.63 \cdot 10^{-6}$</td>
<td>$484.16685 \cdot 10^{-6}$</td>
</tr>
<tr>
<td>$J_4 = J_{40}$</td>
<td>$-2.090948 \cdot 10^{-6}$</td>
<td>$-0.79030406 \cdot 10^{-6}$</td>
</tr>
<tr>
<td>$J_6 = J_{60}$</td>
<td>$+0.00506175 \cdot 10^{-6}$</td>
<td>$+0.00168725 \cdot 10^{-6}$</td>
</tr>
</tbody>
</table>

which contains only even coefficients $J_{2n} = J_{2n,0}$ as the normal field is symmetric about the equatorial plane.

The coefficients for the GRS80 normal gravitational potential are found in table 3.2. Higher terms are usually not needed. The relationship between fully normalized and non-normalized coefficients is $J_n = J_n \sqrt{2n + 1}$.

(Notes that in the expansion of the same field into ellipsoidal harmonics only the degree zero and degree two coefficient are non-zero! This is the main reason why these functions are used at all.)

Instead of using an ellipsoidal model, we may use as a normal gravity potential formula also the first two, three terms of a spherical harmonic expansion. Then we obtain (taking the centrifugal potential along):

$$U = \frac{Y_0}{r} + \frac{Y_2(\phi, \lambda)}{r^3} + \frac{1}{2} \omega^2 (X^2 + Y^2),$$

with the corresponding equipotential surface being the “Bruns spheroid”; or

$$U = \frac{Y_0}{r} + \frac{Y_2(\phi, \lambda)}{r^3} + \frac{Y_4(\phi, \lambda)}{r^5} + \frac{1}{2} \omega^2 (X^2 + Y^2),$$

the “Helmert spheroid”.

These equations are easy to compute, but their equipotential surfaces are not exactly ellipsoids of revolution. The are, in fact, quite complicated surfaces (Heiskanen and Moritz, 1967, 2-12) By including a couple more terms ($Y_6, Y_8$) we obtain an already very precise practical approximation to the ellipsoidal gravity field.

However, in geometric geodesy we always use a reference ellipsoid, so this is also a wise thing to do in physical geodesy.

### 3.10 The disturbing potential

Write the gravity potential

$$W = V + \Phi,$$
where $\Phi$ is the centrifugal potential (see above), and the normal potential

$$U = \Psi + \Phi.$$  

The difference between them is

$$T = W - U = V - \Psi,$$  

the \textit{disturbing potential}.

Both $V$ and $\Psi$ can be expanded into spherical harmonics; if we write the gravity potential

$$W = \Phi + \frac{GM}{r}$$

$$= \Phi + GM \left\{ 1 - \sum_{n=2}^{\infty} \left( \frac{a}{r} \right)^n \sum_{m=0}^{n} P_{nm} (\sin \phi) \left[ J_{nm} \cos m\lambda + K_{nm} \sin m\lambda \right] \right\},$$

and the normal potential

$$U = \Phi + \frac{GM}{r}$$

$$= \Phi + \frac{GM}{r} \left\{ 1 - \sum_{n=2,\text{even}}^{\infty} \left( \frac{a}{r} \right)^n J_n^* P_n (\sin \phi) \right\},$$

we obtain by subtraction for the disturbing potential

$$T = W - U =$$

$$= - \frac{GM}{r} \left\{ \sum_{n=2}^{\infty} \left( \frac{a}{r} \right)^n \sum_{m=0}^{n} P_{nm} (\sin \phi) \left[ J_{nm} \cos m\lambda + \delta K_{nm} \sin m\lambda \right] \right\},$$

where

$$\left\{ \begin{array}{ll} \delta J_{n0} = J_{n0} - J_n^* & \text{if } n \text{ even}, \\
\delta J_{nm} = J_{nm} & \text{if } n \text{ odd}; \text{ and} \\
\delta K_{nm} = K_{nm}. \end{array} \right.$$  

The above equation for the disturbing potential $T$ is shortened as follows (Heiskanen and Moritz, 1967, equation 2-152):

$$T (\phi, \lambda, r) = \sum_{n=0}^{\infty} \left( \frac{a}{r} \right)^{n+1} T_n (\phi, \lambda), \quad (3.8)$$

where in every term $T_n$ has the same dimension as $T$. On the surface of the reference sphere of radius $a$:

$$T = \sum_{n=0}^{\infty} T_n,$$

from which we see, that the terms $T_n (\phi, \lambda)$ are really the \textit{partial potentials} for a certain degree number $n$ on the reference level.

\footnote{Earlier on we have used for this reference radius (in spherical approximation) also the notation $R$.}
3.11 Exercises

3.11.1 Somigliana-Pizetti formula

1. Given gravity on the equator \( \gamma_a \) and on the poles \( \gamma_b \). What is gravity on geodetic latitude \( \varphi = 45^\circ \)?

2. And what is gravity on the reduced latitude \( \beta = 45^\circ \)?

3. Given the semi-major axis \( a \) and semi-minor axis \( b \), what are the differences between the different latitudes (geodetic \( \varphi \), geocentric \( \phi \), and reduced \( \beta \)) at most? You may assume that the maximum happens at latitudes \( \pm 45^\circ \).

4. Compute for the above quantities numerical values in the case of GRS80.
Chapter 3. The normal gravity field
4. Anomalous quantities of the gravity field

4.1 Disturbing potential, geoid height, deflections of the plumbline

The first anomalous quantity, which we already discussed above, is the difference between the true gravity potential $W$ and the normal gravity potential $U$:

$$T = W - U.$$ 

All other anomalous quantities are various functions of this disturbing potential, like the geoid height $N$ and the plumbline deflections $\xi, \eta$. They are generally obtained by subtracting from each other

1. natural quantities related to the Earth’s real gravity field, and
2. corresponding quantities related to the normal gravity potential of the reference ellipsoid of the Earth.

For example deflections of the plumbline

$$\xi = \Phi - \varphi, \quad \eta = (\Lambda - \lambda) \cos \varphi,$$

where $(\Phi, \Lambda)$ are astronomical latitude and longitude, i.e., the direction of the local plumbline, and $(\varphi, \lambda)$ is correspondingly the direction of the normal on the reference ellipsoid. See figure 4.1.

The geoid height or geoid undulation is:

$$N = H - h,$$

where $H$ is the orthometric height (from mean sea level) and $h$ the height from the reference ellipsoid.
Chapter 4. Anomalous quantities of the gravity field

Figure 4.1. Geoid undulations and deflections of the plumbline.

Deflections of the plumbline are in Finland a few seconds of arc in magnitude, geoid undulations range from 15 to 32 m (for comparison, globally the range is $-107 \ldots +85$ m), relative to the GRS80 ellipsoid as is today customary. At sea level, the plumbline deflections equal the horizontal gradients of the geoid undulation. See figures 4.1, 4.2.

For any reference ellipsoid, e.g., the GRS80 ellipsoid, there exists its own mathematically exact standard or normal gravity field, of which one equipotential surface is precisely that reference ellipsoid. With the aid of this field we may calculate for each gravity field quantity the corresponding normal quantity, and by subtracting the two from each other we obtain again the corresponding anomalous quantity.

For heights above the reference ellipsoid there exists an analogous formula to that for orthometric heights, where $U$ is the normal potential and $\gamma$ normal gravity:

$$h_P = -\int_{U_0}^{U_P} \frac{1}{\gamma(U)} dU.$$  

The geoid height of point $P$ is now

$$N_P = h_P - H_P = \int_{W_0}^{W_P} \frac{1}{\gamma} dW - \int_{U_0}^{U_P} \frac{1}{\gamma} dU = \int_{W_0}^{W_P} \frac{1}{\gamma} dW - \int_{W_0}^{W_P} \frac{1}{\gamma} dU - \int_{W_0}^{U_P} \frac{1}{\gamma} dU + \int_{W_0}^{U_0} \frac{1}{\gamma} dU$$

by re-naming the integration variables $W, U \rightarrow W', U'$ and changing it to a metric one: $dW' = g dH$.

In equation (4.1) the last term vanishes if we assume $U_0 = W_0$. Now in the

This is not self-evident! In a local vertical datum the potential of the zero point
4.1. Disturbing potential, geoid height, deflections of the plumbline

Figure 4.2. A geoid model for Finland from 1984. Deflections of the plumbline from observations in red (Vermeer, 1984).

definition of geoid heights, point $P$ is at the level of mean sea level (the zero point of the height system). It follows that also the first term vanishes (it would in any case always be small, except in the mountains). So

$$N_P = -\int_{W_P}^{U_P} \frac{1}{\gamma} dU \approx \frac{1}{\gamma_P} (W_P - U_P) = \frac{T_P}{\gamma_P} \quad \text{or} \quad N = \frac{T}{\gamma} \quad (4.2)$$

where we have substituted $T = W - U$, the disturbing potential. All quantities are assumed to be at sea level. This is the famous Bruns² equation (Heiskanen and Moritz, 1967, equation 2-144).

could well differ even as much as a metre from the normal potential of a global reference ellipsoid.

²Ernst Heinrich Bruns (1848–1919) was an eminent German mathematician and mathematical geodesist.
Chapter 4. Anomalous quantities of the gravity field

\[ \vec{\gamma} = \nabla U \]

We give figure 4.3 to even better depict the situation. In this figure, the gradient vectors \( g = \nabla W \) and \( \vec{\gamma} = \nabla U \) have lengths \( \partial W / \partial H \) and \( \partial U / \partial h \), from which it follows, with equation \( T = W - U \), that the separation between “matching” surfaces \( W = W_P \) and \( U = U_Q \), when \( W_P = U_Q \), is

\[
N = \frac{U_Q - U_P}{\gamma} = \frac{W_P - U_P}{\gamma} = \frac{T}{\gamma}.
\]

4.2 Gravity disturbances

The difference between the true and normal gravity accelerations is called the gravity disturbance, \( \delta g \). An exact equation would be

\[
\delta g = - \left( \frac{\partial W}{\partial H} - \frac{\partial U}{\partial h} \right);
\]

where differentiation takes place along the plumbline for \( W \), and along the ellipsoidal normal for \( U \). The directions of plumbline and ellipsoidal normal are actually very close to each other.

In spherical approximation we have

\[
\delta g = - \left( \frac{\partial W}{\partial r} - \frac{\partial U}{\partial r} \right) = - \frac{\partial T}{\partial r}.
\]

We already expanded the disturbing potential into components for different spherical harmonic degree numbers – equation (3.8) –, and now we obtain
4.3. Gravity anomalies

by differentiating with respect to $r$:

$$
\delta g(\phi, \lambda) = -\frac{\partial T(\phi, \lambda)}{\partial r} = -\frac{\partial}{\partial r} \left[ \sum_{n=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} T_n(\phi, \lambda) \right] = 
$$

$$
= \frac{1}{r} \sum_{n=0}^{\infty} (n+1) \left( \frac{R}{r} \right)^{n+1} T_n(\phi, \lambda),
$$

or on the Earth’s surface ($r = R$):

$$
\delta g(\phi, \lambda) = \frac{1}{R} \sum_{n=0}^{\infty} (n+1) T_n(\phi, \lambda).
$$

This is the spectral representation of the gravity disturbance on the Earth’s surface – more precisely, on a sphere of radius $R$. As a mean value for $R$ one may take the Earth’s equatorial radius $a$.

We can observe gravity disturbances only if, in addition to measuring the acceleration of gravity $g_P = \left( \frac{\partial W}{\partial H} \right)_{P}$ in point $P$, we have a way to measure $P$’s location in space, relative to the geocentre, so one may calculate normal gravity $\gamma_P = \left( \frac{\partial U}{\partial h} \right)_{P}$. Nowadays this is even easy using GPS, but traditionally it has been impossible. For this reason gravity disturbances are little used. One rather uses gravity anomalies, about which more below.

4.3 Gravity anomalies

Normal gravity is calculated as a function of geodetic co-ordinates $(X, Y, Z)$, or $(\phi, \lambda, h)$. However, in traditional gravimetric field work, before the GPS era, one only had access to the geodetic co-ordinates $\phi$ and $\lambda$, not the height $h$ above the reference ellipsoid. One only had access to the height $H$ above sea level (the geoid), obtained, e.g., though a national levelling network – or, worst case, barometrically.

This means that, though the true gravity $g$ is measured in point $P$ the height of which above sea level is $H_P$, normal gravity $\gamma$ is calculated in another point $Q$, the height of which above the reference ellipsoid is $h_Q = H_P$. See figure 4.4.

In other words, the measured height of point $P$ above sea level is substituted, brute-force style, into the normal gravity formula, that however expects a height above the reference ellipsoid! This special trait of the definition of gravity anomalies may be called a “free boundary-value problem”.
Chapter 4. Anomalous quantities of the gravity field

According to this we calculate gravity anomalies as follows:

\[
\Delta g_P = g_P - \gamma_Q = (g_P - \gamma_P) + (\gamma_P - \gamma_Q) = \\
\frac{-\partial W_P}{\partial H} - \frac{\partial U_P}{\partial h} - \frac{\partial U_Q}{\partial h} \\
\approx -\frac{\partial (W_P - U_P)}{\partial H} + (h_P - h_Q) \frac{\partial \gamma}{\partial h} = \\
\frac{-\partial T_P}{\partial H} + (h_P - H_P) \frac{\partial \gamma}{\partial H} = \left[ \frac{\partial T}{\partial H} + \frac{T}{\gamma} \frac{\partial \gamma}{\partial H} \right]_P,
\]

using almost all equations above.

That last equation looks familiar: it is the boundary condition of the third boundary-value problem (Heiskanen and Moritz, 1967, section 1-17). It enables the solution of \( T \) in the exterior space, if \( \Delta g \) is given everywhere on the Earth’s surface.

If we assume that the Earth’s normal gravity field is spherically symmetric, we may approximate (exercise: show this!):

\[
\Delta g = \frac{-\partial T}{\partial r} - \frac{2}{r} T,
\]

where \( r = R + H \) is the distance from the Earth’s centre.

By substituting into this the equation for \( \delta g \), and \( r = R \), we obtain on the Earth’s surface:

\[
\Delta g = \delta g - \frac{2}{R} T.
\]

from this we obtain directly by using the above spectral representations for \( T \) and \( \delta g \):

\[
\Delta g = \frac{1}{R} \sum_{n=0}^{\infty} ((n + 1) - 2) T_n = \frac{1}{R} \sum_{n=2}^{\infty} (n - 1) T_n.
\]
(Here it is assumed that $T_0$, the average of the disturbing potential over the whole Earth’s surface, is zero. Obviously also $T_1$ may be neglected.) Sometimes the following presentation is chosen:

$$\Delta g = \sum_{n=2}^{\infty} \Delta g_n,$$

where

$$\Delta g_n = \frac{n-1}{R} T_n.$$  \hfill (4.6)

From this it is seen that gravity anomalies cannot contain $n = 1$ components. It is always wise to choose the origin of the co-ordinate system to be in the centre of mass of the Earth, but if it is not, at least gravity anomalies do not change.

Equation (4.6) applies only on a spherical Earth of radius $R$. Outside the Earth we obtain, using equations (4.3) ja (4.4), the corresponding equation:

$$\Delta g = \frac{1}{r} \sum_{n=2}^{\infty} \left( \left( \frac{R}{r} \right)^{n+1} T_n \right) = \frac{1}{r} \sum_{n=2}^{\infty} (n-1) \left( \frac{R}{r} \right)^{n+1} T_n = \left( \frac{R}{r} \right)^{n+2} \Delta g_n.$$  \hfill (4.7)

### 4.4 The boundary-value problem of physical geodesy

As we explained in the above section, gravimetric measurement is a bit more complicated than just measuring the quantity $\frac{\partial W}{\partial r}$. This is because, even if we measure the radial derivative of the potential, we do it in a place we don’t precisely know. And even if we knew the height of the measurement location above sea level, that still doesn’t give us the measurement point’s location in space, which depends additionally on the location in space of sea level, i.e., the geoid – an equipotential surface of the geopotential field –, specifically its height above or below the reference ellipsoid.

This is how we arrive at the third boundary-value problem. The **boundary-value problem of physical geodesy** is to determine the potential $V$ outside a body if given on its surface is a linear combination

$$aV + b \frac{\partial V}{\partial n},$$

\hfill (4.8)

The third or mixed boundary-value problem is associated with Victor Gustave Robin (1855–1897), a French mathematician. Then, the Dirichlet problem could be called the first, the Neumann problem the second boundary-value problem.
with $a, b$ suitable constants. The variable $n$ represents here direction of the normal to the Earth surface, in practice the same as $r$ or $h$.

In physical geodesy is given the following linear combination (gravity anomaly, equation (4.5)):

$$
\Delta g = -\frac{\partial T}{\partial n} - \frac{2}{R} T. \tag{4.8}
$$

The equation, or boundary condition, (4.8) is called the fundamental equation of physical geodesy.

Above we already obtained equations (2.12) and (2.15), that apply equally well to the disturbing potential $T$ as to the general potential $V$:

$$
\frac{\partial T}{\partial n} = - \sum_{n=0}^{\infty} \left(\frac{n+1}{r} \left(\frac{R}{r}\right)^{n+1} T_n (\phi, \lambda) \right);
$$

$$
T = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} T_n (\phi, \lambda).
$$

By combining these we obtain

$$
\Delta g = - \sum_{n=0}^{\infty} \left[\frac{n+1}{r} - \frac{2}{R} \right] \left(\frac{R}{r}\right)^{n+1} T_n (\phi, \lambda),
$$

or on the Earth’s surface ($R = r$):

$$
\Delta g = - \sum_{n=0}^{\infty} \frac{n-1}{R} T_n (\phi, \lambda) \equiv \sum_{n=0}^{\infty} \Delta g_n (\phi, \lambda),
$$

where the quantities $\Delta g_n = \frac{n-1}{R} T_n (\phi, \lambda)$ are defined in a logical way. Remember that the functions $\Delta g_n (\phi, \lambda)$ are computable with the help of the degree constituent equation (2.19) when $\Delta g (\phi, \lambda)$ is known all over the Earth.

Observe also that the term $n = 1$ vanishes: $\Delta g_1 = 0$. We assume also $\Delta g_0 = \tau_0 / R = 0$, i.e., the true external geopotential, and thus the total mass of the Earth $GM$ (and her volume$^4$), is on average the same as the normal potential and its assumed total mass (and ellipsoidal volume). The assumption is largely justified because $GM$ can be, and has been, determined very precisely by satellites, and modern models for the normal potential are based on these determinations.

Thus we obtain the solution also of this boundary-value problem in spectral representation (which is thus valid in the whole exterior space) by using the

$^4$In fact, the atmosphere complicates this matter.
degree constituent equation (2.19):

\[
T(\phi, \lambda, r) = R \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \frac{\Delta g_n (\phi, \lambda)}{n-1}
\]

\[
= \frac{R}{4\pi} \sum_{n=2}^{\infty} \frac{2n + 1}{n - 1} \left( \frac{R}{r} \right)^{n+1} \int_{\sigma} \Delta g (\phi', \lambda') P_n (\cos \psi) \, d\sigma'.
\] (4.9)

This is precisely the boundary-value problem that is created if everywhere on the Earth, land and sea, surface gravity anomalies are given.

The integral equation corresponding to the above spectral equation (4.9) is known as the Stokes\textsuperscript{5} equation:

\[
T(\phi, \lambda, r) = \frac{R}{4\pi} \int_{\sigma} S(\psi, r) \Delta g (\phi', \lambda') \, d\sigma',
\]

where

\[
S(\psi, r) = \sum_{n=2}^{\infty} \frac{2n + 1}{n - 1} \left( \frac{R}{r} \right)^{n+1} P_n (\cos \psi).
\] (4.10)

In section 7.1 we will give a closed formula for this function (for the case \( r = R \)) and a graph.

### 4.5 The telluroid mapping and the “quasi-geoid”

If we measure the astronomical latitude and longitude \( \Phi, \Lambda \) and interpret them as geodetic (geographical) co-ordinates \( \varphi, \lambda \), and also interpret the potential difference \( W - W_0 \) as a measure for the height above the reference ellipsoid \( h \), we perform, as it were, a mapping, which adds to every point \( P \) a corresponding point \( Q \), the geodetic co-ordinates of which are the same as the natural co-ordinates of point \( P \).

This approach is called the telluroid mapping. The telluroid is the surface that follows the shapes of Earth’s topography, but is everywhere below the topography by an amount \( \zeta \) if positive, or above it by an amount \( -\zeta \) otherwise. This quantity is called a height anomaly.

The telluroid mapping is an important tool in Molodensky’s gravity field theory. It is however a pretty abstract concept. One may say that the telluroid is a model of the Earth surface, obtained by starting from the assumption that

- the true potential field of the Earth is the normal potential; and

---

\textsuperscript{5}Sir George Gabriel Stokes (1819–1903) was an Irish-born, gifted mathematician and physicist making his career in Cambridge.
the mathematical mean sea surface or geoid, the reference surface for height measurement, coincides with the reference ellipsoid.

In other words, the telluroid is a model for the Earth’s topographic surface that is obtained by taking levelled heights – more precisely, geopotential numbers obtained from levelling – as if they represented potential differences with the reference ellipsoid.

In practice, a map of values $\zeta$ is often called a “quasi-geoid” model. The quasi-geoid is usually close to the geoid, except in the mountains, where the differences can exceed a metre.

One should however remember, that the height anomaly $\zeta$ is defined on the topographic surface, a surface that may be quite rough. This means also that all variations in topographic height will be reflected also as variations in the “quasi-geoid”, in such a way, that the quasi-geoid correlates strongly with the small details in the topography. One can thus not say that the shape of the quasi-geoid only describes the Earth’s potential field. In it, variations in potential and variations in topographic height are hopelessly mixed up. This is why the quasi-geoid is an unfortunate compromise, a concession to “reference surface thinking”, which only really works within the classical geoid concept. Better stick – within Molodensky’s theory – to the concept height anomaly, which is a three-dimensional function or field

$$\zeta (X, Y, Z) = \zeta (\phi, \lambda, h).$$

### 4.6 Free-air anomalies

If we measure gravity $g$ in point $P$, the height of which over sea level is $H$ and its latitude $\Phi$, we may compute the gravity anomaly as follows:

$$\Delta g_p = g_p - \gamma (H, \Phi),$$

where $\gamma (H, \Phi)$ is normal gravity computed at height $H$ and latitude $\Phi$. This is how we define free-air anomalies.

We linearize this as follows:

$$\Delta g_p = g_p - \gamma (H, \Phi) \approx g_p - \gamma (h, \varphi) - (H - h) \frac{\partial \gamma}{\partial h} - (\Phi - \varphi) \frac{\partial \gamma}{\partial \varphi} \approx$$

$$\approx g_p - \gamma (0, \varphi) - h \frac{\partial \gamma}{\partial h} - (H - h) \frac{\partial \gamma}{\partial h} = g_p - \gamma (0, \varphi) - H \frac{\partial \gamma}{\partial h},$$
where we make the approximation, that the vertical gradient $\partial \gamma / \partial h$ of normal gravity is constant$^6$.

Thus, free-air anomalies can be calculated in a simpler way. The gravity formula of the normal field (3.7) gives for latitude $60^\circ$:

$$\gamma = 974147.516 - 0.3084494 H + \ldots \text{mGal}.$$ 

So, in linear approximation (close to the Earth surface) gravity attenuates some 0.3 mGal for every metre in height. This value is worth memorizing.

An approximate formula for calculating free-air anomalies then is

$$\Delta g_P = g_P - \gamma_0(\phi) + 0.3084 \left[\text{mGal/m}\right] H, \quad (4.11)$$

where $\gamma_0(\phi) \approx \gamma(0, \phi)$, normal gravity at sea level, is only a function of latitude. In a country like Finland, formula (4.11) is often sufficiently precise, though today also the evaluation of original formula (3.7) is easy.

Free-air anomalies are widely used. Generally, when one discusses gravity anomalies, one means just this, free-air anomalies. They describe the Earth’s external gravity field, including mountains, valleys and everything.

Questions:

1. If gravity on the Earth’s surface is $9.8 \text{m/s}^2$, at what height will gravity disappear according to the above mentioned vertical gravity gradient $-0.3 \text{mGal/m}$?

2. How physically realistic is this?

Answers:

1. At $-0.3 \text{mGal/m}$, it takes $\frac{10^3 \times 9.8}{0.3} \text{m} = 3267 \text{ km}$ to go to zero.

2. Not very. The gravity gradient itself drops quickly from the value of $-0.3 \text{mGal/m}$ going up, so this linear extrapolation is simply wrong.

---

$^6$So, for greatest precision one should consider that also the latitude $\Phi$ may not be a latitude on a geocentric reference ellipsoid, but, e.g., astronomical latitude, or latitude in some old national co-ordinate system computed on a non-geocentric ellipsoid, like in Finland KKJ, the National Grid Co-ordinate System, which was computed on the Hayford ellipsoid. The error caused by this is however of order $10^3$ smaller than the effect caused by $H - h$. 
Chapter 4. Anomalous quantities of the gravity field

4.7 Exercises

4.7.1 The spectrum of gravity anomalies

Use equation (4.6). If we assume that the mean magnitude of the spectral components $\Delta g_n$ of gravity anomalies

$$\|\Delta g_n\| \equiv \frac{1}{4\pi} \iint \Delta g_n(\phi, \lambda) \ d\sigma$$

does not depend on the chosen degree number $n$; how then does $\|T_n\|$ depend on $n$?

In other words: which degree numbers of the gravity field are relatively strongest in the disturbing potential, and which in the gravity anomalies?
5. Geophysical reductions

5.1 General

We see that integral equations, like Green’s third theorem (1.24), offer a possibility to calculate the whole exterior potential field of the Earth (as well as all quantities that may be calculated from the potential, like the acceleration of gravity, etc.) from observed values \( V \) and \( \frac{\partial V}{\partial n} \) on the boundary surface only. Green III is but one example out of many: every integral theorem is the solution of some boundary-value problem.

There are three alternatives concerning the choice of boundary surface:

1. choose the topographic surface of the Earth.
2. Choose mean sea level, more precisely, a equipotential surface close to mean sea level called the geoid.
3. Choose the reference ellipsoid.

- Alternative 1 has been developed most of all by the Molodensky school (Molodensky et al., 1962) in the Soviet Union. The advantage of the method is that we need no gravity reduction, because all significant masses are already inside the boundary surface. Its disadvantage is, that the, often complex, shape of the topography must be taken into account when the boundary-value problem is formulated and solved.
- Alternative 2 is classical geoid or geopotential determination. In this case geophysical reductions are needed to the input gravity data: some masses are outside the computation boundary and need to be moved to the inside.
A further complication of the method then is, that the geopotential or
geoïd solution obtained is not that of the original mass distribution,
but of the reduced one. This surface is called the co-geoïd. We need
a “restoration step” where this influence of the reduction step on the
geopotential and geoïd is determined and reversed\(^1\).
In the literature this method is also referred to as the Remove-Restore
method.

- Alternative 3 has been used rarely, because it has not been traditionally
  possible to do gravity measurements in a location known in the absolute
  sense, relative to the geocentre or the reference ellipsoid. Nowadays this
  would be possible using GNSS, e.g., in Antarctica and Greenland where
  there is no sea level bound height system.

### 5.2 Bouguer anomalies

Free-air anomalies depend on the topography. This is clear, because gravity
itself contains the attractive effect of topographic masses. A map of free-air
anomalies shows the same small details as seen in the topography. One way
of removing the effect of the topography is the so-called Bouguer\(^2\) reduction.

#### 5.2.1 Calculation

We calculate the effect of a homogeneous plate on gravity. Assume that the
plate is infinite in size; thickness \(d\), matter density \(\rho\), and height of point \(P\)
above the lower surface of the plate \(H\). See figure 5.1. The attraction in point
\(P\) (which is directed straight downward for symmetry reasons) is obtained
by integrating\(^3\):

\[
a \equiv |\mathbf{a}| = G \iiint \frac{\cos \beta}{\ell^2} \rho d\mathcal{V} = G\rho \int_0^d \int_0^{2\pi} \int_0^{-\pi/2} \frac{\cos \beta}{\ell^2} \cdot \frac{\ell}{\cos \beta} \cdot s \cdot d\alpha \cdot dz = 2\pi G\rho \int_0^d \int_0^{\pi/2} \sin \beta \cdot d\beta \cdot dz.
\]

\(^1\)This influence is called the “indirect effect”.

\(^2\)Pierre Bouguer (1698 – 1758) was a French professor in hydrography, who participated
in the public discussion on the figure of the Earth, and in 1735–1743 led an
expedition of the French Academy of Sciences doing a grade measurement in Peru,
South America, at the same time when De Maupertuis executed a similar grade
measurement in Lapland. In addition to geodesy, he was also active in astronomy.

\(^3\)We used here \(\cos \beta ds = \ell d\beta \Rightarrow ds = \frac{\ell}{\cos \beta} d\beta\), as is correct when transforming from
\((z, s, \alpha)\) co-ordinates to \((z, \beta, \alpha)\) co-ordinates.
Here, the integral
\[
\int_0^{\pi/2} \sin \beta \, d\beta = \left[ -\cos \beta \right]_0^{\pi/2} = 1,
\]
and the end result is
\[
a = 2\pi \rho \, d. \tag{5.1}
\]
This is the formula for the attraction of a Bouguer plate. As a side result we obtain the attraction of a circular disk of radius \(r\):
\[
\int_0^{\beta_0(z)} \sin \beta \, d\beta = \left[ -\cos \beta \right]_0^{\beta_0(z)} = 1 - \cos (\beta_0(z)),
\]
and the whole integral
\[
2\pi \rho \int_0^d \left[ 1 - \frac{H - z}{\sqrt{(H - z)^2 + r^2}} \right] \, dz.
\]
The indefinite integral is
\[
\int \frac{H - z}{\sqrt{(H - z)^2 + r^2}} \, dz = -\sqrt{(H - z)^2 + r^2}.
\]
Substituting the bounds yields
\[
\int_0^d \left[ 1 - \frac{H - z}{\sqrt{(H - z)^2 + r^2}} \right] \, dz = d + \sqrt{(H - d)^2 + r^2} - \sqrt{H^2 + r^2}.
\]
If we define
\[
\ell(z) = \sqrt{(H - z)^2 + r^2}
\]
we obtain for the whole integral
\[
2\pi \rho \left[ d + \ell(d) - \ell(0) \right].
\]
In the limit \( r \to \infty \) (and thus \( \ell (d) \to \ell (0) \)) this is identical to equation (5.1).

Bouguer anomalies are computed in order to remove the attraction of masses of the Earth’s crust above sea level or the geoid. The true topography is approximated by a Bouguer plate, see figure 5.2. There is no standard way to treat sea-covered areas; sometimes maps are drawn on which there are Bouguer anomalies over land and free-air anomalies over the sea. The calculation goes as follows:

\[
\Delta g_B = \Delta g_{FA} - 2\pi G \rho H = \Delta g_{FA} - 0.1119 H, \quad (5.2)
\]

where we assume for the density of the plate an often used value for the average density of the Earth’s crust, \( \rho = 2670 \text{ kg/m}^3 \). By substituting into this equation (4.11) we obtain

\[
\Delta g_B = g_P - \gamma_0 (\varphi) + [0.3084 - 0.1119] H = g_P - \gamma_0 (\varphi) + 0.1965 H. \quad (5.3)
\]

The quantities \( \Delta g_B \) are called (simple) Bouguer anomalies.

The difference between the attraction of a Bouguer plate and the true topography is called the terrain correction (areas I and II in the figure). We shall return to its computation later on.

### 5.2.2 Properties

Bouguer anomalies are, unlike free-air anomalies that vary on both sides of zero, especially in the mountains strongly negative. E.g., if the mean elevation of a mountain range is \( \overline{H} = 1000 \text{ m} \), the Bouguer anomalies will, as a consequence of this, contain a bias of \( 1000 \times (-0.1119 \text{ mGal}) = -112 \text{ mGal} \); about \(-100 \text{ mGal} \) for every kilometre of elevation.

The advantage of Bouguer anomalies is their smaller variation with place. For this reason they are suited especially for interpolation and prediction of
5.3 Terrain effect and terrain correction

Using the simple Bouguer correction does not remove precisely the attractive effect of the whole topography. Looking at figure 5.2, we see that we make two types of error:

- the attraction of areas I is taken along, though there is nothing there
- the attraction of areas II, where there actually is stuff, is ignored.

Both errors work in the same direction! Because areas I are below the point of evaluation, their attraction acts downward. And because areas II are above the point of evaluation, their attraction – which in the simple Bouguer reduction is not taken into account – would act upward, and the error made is in the same direction as in the previous case.

The terrain correction is always positive!

We write

\[ \Delta g'_B = \Delta g_B + TC, \]

where \( TC \) – the terrain correction, is positive. \( \Delta g'_B \) is called the terrain corrected Bouguer anomaly.
The terrain correction is computed by numerical integration. Figure 5.5 shows the prism method, and how both prisms, I and II, lead to a positive correction, because prism I is computationally added and prism II removed when applying the terrain correction. One needs a digital terrain model, DTM, which must be, especially around the evaluation point, extremely dense: according to experience, 500 m is the maximum inter-point separation in a country like Finland, in the mountains one needs even 50 m. The systematic nature of the terrain correction makes a too sparse terrain model cause, possibly serious, biases in the insufficiently corrected gravity anomalies.

For computing the terrain correction with the prism method we use the following formula (assuming a constant crustal density $\rho$ and a flat Earth) in rectangular map co-ordinates $x, y$:

$$TC(x, y) = \frac{1}{2} G \rho \int_{-D}^{+D} \int_{-D}^{+D} (H(x', y') - H(x, y))^2 \epsilon^{-3} \, dx' \, dy',$$
where
\[ \ell = \sqrt{(x' - x)^2 + (y' - y)^2 + \left\{ \frac{1}{2} (H(x', y') - H(x, y)) \right\}^2} \]
is the distance between the evaluation point \([x \ y \ H(x, y)]^T\) and the centre point of the prism \([x \ y \ \frac{1}{2} (H(x, y) + H(x', y'))]^T\). Of course this is only an approximation, but it works well enough in terrain where slopes generally do not exceed 45°. In the integral above, the limit \(D\) is typically tens or hundreds of kilometres. In the latter case, the curvature of the Earth already starts having an effect, which the formula does not consider.

The values of the terrain correction \(TC\) vary from fractions of a milligal (Southern Finland) to hundreds of milligals (high mountain ranges). In the “arm” or Finland (the North-Western, somewhat mountainous border area with Sweden and Norway), the terrain correction may be tens of milligals.

In figure 5.6 we depict the stages of computing Bouguer anomalies from
5. Geophysical reductions

Figure 5.7. A special terrain shape. The vertical rock wall at PQ is also straight on a map and extends to infinity in both directions.

gravity observations through terrain correction, Bouguer plate correction and free-air reduction.

5.3.1 Example: applying the terrain correction in a special case

Given the special terrain shape rendered in quasi 3D in figure 5.7. Here, the height differences are \( PQ' = 300 \) m and \( QQ' = 200 \) m. Rock density is the standard crustal density, \( 2670 \text{ kg/m}^3 \).

Questions:

1. Calculate the terrain correction at point P (hint: use the attraction formula for the Bouguer plate). Algebraic sign?

2. Calculate the terrain correction at point Q. Algebraic sign?

3. If in point P is given that the free-air anomaly is 50 mGal, how much is then the Bouguer anomaly in the point?

4. If in point Q is given that the Bouguer anomaly is 22 mGal, how much is then the free-air anomaly in the point?

Answers:

1. The terrain correction at point P is the change in gravity, if the terrain is filled up on the left side up to level 300 metres. This means the adding of half a Bouguer plate, thickness 100 m, below the level of P. The effect (projected onto the vertical direction) is

\[
TC = \frac{1}{2} \cdot 2\pi G \rho \cdot H = \frac{1}{2} \cdot 0.1119 \text{ mGal/m} \cdot 100 \text{ m} = 5.595 \text{ mGal}.
\]
2. The terrain correction at point $Q$ is the change in gravity, if we remove the half Bouguer plate to the right of the point, which is 100 m thick. Its vertical gravity effect is, as calculated above,

$$TC = 5.595 \text{ mGal},$$

and, because a semi-plate is removed that is above the level of point $Q$, the algebraic sign of $TC$ is again positive.

3. Free air to Bouguer:

$$\Delta g_{FA} (P) = 50.000 \text{ mGal}$$
$$TC = +5.595 \text{ mGal}$$
$$\text{Bouguer plate removal,} -33.570 \text{ mGal}$$

300 m

$$\Delta g_{B} (P) = 22.025 \text{ mGal}$$

4. Bouguer to free air:

$$\Delta g_{B} (Q) = 22.000 \text{ mGal}$$
$$\text{Bouguer plate addition,} +23.800 \text{ mGal}$$
$$TC \text{ “uncorrection”} = -5.595 \text{ mGal}$$

$$\Delta g_{FA} (Q) = 40.205 \text{ mGal}$$

5.4 Spherical Bouguer anomalies

More recently, also spherical Bouguer anomalies have been calculated, e.g., Balmino et al. (2012), Kuhn et al. (2009), Hirt and Kuhn (2014). In this calculation, the topography and bathymetry of the whole Earth is taken into account, in spherical geometry (the error caused by neglecting the Earth’s flattening is in this calculation negligible). This causes four differences with the Bouguer plate anomalies:
1. The attraction of a Bouguer shell is $4\pi G \rho H$, twice as much as the Bouguer plate attraction.

2. The bathymetry of the oceans is accounted for\(^4\) by replacing the water by standard-density crustal rock; this contribution to the anomalies is positive.

3. Also the topography and bathymetry of remote parts of the globe are taken into account realistically. As most of the Earth is covered by ocean, this causes a positive general bias, which in moderately elevated areas like Southern Finland more than cancels the negative one caused by the local topography!

4. As now also the terrain correction is computed over the whole globe – in spherical geometry – it is no longer a small number, and may be strongly negative as well as positive.

Between the planar and spherical Bouguer anomalies exists a large systematic difference, which however is very long-wavelength in nature, and even in an area the size of Australia almost a constant, $-18.6$ mGal within a few milligals. The details in the Bouguer maps look the same (Kuhn et al., 2009).

### 5.5 Helmert condensation

An often used method, proposed by Friedrich Robert Helmert\(^5\), for removing the effect of the masses external to the geoid is condensation. In this method, we shift mathematically all the continental masses vertically downward to mean sea level into a simple mass density layer $\kappa = H \rho$ – more precisely:

$$\kappa = H \rho \left( 1 + \frac{H}{R} \right)$$

(5.4)

\hspace{1cm}

on a spherical Earth, radius $R$ – where $H$ is the height of the topography above sea level and $\rho$ its mean density. The advantage of Helmert condensation over Bouguer reduction is, that no mass is being removed. The Bouguer reduction is the computational removal of topographic masses on a large

\(^4\)One can do so, and often does, also in connection with the Bouguer plate correction.

\(^5\)Friedrich Robert Helmert (1843 – 1917) was an eminent German geodesist known for his work on mathematical and statistical geodesy.
scale. Therefore, unlike with Bouguer reduction, in Helmert condensation gravity anomalies will not change systematically.

In appendix D we derive series expansions in spherical geometry, which describe both the external and the internal potential as functions of the “degree constituents” of the various powers of the topography $H(\phi, \lambda)$. The extensively presented derivation in the appendix is much used in gravity field theory to model the gravity effect of the topography. In this theory, issues of convergence are difficult, though we gloss over those here.

As the subject involves rather a lot of math and is somewhat specialized, we have chosen to place it in an appendix.

5.6 Isostasy

5.6.1 Classical hypotheses

Already in the 18th and 19th centuries, e.g., thanks to Bouguer’s work in South America, as well as that of British geodesists in the Indian Himalayas, it was understood that mountain ranges weren’t just piles of rock on top of the Earth crust; the gravity field surrounding the mountains, specifically the plumbline deflections, could only be explained by assuming that under every mountain range there was also a “root” made from lighter rock species. The origin of this root was speculated to be the almost hydrostatic behaviour
of the Earth crust over geological time scales. This assumption of hydrostatic equilibrium was called the *hypothesis of isostasy*; also *isostatic compensation*.

Back then, unlike now, it was not yet possible to get a precise or even correct picture using physical methods (seismology) of how these mountain roots were really shaped. That’s why simplified working hypotheses were formulated.

One older isostatic hypothesis is the Pratt-Hayford hypothesis. This was proposed by J.H. Pratt\(^6\) in the middle of the 19th century (Pratt, 1855, 1858, 1864), and J.F. Hayford\(^7\) developed the mathematical tools needed for computation. According to this hypothesis, the density of the “root” under a mountain would vary with the height of the mountain, so that under the highest mountains would be the lightest material, and the boundary between this light root material and the denser Earth mantle material would be at a fixed depth. This model, which nowadays finds little acceptance anymore, is illustrated in figure 5.11.

---

6John Henry Pratt (1809–1871) was a British clergyman and mathematician who worked as the archdeacon of Kolkata, India. [http://tigger.uic.edu/~rdemar/geol107/pratt.htm](http://tigger.uic.edu/~rdemar/geol107/pratt.htm).

7John Fillmore Hayford (1868–1925) was a United States geodesist who studied isostasy and the figure of the Earth.
5.6. Isostasy

Another classical isostatic hypothesis is due to G.B. Airy. Because W.A. Heiskanen used it extensively and developed its mathematical form, it is called the Airy-Heiskanen model. In this model it is assumed that the mass density of the “root” is fixed, and that the isostatic compensation is realized by varying the depth to which the root extends into the Earth mantle. In our

8 George Biddell Airy (1801–1892) was an English mathematician and astronomer, “Astronomer Royal” 1835–1881.

9 Weikko Aleksanteri Heiskanen (1895–1971), “the great Heiskanen” (http://en.wikipedia.org/wiki/Beyond_Sleep) was an eminent Finnish geodesist who also worked in Ohio, U.S.A. He is known for this work on isostasy and the global geoid. See Kakkuri (2008).
current understanding this corresponds better to what is really happening inside the Earth. This hypothesis is illustrated in figure 5.12.

5.6.2 Computation formulas

Airy’s isostasy hypothesis assumes that in every place the total mass of a column of matter is the same. So, let the density of the Earth crust be $\rho_c$, the density of the mantle $\rho_m$, and the density of sea water $\rho_w$; sea depth $d$, crustal thickness $t$, and topographic height $H$; we have

$$t\rho_c + d\rho_w - (t + d)\rho_m = c \Rightarrow t = -\frac{d (\rho_m - \rho_w) + c}{\rho_m - \rho_c}$$
on the sea, and

$$t\rho_c - (t - H)\rho_m = c \Rightarrow t = \frac{H\rho_m - c}{\rho_m - \rho_c}$$
on land; where $c$ is a suitable constant\(^\text{10}\). Here we have conveniently forgotten about the curvature of the Earth, i.e., we use the “flat Earth model”.

Under land, the depth of a mountain root is

$$r = t - H = \frac{H\rho_m - c}{\rho_m - \rho_c} - \frac{H\rho_m - h\rho_c}{\rho_m - \rho_c} = \frac{H\rho_c - c}{\rho_m - \rho_c}.$$

Similarly under the sea:

$$r = t + d = -\frac{d (\rho_m - \rho_w) + c}{\rho_m - \rho_c} + \frac{d\rho_m - d\rho_c}{\rho_m - \rho_c} = -\frac{d (\rho_c - \rho_w) + c}{\rho_m - \rho_c}.$$
Note that the constant \( c \) is arbitrary and expresses the fact that the level from which one computes the depth of the root – or, less precisely, the “average thickness of the crust” – can be chosen arbitrarily.

Another approach: instead of \( c \), use the “zero topography compensation level” \( t_0 \), to be computed from the above equation by setting \( H = d = 0 \):

\[ t_0 (\rho_c - \rho_m) = c. \]

This yields under the land:

\[ r = \frac{H \rho_c - t_0 (\rho_c - \rho_m)}{\rho_m - \rho_c} = t_0 + H \frac{\rho_c}{\rho_m - \rho_c}, \tag{5.5} \]

and under the sea:

\[ r = -d \frac{(\rho_c - \rho_w) + t_0 (\rho_c - \rho_m)}{\rho_m - \rho_c} = t_0 - d \frac{\rho_c - \rho_w}{\rho_m - \rho_c}, \tag{5.6} \]

somewhat simpler equations that are also more intuitive:

\[ H \rho_c + (-r) (\rho_m - \rho_c) = -t_0; \]

\[ (-d) (\rho_c - \rho_w) + (-r) (\rho_m - \rho_c) = -t_0. \]

In other words,

\[ \sum_{\text{interfaces}} (\text{deviation} \times \text{density contrast}) = \text{const}. \]

However, the effect of the different isostatic hypotheses on gravity is pretty much the same; the hypotheses can not be distinguished based on gravity measurements only. The effect of the hypothesis on the geoid is stronger.

\[ ^{10} \text{Its dimension is “pressure”: according to Archimedes’ law, the pressure of the crustal (plus sea-water) column minus the pressure of the column of displaced mantle material.} \]
5.6.3 Example: Norway

The Southern Norwegian *Hardangervidden* plateau is a highland at, on average, 1100 m above sea level. It is the largest peneplain in Europe, a national park and a popular tourist attraction, being traversed by *Bergensbanen*, the highest regular railway in Northern Europe.

The *Norwegian Sea* is the part of the Atlantic Ocean adjoining Norway, and does not belong to the continental shelf. It is on average 2 km deep.

Questions:

1. What is the depth of the root under the Hardanger plateau, relative to the compensation depth $t_0$?
2. What is the negative depth of the anti-root under the Norwegian Sea, relative to the same compensation depth?
3. What is the relative depth of the root of the Hardanger plateau, compared to the nearby Norwegian Sea?

Answers:

1. We use the equation (5.5), finding

\[
 r - t_0 = H \frac{\rho_c}{\rho_m - \rho_c} = 1100 \text{ m} \times \frac{2670 \text{ kg/m}^3}{3370 - 2670 \text{ kg/m}^3} = 4196 \text{ m}. 
\]

Here we have used standard densities for crustal and mantle rock, respectively.

2. We use the equation (5.6), finding

\[
 r - t_0 = -d \frac{\rho_c - \rho_w}{\rho_m - \rho_c} = -2000 \text{ m} \times \frac{2670 - 1027 \text{ kg/m}^3}{3370 - 2670 \text{ kg/m}^3} = -4694 \text{ m},
\]

using the standard density value for sea water.

3. The depth contrast between root and anti-root is $4196 - (-4694) = 8890 \text{ m}$ (for perspective, Mount Everest is 8848 m above sea level).

5.6.4 The current understanding of isostasy

Nowadays we have a much better understanding of the internal situation in the Earth. However, isostasy continues to be a valid concept. A more realistic picture of the internal structure of the Earth is given in figure 5.14.
An important subject for current research is the effect on vertical motion of the Earth crust of the growing and melting of the ice masses of the Earth, like the continental ice sheets. To this belongs both the direct effect of the varying ice masses, and the effect of the changes caused in the water masses of the ocean. So-called paleo-research concentrates on the changes over the glacial cycle, while modern retreats of glaciers, e.g., in Alaska and on Spitzbergen, cause their own, observable local uplift of the Earth’s crust. More in chapter 11.

5.6.5 Example: Fennoscandian land uplift

During the last glacial maximum, some 20,000 years ago, Fennoscandia was covered by a continental ice sheet of thickness up to 3 km.

Questions:

1. How much was the Earth’s surface depressed by this load, assuming isostatic equilibrium?

2. Currently the land is rising in central Fennoscandia, where the ice thickness was maximal, at a rate of $10 \text{ mm/a}$. How long would it take at this rate for the depression to disappear?

Answers:
1. We assume for the ice density a value of $920\,\text{kg/m}^3$. Then, with an upper mantle density of $3370\,\text{kg/m}^3$ – note that it is Earth’s mantle material that is being displaced by the ice, the Earth’s crust just transmits the load! See figure 11.1a – we find for the depression:

$$\Delta H = 3000\,\text{m} \times \frac{920\,\text{kg/m}^3}{3370\,\text{kg/m}^3} = 819\,\text{m}.$$ 

2. At the rate of $10\,\text{mm/yr}$ it will take $819\,\text{m}/0.01\,\text{m/yr} = 81,900$ years total. Part of this uplift has already taken place since the last deglaciation. In reality, of course, the rate has decreased substantially, and will continue to decrease, over time.

### 5.7 Isostatic reductions

The computational removal of both the topography and its isostatic compensation from the measured quantities of the gravity field is called isostatic reduction. It serves two purposes.

1. By removing as many as possible “superficial” effects from the gravity field, we are left with a field where only the effect of the Earth’s deep layers remains.

2. These “superficial” effects are also generally very local: in spectral language, very short wavelength. By removing those, we are left with a residual field that is much smoother, and that can be interpolated or predicted better. This is important especially in areas where there is a paucity of real measurement data, like the oceans, deserts, polar areas etc.

For example, isostatic anomalies, i.e., free-air anomalies to which isostatic reduction has been applied, are very smooth (like also Bouguer anomalies), and their predictive properties are good. However, unlike Bouguer anomalies, isostatic anomalies are on average zero. They lack the large bias that makes Bouguer anomalies strongly negative especially in mountainous areas (see section 5.2). This of course is because isostatic reduction is only the shifting of masses from one place to another – from mountains to roots beneath the same mountains, the mass deficit of which is pretty precisely the same as the mass of the mountains themselves sticking out above sea level – rather than removal of masses, which is what Bouguer reduction does.
The reduction methods used in isostatic calculations are the same as in other reductions, and we will discuss them later: numerical integration in the space domain – grid integration, “spherical cap”, least-squares collocation (LSC), finite elements, etc. – or in the spectral domain (FFT, “fast collocation”, etc.).

The question of the hypothesis assumed to apply is a more interesting one. Traditionally, the Pratt or Airy hypotheses have been used, developed into methodologies by Hayford or Heiskanen or Vening Meinesz. A newer approach has been to use real measurement data from seismic tomography in order to model the interior structure of the Earth. With real measurement data, if reliable, one should get better results.

5.8 The “isostatic geoid”

Let us look at how the “isostatic geoid”, more precisely the co-geoid of isostatic reduction, is computed. Isostatic reduction is one possible method for computationally removing the masses outside the geoid, in order to formulate a
boundary-value problem on the geoid.

We can show (Heiskanen and Moritz, 1967 p. 142), that the isostatic co-geoid is under the continents as much as several metres below the geoid, i.e., the indirect effect (“Restore step”) is of this order. Under the oceans, similarly the isostatic co-geoid is somewhat above the geoid.

As one of the requirements for geoid determination methods is a small indirect effect, it follows that isostatic methods are not (contrary to what is said on Heiskanen and Moritz page 152) the best possible if the intent is to calculate the geoid or quasi-geoid representing the exterior potential\textsuperscript{11}. However, isostatic methods are very suitable for elucidating the interior structure of the Earth, because both the topography and the “impression” it makes on the Earth’s mantle, the isostatic compensation, are computationally removed. Research has shown that the great topographic features of the Earth are some 85–90\% isostatically compensated. This is valuable information if no other knowledge is available.

This is the second reason why the isostatic geoid is of interest: the gravity field of an Earth from which the effect of mountains has been removed completely – roots and all – can uncover physical unbalances existing in deeper layers, and processes causing these. Such processes are especially convection currents in the Earth’s mantle as well as the possible effect of the liquid core of the Earth on these currents. Interesting correlations have been found between mantle convection patterns, the global map of the geoid, and the electric current patterns in the core causing the Earth’s magnetic field (Wen and Anderson, 1997; Prutkin, 2008; Kogan et al., 1985).

Isostatic reduction consists of two parts:

1. computational removal of the topography

2. computational removal of the isostatic compensation of the topography.

It is possible to calculate both these parts exactly using prism integration, see section 5.3. Here however we shall gain understanding by a qualitative approach. We approximate both parts with a single mass density layer, density, e.g., $\kappa = \rho H$ for the topography. We place the first layer at level $H = 0$,

\textsuperscript{11}Of course Bouguer reduction is even worse! The indirect effect can be hundreds of metres.
and the second, density $\kappa$, at compensation depth $H = -D$. The situation is depicted in figure 5.16.

In the following we use the “generating function” equation (7.4),

$$\frac{1}{\ell} = \frac{1}{R} \sum_{n=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} P_n \left( \cos \psi \right),$$

together with the single mass density layer equation (1.15):

$$V = G \int_{\text{Surface}} \kappa \frac{dS}{\ell} = GR^{2} \int_{\text{Surface}} \kappa \frac{d\sigma}{\ell}.$$

We obtain for the potential field when the density layer is placed at sea level ($H = 0 \Rightarrow r = R$):

$$T_{\text{top}} = GR \int_{\sigma} \kappa \sum_{n=0}^{\infty} P_n \left( \cos \psi \right) d\sigma$$

and with the density layer at compensation depth (source level $R-D$, evaluation level $R$):

$$T_{\text{comp}} = \frac{GR^{2}}{R-D} \int_{\sigma} (-\kappa) \sum_{n=0}^{\infty} \left( \frac{R-D}{R} \right)^{n+1} P_n \left( \cos \psi \right) d\sigma =$$

$$= -GR \int_{\sigma} \kappa \sum_{n=0}^{\infty} \left( \frac{R-D}{R} \right)^{n} P_n \left( \cos \psi \right) d\sigma,$$

from which the combined effect ($n = 0$ drops out)

$$\delta T_{\text{iso}} = -(T_{\text{top}} + T_{\text{comp}}) = -GR \int_{\sigma} \kappa \sum_{n=1}^{\infty} \left[ 1 - \left( \frac{R-D}{R} \right)^{n} \right] P_n \left( \cos \psi \right) d\sigma.$$

Here, the mass density per surface area $\kappa$ is

$$\kappa = \begin{cases} \rho_c H & \text{if } H \geq 0, \\ (\rho_c - \rho_w) H & \text{if } H < 0, \end{cases}$$

i.e., we replaced ocean depths by equivalent “dry” depths\(^{12}\). Now we use again the degree constituent equation Heiskanen and Moritz (1967) equation 1-71, or our equation (2.19) in the following form:

$$\kappa_n (\phi, \lambda) = \frac{2\pi n + 1}{4\pi} \int_{\sigma} \kappa (\phi', \lambda') P_n \left( \cos \psi \right) d\sigma,$$

or

$$-\frac{4\pi GR}{2\pi + 1} \left[ 1 - \left( \frac{R-D}{R} \right)^{n} \right] \kappa_n (\phi, \lambda) =$$

$$= -GR \int_{\sigma} \kappa (\phi', \lambda') \left[ 1 - \left( \frac{R-D}{R} \right)^{n} \right] P_n \left( \cos \psi \right) d\sigma.$$

\(^{12}\)This works on dry land and on the ocean. Lakes, glaciers and areas like the Dead Sea are more complicated.
Summation yields the expression (5.7) above:

$$\sum_{n=1}^{\infty} \frac{4\pi G R}{2n+1} \left[ 1 - \left( \frac{R - D}{R} \right)^n \right] \kappa_n (\phi, \lambda) =$$

$$= -G R \int_\sigma \kappa (\phi', \lambda') \sum_{n=1}^{\infty} \left[ 1 - \left( \frac{R - D}{R} \right)^n \right] P_n (\cos \psi) \, d\sigma,$$

so that

$$\delta T_{iso} = -\sum_{n=1}^{\infty} \frac{2}{2n+1} R \left[ 1 - \left( \frac{R - D}{R} \right)^n \right] 2\pi G \kappa_n =$$

$$= -\sum_{n=1}^{\infty} \frac{2}{2n+1} R \left[ 1 - \left( \frac{R - D}{R} \right)^n \right] [A_B]_n.$$

Here we have used the notation $A_B = 2\pi G \kappa$, the Bouguer attraction of mass density layer $\kappa$.

Let us first look at the contribution from $0 < n \leq N = \frac{R}{D}$. Then the following approximation holds, as $\left( \frac{R - D}{R} \right)^n \approx 1 - \frac{nD}{R}$:

$$\delta T_{iso} \approx -\sum_{n=1}^{N} \frac{2nD}{2n+1} [A_B]_n \approx -\sum_{n=1}^{N} D [A_B]_n \approx -D A_B,$$

and

$$\delta N_{iso} = \frac{\delta T_{iso}}{\gamma} \approx -\frac{D A_B}{\gamma}.$$

This is the indirect effect of isostatic reduction.

---

The contribution from degree numbers $n > \frac{R}{D}$ is

$$\delta T_{iso} \approx -\sum_{n=N+1}^{\infty} \frac{2R}{2n+1} [A_B]_n,$$

where the terms are small and rapidly falling to zero. In this region also the mass density layer approximation for the topography breaks down.
5.8. The “isostatic geoid”

Let’s substitute realistic values. Let the Mohorovičić\textsuperscript{14} discontinuity’s depth be on average \( \sim 20 \text{ km}\).\textsuperscript{15}

On land: \( H \approx 0.8 \text{ km} \) (The Earth’s mean topographic height): \( \Rightarrow \delta N_{\text{iso}} \approx -1.8 \text{ m} \).

On the ocean: \( H \approx -3.7 \text{ km} \) on average (multiply by \( \frac{1.67}{2.67} \) because of the water): \( \Rightarrow \delta N_{\text{iso}} \approx +5.1 \text{ m} \).

In other words, this effect can be big!

Note that equation (5.8) is linear in the height \( H \). This means that, under the continents, the isostatic co-geoid will run on the order of a couple of metres below the classical geoid, when on the oceans again it must be a few metres above the geoid (mean sea level). We may also conclude that in the isostatic reduction’s effect on the geoid – at least for longer wavelengths \( \frac{2\pi R}{n} \), longer than the compensation depth \( D \) – all wavelengths are represented in the spectrum in approximately in the same proportions as in the topography itself, and the effect is in fact proportional to the topography.

\textsuperscript{14}Andrija Mohorovičić (1857–1936) was a Croatian meteorologist and a pioneer of modern seismology.

\textsuperscript{15}Under the continents 35 km, under the oceans 7 km below the sea floor, according to Encyclopædia Britannica. Using these values, we find \( \delta N_{\text{iso}} = -3.2 \text{ m} \) on land, +2.8 m on the ocean.
6. Vertical reference systems

6.1 Levelling, orthometric heights and the geoid

Heights have traditionally been determined by levelling. Levelling is a technique for determining height differences using a level (levelling instrument) and two rods or staffs. The level comprises a telescope and a spirit level, and its optical axis is aligned with the local horizon. Levelling staffs are placed on two measurement points, and through the telescope, measurement values are read off them. The difference between the two values gives the height difference between the two points in metres.

The distance between level and staffs is 40–70 m; longer distances would cause too large errors due to the effect of atmospheric refraction. Longer distances are measured by repeat measurements using several instrument stations and intermediate points.

Figure 6.1. The principle of levelling.
The height differences $\Delta H$ thus obtained are not, however, directly useable. The “height difference” obtained by directly summing height differences $\Delta H$ between two points $P$ and $Q$, depends namely also on the path chosen. Also the sum of height differences $\sum \Delta H$ around a closed path is generally not zero.

*Geometric height is not a conservative field.*

This is why, in precise levelling, the height differences are always converted to potential differences: $\Delta W = -\Delta H \cdot g$, where $g$ is the local gravity, which is either measured or (e.g., in Finland) interpolated from an existing gravity survey data base. The sum of potential differences around a closed loop is always zero: $\sum \Delta W = 0$.

For the potential of an arbitrary terrain point $P$ we find:

$$ W_P = W_0 - \sum (\Delta H \cdot g), $$

where the summation is done directly from sea level (potential $W_0$) up to point $P$. The quantity

$$ C_P = -(W_P - W_0) = \sum_{\text{Sea level}}^{P} (\Delta H \cdot g) $$

(positive above sea level) is called the *geopotential number* of point $P$.

$W_0$ is the potential of the national height reference level. In Finland, the reference level of the old N60 system is in principle mean sea level in Helsinki harbour at the beginning of 1960, which is why the system is called N60. However the precise realization is a special pillar in the garden of Helsinki astronomical observatory in Kaivopuisto\(^1\). The new Finnish height system is called N2000, and the realization of its reference level is a pillar at the Metsähovi research station (in practice, N2000 heights are at the decimetric precision level heights over the Amsterdam NAP datum). Other countries have their own, similar height reference or datum points: Russia has Kronstadt, Western Europe the widely used Amsterdam datum NAP, Southern Europe has Trieste, North America the North American Vertical Datum 1988 (NAVD88), datum point Pointe-au-Père in Rimouski, Quebec, Canada, etc.

\(^1\)However, the value engraved in the pillar is the reference height of the still older system NN, not of N60. The correct reference value for N60 for this pillar, 30.51376 m, is given in the publication Kääriäinen (1966).
6.2 Orthometric heights

For creating a vertical reference, it would be simplest to use the original geopotential differences from sea level, i.e., the geopotential numbers defined above, \( C = - (W - W_0) \), directly as height values. However, this is psychologically and practically difficult: people want their heights to be in metres. Geopotential numbers have their clear advantages: they represent the amount of energy that is needed (for a unit test mass) to move from one point to another. Fluids (sea water, but also air, or, on geological time scales, even bedrock!) flows always downward and seeks the state of minimum energy. 

In Finland, as in many other countries, orthometric heights have been long in use. They are physically defined heights above “mean sea level” or the geoid. See figure 6.3.

The classical geoid is defined as

“The equipotential surface of the Earth’s gravity field which fits on average best to mean sea level.”

The orthometric height of point \( P \) is defined as the height obtained by measuring along the plumpline the distance of \( P \) from the geoid.

This is a very physical definition, however not a very operational one, because we (generally) do not get to measure along a plumpline inside the Earth, and the geoid isn’t visible there. This is why orthometric heights are
Figure 6.3. Levelled heights and geopotential numbers. The height obtained by summing levelled height differences, $\sum_{i=1}^{3} \Delta H_i$, is not the “correct” height above the geoid, i.e., $\sum_{i=1}^{3} \Delta H'_i$ computed along the plumbline.

Note how the level surfaces of the geopotential, or equipotential surfaces, are not parallel: because of this, a journey along the Earth surface may well go “upward”, i.e., to increasing heights above the geoid, although the geopotential number decreases. Thus, water may flow “upward”.

The gravity vector $\mathbf{g}$ is everywhere perpendicular to the equipotential surfaces, and its length is inversely proportional to the distance separating the surfaces.

calculated from geopotential numbers: if the geopotential number of point $P$ is $C_P$, we calculate the orthometric height with the formula

$$H = \frac{C_P}{\bar{g}},$$

where $\bar{g}$, the average gravity along the plumbline, is

$$\bar{g} = \frac{1}{H} \int_0^H g(z) \, dz,$$

and $z$ is the measured distance along the plumbline. Because the formula for $\bar{g}$ already itself contains $H$, we obtain the solution iteratively, using initially a crude estimate for $H$; the iteration converges fast.

We shall see that determining very precise orthometric heights is challenging, especially in the mountains.

### 6.3 Normal heights

In Finland, currently, with the height system N2000, normal heights are used. They are, like orthometric heights, heights above mean sea level. The mathematical representation of mean sea level in this case is the quasi-geoid. In sea areas, the quasi-geoid is identical to the geoid; over land, it differs a
little from the geoid, and in mountainous areas the difference may even be substantial.

6.3.1 Molodensky’s theory

M.S. Molodensky\(^2\) developed a theory in which the height of a point from “mean sea level” would be defined by the following equation:

\[
H^* = \frac{C}{\gamma_0H},
\]

where \(\gamma_0\) is the average normal gravity computed between the zero level (reference ellipsoid) and \(H^*\) along the ellipsoidal normal. I.e., the same way of computing as in the case of orthometric heights, but using the normal gravity field instead of the true gravity field.

Heights “above sea level” are for practical reasons given in metres. For large, continental networks we want to give heights above a computational reference ellipsoid in metres, and thus also heights above “sea level” have to be in metres.

Molodensky proposed also that instead of the geoid, height anomalies would be used, the definition of which is

\[
\zeta = \frac{T}{T_Hh},
\]

where now \(T_H\) is the average normal gravity at terrain level, more precisely: between the heights \(H^*\) (but reckoned from the ellipsoid) and \(h\). \(T\) is the disturbing potential.

\(^2\)Mikhail Sergeevich Molodensky (1909–1991) was an illustrious Soviet physical geodesist.
Based on these assumptions, he showed that

\[ H^* + \zeta = h, \]

where \( h \) is the height of a point above the reference ellipsoid. This equation is very similar to the corresponding one for orthometric heights and geoid heights

\[ H + N = h. \]

Also otherwise \( \zeta \), the height anomaly, also called “quasi-geoid height”, is very close to \( N \), and correspondingly \( H^* \) close to \( H \).

### 6.3.2 Molodensky’s proof

The realization of the Molodensky school was, that, because normal gravity along the plumbline is very close to a linear function of place, one could define a height type that can be computed directly from geopotential numbers, and that also would be compatible with similarly defined, so-called height anomalies, and with geometric heights \( h \) reckoned from the reference ellipsoid.

The geometric height \( h \) from the reference ellipsoid may be connected to the potential \( U \) of the normal gravity field indirectly, though the following integral equation:

\[ U = U_0 - \int_0^h \gamma(z)dz. \]

Here \( U \) is the normal potential and \( \gamma \) normal gravity. One equipotential surface of \( U \), \( U = U_0 \), is also the reference ellipsoid. The variable \( z \) is the distance from the ellipsoid along the local normal.
By defining
\[ \gamma_0 h = h \int_0^h \gamma(z) dz \]
we obtain
\[ h = - \frac{U - U_0}{\gamma_0 h}. \]
By using \( W = U + T \) and dividing by \( \gamma_0 h \) we obtain:
\[ \frac{W - W_0}{\gamma_0 h} = \frac{T}{\gamma_0 h} - h \]
assuming \( W_0 = U_0 \), the normal potential on the reference ellipsoid.

Next, one could define
\[ H^+ \equiv - \frac{W - W_0}{\gamma_0 h} \]
as a new height type, and
\[ N^+ = h - H^+ = \frac{T}{\gamma_0 h} \]
as the corresponding new geoid height type. It has however the esthetic flaw, that we divide here by the average normal gravity computed between the levels 0 and \( h \). This quantity is not operational without a means of determining the ellipsoidal height \( h \). This suggests the following improvement. Define the more operational quantity
\[ \gamma_0 H = \frac{1}{H^+} \int_0^{H^+} \gamma(z) dz \approx \frac{1}{H^+} N^+ \frac{d\gamma}{dr} \approx \gamma_0 \left( 1 - \frac{N^+}{R} \right) \]  
(6.1)
(R is the Earth radius, and \( \frac{d\gamma}{dr} \approx \frac{2\gamma}{R} \)), and
\[ \gamma_{Hh} \approx \gamma \left( H^+ + \frac{1}{2} N^+ \right) \approx \gamma_0 h + \frac{1}{2} H^+ \frac{d\gamma}{dr} \approx \gamma_0 \left( 1 + \frac{H^+}{R} \right), \]  
(6.2)
(all the time exploiting the circumstance, that \( \gamma(z) \) is a nearly linear function), from which follows
\[ H^* \equiv - \frac{W - W_0}{\gamma_{Hh}} \approx H^+ \left( 1 + \frac{N^+}{R} \right) = H^+ + \frac{N^+ H^+}{R}, \]
\[ \zeta \equiv - \frac{T}{\gamma_{Hh}} \approx N^+ \left( 1 - \frac{H^+}{R} \right) = N^+ - \frac{N^+ H^+}{R}, \]
and, as the correction terms \( \frac{N^+ H^+}{R} \) sum to zero:
\[ H^* + \zeta = H^+ + N^+ = h. \]  
(6.3)
The quantity \( \gamma_{Hh} \), and thus also normal height \( H^* \), can be, unlike \( \gamma_0 h \), computed using only information obtained by (spirit or trigonometric) levelling, without knowing the height \( h \) above the ellipsoid, which would require again knowledge of the local geoid.
This was Molodensky’s realization (Molodensky et al., 1962) already in 1945, long before GPS, or a global, geocentric reference ellipsoid, existed. Back then, continental triangulation networks were computed on their own, regional reference ellipsoids.

The size of the correction term $N^+ H^+ / R$ is, taking for the heights of the global geoid up to 110 m, 17 mm for each kilometre of terrain height. The errors remaining after applying this term are microscopically small, because normal gravity is (unlike true gravity) extremely linear along the plumbline – as equations (6.1) and (6.2) already assumed.

Figure 6.6 attempts to visualize the derivation.

6.3.3 Normal height and height anomaly

Normal height:

$$H^* = \frac{C}{\bar{\gamma}} = -\frac{W - W_0}{\bar{\gamma}},$$

(6.4)

where (recursive definition!)

$$\bar{\gamma} = \bar{\gamma}_{0H} = \frac{1}{H^*} \int_0^{H^*} \gamma(z) \, dz.$$
Figure 6.7. Geoid, quasigeoid, telluroid and topography. Note the correlation between quasigeoid and topography.

where

\[ \bar{\gamma}_{THh} = \frac{1}{\zeta} \int_{H^*}^{h} \gamma(z) \, dz. \]

The height anomaly \( \zeta \), which otherwise is a quantity similar to the geoid height \( N \), however is located on the level of the topography, not sea level. The surface formed by points which are a distance \( H^* \) above the reference ellipsoid (and thus a distance \( \zeta \) below the topography), is called the telluroid. It is a map of sorts of the topographic surface: the set of points \( Q \), whose normal potential \( U_Q \) is the same as the true potential \( W_P \) of the true topography’s corresponding point \( P \). See figure 4.4.

Often, as a concession to old habits, we construct a surface that is at a distance \( \zeta \) above the reference ellipsoid. This surface is called the quasi-geoid. It lacks any physical meaning; it is not an equipotential surface, although out at sea it coincides with the geoid. Its short-wave features, unlike those of the geoid, correlate with the short-wavelength features of the topography.

**Height above the ellipsoid** (assumed \( U_0 = W_0 \)):

\[ h = \frac{U - U_0}{\bar{\gamma}_{0h}}, \]

where

\[ \bar{\gamma}_{0h} = \frac{1}{h} \int_0^h \gamma(z) \, dz. \]

The relationship between the three quantities is

\[ h = H^* + \zeta. \]

In all three cases, the quantity is defined by dividing the potential difference by some sort of “average normal gravity”, suitably computed along a segment of the local plumbline. In the case of the height anomaly \( \zeta \), a piece
Chapter 6. Vertical reference systems

of plumbline is used high up, close to the topographic surface, between level $H^*$ (telluroid) and level $h$ (topography).

6.4 Difference between geoid height and height anomaly

Normal heights are very operational. They are always used together with so-called “quasi-geoid” heights (more correctly: height anomalies) $\zeta$. Orthometric heights (more precisely: Helmert heights) on the other hand are always used together with geoid heights $N$. For computing both, $H$ and $N$, one needs the topographic mass density $\rho$, which generally is assumed constant ($2670 \text{ kg/m}^3$), and the local gradient of gravity, for which generally the standard free-air gradient ($-0.3086 \text{ mGal/m}$) is assumed.

The difference between geoid height and height anomaly is calculated as follows.

1. First, calculate the separation between the quasi-geoid and the “free-air geoid”. The free-air geoid is an equipotential surface of the harmonically downward continued, external potential field. If $T$ is the disturbing potential of the external, harmonically continued field, then its difference between topography level and sea level is:

$$T_H - T_0 = \int_0^H \frac{dT}{dh} dh \approx -\Delta g_{FA} H, \quad (6.5)$$

and by using definitions $\zeta = \frac{T_H}{\gamma}$ (height anomaly or quasi-geoid height) and $N_{FA} = \frac{T_0}{\gamma}$ (“free-air geoid height”, FA = Free Air) we obtain

$$\zeta - N_{FA} = -\frac{\Delta g_{FA} H}{\gamma}. \quad (6.6)$$

2. Thus we have obtained the difference between height anomalies and heights of the “free-air geoid”; what is left is determining the separation between “free-air geoid” and geoid.

Let us approximate the topography by a Bouguer plate. Then

- in the case of the “free-air geoid” $N_{FA}$ the thickness of the plate is the height $H$ of point $P$. This is because the free-air geoid is based on the harmonically downward continued external field, meaning that also the Bouguer-plate attraction acting at $P$ must be continued downward, i.e., taken fully into account.
Because the surface mass density of the plate is $H\rho$, its assumed attraction is everywhere on point $P$’s plumbline:

$$2\pi GH\rho.$$ 

Now in the case of the geoid, we have to be physically realistic: in an arbitrary point $P'$ on the plumbline of point $P$, the Bouguer plate is partly below the point, and partly above the point. The attraction is then only

$$2\pi GH'\rho - 2\pi G (H - H') \rho = 2\pi G (2H' - H) \rho,$$

where $H'$ is now the height of point $P'$.

By integrating the difference, like we did for equation (6.5), we obtain

$$T - T_{FA} = 2\pi G\rho \int_0^H [(2H' - H) - H] \, dH' =$$

$$= 2\pi G\rho \left[ (H')^2 - 2HH' \right]_{H'=0}^H = -2\pi G\rho H^2 \approx -A_B H,$$

where $A_B$ is the attraction of a Bouguer plate of thickness $H$. We obtain again by dividing by the average normal gravity:

$$N - N_{FA} = -\frac{A_B H}{\gamma}.$$

By subtracting this latest result from equation (6.6) we find:

$$\zeta - N = \frac{(-\Delta_{FA} + A_B) H}{\pi} = -\frac{\Delta_B H}{\pi}.$$ (6.7)

See also Heiskanen and Moritz (1967, pp. 8–13). As in the mountains the Bouguer anomaly is strongly negative, it follows that the “quasi-geoid” is there always above the geoid: approximately, using equation (5.2):

$$\zeta - N \approx \frac{0.1119 \text{ [mGal/m]}}{9.8 \text{ [m/s^2]}} H^2 \approx 10^{-7} \text{ [m^{-1}]} H^2.$$

Or, if $H$ is in units of [km] and $\zeta - N$ in units of [m]:

$$\zeta - N \text{ [m]} \approx 0.1 H^2 \text{ [km]}.$$

### 6.5 Difference between normal heights and orthometric heights

The geoid is the level from which orthometric heights are measured. Therefore we may write

$$h = H + N,$$
where $h$ is height above the reference ellipsoid and $H$ orthometric height.

We may also bring back to memory equation (6.3):

$$h = H^* + \zeta,$$

where $\zeta$ is height anomaly and $H^*$ normal height.

We obtain simply:

$$H - H^* = \zeta - N = -\frac{\Delta gB H}{T},$$ (6.8)

using equation (6.7).

### 6.6 Calculating orthometric heights precisely

Orthometric heights are a more traditional way of expressing height “above sea level”. Orthometric heights are heights above a real geoid, i.e., an equipotential surface inside the Earth, and in the mean located at the same level as mean sea level.

We may write

$$W = W_0 - \int_0^H g(z) \, dz,$$

where $g$ is the true gravity inside the topographic masses. From this we obtain

$$H = \frac{C}{\bar{g}} = -\frac{(W - W_0)}{\bar{g}},$$

where mean gravity along the plumbline is

$$\bar{g} = \frac{1}{H} \int_0^H g(z) \, dz.$$

The method is recursive: $H$ appears both on the left and on the right side. This is not critical: both $H$ and $\bar{g}$ are obtained iteratively. Convergence is fast.

In practice one calculates orthometric height using an approximate formula. In Finland, Helmert’s formula has long been used, where gravity measured on the Earth surface, $g(H)$, is extrapolated downward by using the estimated gravity gradient internal to the rock. It is assumed that its standard value outside the rock, $-0.3086 \text{ mGal/m}$ (free-air gradient), changes to a value $+0.2238 \text{ mGal/m}$ greater (double Bouguer plate effect): the end result is the total gravity gradient, $-0.0848 \text{ mGal/m}$. 
This is called the Prey\(^3\) reduction. The end result are the following equations (the coefficient is half the gradient, i.e., the mean gravity along the plumbline is the same as gravity at the midpoint of the plumbline):

\[
\bar{g} = g(H) - 0.0848 \text{[mGal/m]} \left(-\frac{1}{2}H\right) = g(H) + 0.0424 \text{[mGal/m]} H, \text{ thus }
\]

\[
H = \frac{C}{g(H) + 0.0424 \text{[mGal/m]} H'}, 
\]

where \(C\) is the geopotential number (potential relative to mean sea level) and \(g(H)\) is gravity at the Earth’s surface. See also Heiskanen and Moritz (1967) pp. 163–167. Note that the term \(0.0424 \text{ mGal/m} \cdot H\) is typically much smaller than \(g(H)\), which is about \(9.8 \text{ m/s}^2 = 980,000 \text{ mGal}\)!

So, iteration, where the above denominator is first calculated using a crude \(H\) value, converges really fast.

The use of Helmert heights as approximations to orthometric heights is imprecise for the following reasons:

- The assumption that gravity changes linearly along the plumbline. This is not the case, especially not because of the terrain correction. In the precise computation of orthometric heights, one ought to compute the terrain correction separately for every point on the plumbline.
- The assumption that the free-air gradient is a constant, \(-0.3086 \text{ mGal/m}\). This is not the case, the gradient can easily vary by \(\pm 10\%\).
- The assumption that rock density is \(\rho = 2.67 \text{ g/cm}^3\). The true density value may easily vary by \(\pm 10\%\) or more around this assumed value.

The first approximation, neglecting the terrain effect, can be corrected by using Niethammer’s method (see Heiskanen and Moritz (1967) p. 167). It requires that, also in geoid computation, the terrain is correspondingly taken into account.

The third approximation, the density, can be removed as a problem by conventionally agreeing to use also in the corresponding geoid computation a standard density \(\rho = 2.67 \text{ g/cm}^3\). The surface thus obtained isn’t any more a true geoid then, but a “fake geoid”, for which no suitable name comes to mind.

\(^3\)Adalbert Prey (1873–1949) was an Austrian astronomer and geodesist and an author of textbooks.

\(^4\)Theodor Niethammer (1876–1947) was a Swiss astronomer and geodesist who was the first to map the gravity field of the Swiss Alps.
The second approximation could be eliminated by using the true free-air gradient instead of a standard value. However, the true gradient depends on local density variations. One can use, e.g., the Poisson equation for computing the gradient, on which more later.

The exact calculation of orthometric heights is thus laborious. Just as laborious as the exact computation of the geoid, and for the same reasons. Fortunately in non-mountainous countries Helmert heights are good enough. In Finland they were even computed using as \( \rho \) values “true” crustal densities according to a geological map…

### 6.7 Calculating normal heights precisely

For this we use the equation (6.4):

\[
H^e = \frac{C}{\bar{\gamma}} = -\frac{W - W_0}{\bar{\gamma}}, \tag{6.10}
\]

where the average value of normal gravity along the plumbline is

\[
\bar{\gamma} = \gamma_{0H} = \frac{1}{H^e} \int_0^{H^e} \gamma(z) \, dz.
\]

Because normal gravity is in good approximation a linear function of \( z \), we may write

\[
\bar{\gamma} = \gamma(H^e) - \frac{1}{2} H^e \frac{\partial \gamma}{\partial z},
\]

where \( \frac{\partial \gamma}{\partial z} = -0.3086 \text{ mGal/m} \). We obtain

\[
\bar{\gamma} = \gamma(H^e) + 0.1543 \left( \text{mGal/m} \right) \cdot H^e.
\]

The solution is again obtained iteratively:

\[
H^e = \frac{C}{\gamma(H^e) + 0.1543 \left( \text{mGal/m} \right) \cdot H^e} \tag{6.11}
\]

where \( \gamma(H^e) \) can be calculated exactly when the height \( H^e \) (and the local latitude) is known. \( H^e \) appears on both sides of the equation; it converges fast because again, the first term of the denominator \( \gamma(H^e) \), some \( 9.8 \text{ m/s}^2 = 980,000 \text{ mGal} \), is a lot larger than \( 0.1543 \left( \text{mGal/m} \right) \cdot H^e \).

Calculating normal heights is not in the same way sensitive to Earth crustal density and similar hypotheses, like calculation of orthometric heights is. It depends however on the choice of normal field, i.e., the reference ellipsoid.
6.8 Calculation example for heights

At point $P$ the potential difference with sea level is $C = 5000 \text{ m}^2/\text{s}^2$. Local gravity is $g = 9.820,000 \text{ m/s}^2$.

Questions:

1. Calculate the free-air anomaly $\Delta g_{FA}$ of point $P$.
2. Calculate the orthometric height of the point.
3. Calculate the Bouguer anomaly (without terrain correction) $\Delta g_B$ of the point.

Normal gravity calculated at the level of point $P$ (i.e., calculated for a point $Q$ that is as high above the ellipsoid as $P$ is above sea level) is $\gamma = 9.820,492 \text{ m/s}^2$.

4. Calculate the normal height of point $P$.
5. If the geoid height at point $P$ is $N = 25.000 \text{ m}$, how much is then the height anomaly ("quasi-geoid height") $\zeta$?

Answers:

1. The free-air anomaly is
   \[ \Delta g_{FA} = 9.820,000 - 9.820,492 \text{ m/s}^2 = -49.2 \text{ mGal}. \]

2. First attempt: $H^{(0)} = C/g = 5000/9.82 = 519.165 \text{ m}$. Second attempt (equation (6.9)):
   \[ H^{(1)} = \frac{5000 \text{ m}^2/\text{s}^2}{9.820000 \text{ m/s}^2 + 0.0424 \cdot 10^{-5} \text{ [m}^{-2}] \cdot 519.165 \text{ m}} = 509.154 \text{ m}; \]
   after that, the millimetres don’t change any more.

3. The Bouguer anomaly is (equation (5.2)):
   \[ \Delta g_B = \Delta g_{FA} - 0.1119 \text{ [mGal/m]} \cdot H = -106.2 \text{ mGal}. \]

4. The first attempt is again $H^{(0)} = C/\gamma = 509.139 \text{ m}$. The second, equation (6.11):
   \[ H^{(1)} = \frac{5000 \text{ m}^2/\text{s}^2}{9.820492 \text{ m/s}^2 + 0.1543 \cdot 10^{-5} \text{ [m}^{-2}] \cdot 509.139 \text{ m}} = 509.099 \text{ m}, \]
   also final on the millimetre level.

5. The difference equation (6.8) is
   \[ \zeta - N = -\frac{\Delta g_B H}{\gamma} = -0.055 \text{ m}. \]
   Also (check) $H^* - H = -0.055 \text{ m}$. So $\zeta = N - (-0.055 \text{ m}) = 25.055 \text{ m}$. 


6.9 Orthometric and normal corrections

In practical computations, one often adds together at first the height differences $\Delta H$ measured by levelling ("staff reading differences") between points $A$ and $B$ as a tentative or crude height difference

$$\sum_A^B \Delta H,$$

after which the non-exactness of this method is accounted for by applying the "orthometric correction":

$$H_B = H_A + \sum_A^B \Delta H + OC_{AB}.$$

The fact that the difference in orthometric heights between two points $A$ and $B$ is not equal to the sum of the levelled height differences is due to gravity not being the same everywhere.

With $C_A, C_B, \Delta C$ the geopotential numbers at $A$ and $B$, and the geopotential differences along the levelling line, we have $C_B - C_A - \sum_A^B \Delta C = 0$ because of the conservative nature of the geopotential. Dividing by a constant $\gamma_0$ yields

$$\frac{C_B}{\gamma_0} - \frac{C_A}{\gamma_0} - \frac{\sum_A^B \Delta C}{\gamma_0} = 0.$$

On the other hand we have

$$OC_{AB} = H_B - H_A - \sum_A^B \Delta H = \frac{C_B}{g_B} - \frac{C_A}{g_A} - \frac{\sum_A^B \Delta C}{g},$$

with $g_A, g_B$ mean gravities along the plumblines of $A$ and $B$, $g$ gravity along the levelling line. In this expression, we compare $\sum_A^B \Delta H$, the naively calculated sum of levelled height differences, with the difference between the orthometric heights of the end points $A$ and $B$, computed according to the definition.

Subtraction yields

$$OC_{AB} - 0 = \left( \frac{C_B}{g_B} - \frac{C_B}{\gamma_0} \right) - \left( \frac{C_A}{g_A} - \frac{C_A}{\gamma_0} \right) - \sum_A^B \left( \frac{\Delta C}{g} - \frac{\Delta C}{\gamma_0} \right),$$

where

$$\frac{C_B}{g_B} - \frac{C_B}{\gamma_0} = \left( \frac{\gamma_0 - g_B}{\gamma_0} \right) \frac{C_B}{g_B} = \left( \frac{\gamma_0 - g_B}{\gamma_0} \right) H_B,$$

$$\frac{C_A}{g_A} - \frac{C_A}{\gamma_0} = \left( \frac{\gamma_0 - g_A}{\gamma_0} \right) H_A,$$

$$\frac{\Delta C}{g} - \frac{\Delta C}{\gamma_0} = \left( \frac{\gamma_0 - g}{\gamma_0} \right) \Delta H.$$
yielding
\[
OC_{AB} = \sum_{A}^{B} \left( \frac{g - \gamma_{0}}{\gamma_{0}} \right) \Delta H + \left( \frac{\gamma_{A} - \gamma_{0}}{\gamma_{0}} \right) H_{A} - \left( \frac{\gamma_{B} - \gamma_{0}}{\gamma_{0}} \right) H_{B},
\] (6.12)
which is identical to Heiskanen and Moritz (1967) equation 4–33.

The choice of \( \gamma_{0} \) is arbitrary; it is wise to choose it close to the average gravity in the general area of \( A, B \) to keep the numerics small.

Similarly we may also compute the normal correction in calculating normal heights. Start from the equation
\[
NC_{AB} = H_{B}^{*} - H_{A}^{*} - \sum_{A}^{B} \Delta H = \frac{C_{B}}{\gamma_{B}} - \frac{C_{A}}{\gamma_{A}} - \sum_{A}^{B} \frac{\Delta C_{g}}{g},
\]
from which, like above, follows by subtraction
\[
NC_{AB} = \sum_{A}^{B} \left( \frac{g - \gamma_{0}}{\gamma_{0}} \right) \Delta H + \left( \frac{\gamma_{A} - \gamma_{0}}{\gamma_{0}} \right) H_{A}^{*} - \left( \frac{\gamma_{B} - \gamma_{0}}{\gamma_{0}} \right) H_{B}^{*},
\] (6.13)

Note that the identical first term in both equation (6.12) and equation (6.13) derives from
\[
\sum_{A}^{B} \frac{\Delta C_{g}}{g} = \sum_{A}^{B} \Delta H,
\]
the naive summation of height differences \( \Delta H \) in the case of both orthometric and normal correction, which is the generic basis of the concept of both corrections.

What changes between the orthometric and normal corrections is the definition of heights: \( H^{*} \) instead of \( H \), requiring division by the average of normal gravity along the plumbline \( \bar{\gamma} \), not by that of true gravity \( \bar{g} \).

Summarize this as
\[
H_{B}^{*} = H_{A}^{*} + \sum_{A}^{B} \Delta H + NC_{AB}.
\]

Note that both the orthometric correction (6.12) and the normal correction (6.13) can be calculated one levelling interval at a time: one must know, in addition to the levelled height difference \( \Delta H \), local gravity \( g \) along the levelling line, and also at the end points \( g (H) \) or \( \gamma (H^{*}) \) for calculating mean gravity \( \bar{g} \) or \( \bar{\gamma} \) along the plumblines of those end points. This goes well with the formulas given above. Remember that \( g \) along the levelling line is needed anyway in order to reduce the levelled height differences \( \Delta H \) to geopotential number differences \( \Delta C \).
Figure 6.8. An optical lattice clock: the ultra-precise atomic clock of the future operates at optical wavelengths. To the right, the trajectory of the Predehl et al. experiment.

6.10 A vision for the future: relativistic levelling

According to General Relativity, the deeper a clock is inside the potential well of masses, the slower it ticks. This is most easily seen by looking at the Schwartzschild metric for a spherically symmetric field:

$$c^2 d\tau^2 = \left(1 - \frac{2GM}{c^2 r}\right) c^2 dt^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 - r^2 (d\phi^2 + \cos^2 \phi d\lambda^2) =$$

$$\left(1 - \frac{2W}{c^2}\right) c^2 dt^2 - \left(1 - \frac{2W}{c^2}\right)^{-1} dr^2 - r^2 (d\phi^2 + \cos^2 \phi d\lambda^2),$$

in spherical co-ordinates plus time $$(r, \phi, \lambda, t)$$. Here we see, how the rate of the proper time $\tau$ is slowed down compared to stationary co-ordinate time $t$, when the geopotential $W$ increases closer to the mass. The slowing-down ratio is

$$\frac{\partial \tau}{\partial t} \approx \frac{W}{c^2}.$$ 

Now $c^2$ is a huge number: $10^{17} m^2/s^2$. This means that measuring a potential difference of $1 m^2/s^2$ – corresponding to a height difference of 10 cm – using this method, requires a precision of $1 : 10^{17}$. More traditional, microwave based atomic clocks can do precisions of $10^{-12} - 10^{-14}$ (Vermeer, 1983). With the new optical clocks the objective should be achievable.

The clock works in this way, that an extremely cold, so-called Bose-Einstein condensate of atoms is trapped inside an optical lattice formed by six laser beams, an electromagnetic pattern of standing waves; the readout beam of the clock oscillation uses a different frequency. A Bose-Einstein condensate
has the property that all atoms are in precisely the same quantum state – like the photons in an operating laser –, i.e., their matter waves are coherent. In a way, all atoms together act as one virtual atom.

The condensate may consist of millions of atoms, and can actually be seen through the window of the vacuum chamber as a small plasma blob.

Unfortunately it is not enough, that just one laboratory measures time to extreme precision – one also has to be able to compare the ticking rates of different clocks over geographical distances. Also for this, a solution has been found: existing optic-fibre cables already in global use for Internet and telephony are useable for this with small modifications. The modifications concern the amplifiers in the cables at distances of some 100 km, which must be replaced by modified ones (Predehl et al., 2012). In this way both the traditional precise levelling networks and the height systems based on GNSS technology and geoid determination may be replaced by this hi-tech (and hi-science!) solution.
Chapter 6. Vertical reference systems
7. The Stokes equation and other integral equations

7.1 The Stokes equation and the Stokes integral kernel

By suitably combining the equations in section 4.3 one obtains easily

\[ T = R \sum_{n=2}^{\infty} \frac{\Delta g_n}{n-1}, \]

(where for the degree numbers \( n = 0, 1 \), the \( \Delta g_n \) are assumed again to vanish, as \( n = 0 \) means a different total mass for the normal field than for the Earth, and \( n = 1 \) an offset of the co-ordinate origin from the Earth’s centre of mass). This is now the Stokes equation’s spectral form.

Using the degree constituent equation (2.19) one obtains the integral equation:

\[
T = \frac{R}{4\pi} \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \int_{\sigma} \Delta g P_n (\cos \psi) \, d\sigma = \\
= \frac{R}{4\pi} \int_{\sigma} \left[ \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n (\cos \psi) \right] \Delta g d\sigma = \\
= \frac{R}{4\pi} \int_{\sigma} S (\psi) \Delta g d\sigma,
\]

where

\[ S (\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n (\cos \psi), \]

the Stokes kernel function. The angle \( \psi \) is the angular distance between evaluation point and moving data point. The equation above allows the calculation, from global gravimetric data and for every point on the Earth surface, of the disturbing potential \( T \), and from that the geoid height \( N \) using the Bruns equation (4.2), \( N = T/\gamma \), with the result

\[ N (\phi, \lambda) = \frac{R}{4\pi \gamma} \int_{\sigma} S (\psi) \Delta g (\phi', \lambda') \, d\sigma', \quad (7.1) \]
where \((\phi, \lambda)\) and \((\phi', \lambda')\) are the evaluation point and the moving point ("data point") respectively, and the distance between them is \(\psi\). Equation (7.1) is the classical Stokes equation of gravimetric geoid computation.

The above illustrates the correspondence between integral equations and spectral equations. There are other examples of this. Earlier we presented the spectral representation of the function \(1/\ell\) (Heiskanen and Moritz (1967) equation 1-81). Of course \(1/\ell\) is also the kernel function of the integral equation yielding the potential \(V\) if given is the surface layer mass density \(\kappa\).

Also a version of the Stokes equation for the exterior space exists; we gave it earlier, equation (4.9). The spectral form of its kernel function, see equation (4.10), is

\[
S (r, \psi, R) = \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \frac{2n + 1}{n - 1} P_n (\cos \psi).
\]

The Stokes kernel function on the Earth’s surface is depicted in figure 7.3, where the angle \(\psi\) is given in radians (1 rad \(\approx 57^\circ.3\)).

This curve was calculated using the following closed expression (see Heiska-
\[ S(\psi) = \frac{1}{\sin(\psi/2)} - 6\sin\frac{\psi}{2} + 1 - 5\cos\psi - 3\cos\psi \ln\left(\sin\frac{\psi}{2} + \sin^2\frac{\psi}{2}\right). \] (7.2)

This closed expression helps us to understand better how the function behaves close to the origin \( \psi = 0 \): the first term, \( (\sin(\psi/2))^{-1} \), goes to infinity when \( \psi \to 0 \). The next three terms, \(-6\sin\frac{\psi}{2} + 1 - 5\cos\psi\), are all bounded on the whole interval \([0, \pi]\) and the limit for \( \psi \to 0 \) is \(-4\). The last, complicated term \(-3\cos\psi \ln\left(\sin\frac{\psi}{2} + \sin^2\frac{\psi}{2}\right)\) goes also to infinity – positive infinity! – but much more slowly, because of the logarithm.

### 7.2 Plumbline deflections and the Vening Meinesz equations

By differentiating the Stokes equation with respect to place we obtain integral equations for the components of the deflections of the plumbline (Heiskanen and Moritz, 1967 Eq. 2.210’):

\[
\begin{align*}
\left\{ \begin{array}{c}
\zeta \\
\eta
\end{array} \right\} &= \frac{1}{4\pi\gamma} \iint_{\sigma} \Delta g \left. \frac{dS(\psi)}{d\psi} \right| \cos \alpha \\
&\quad \left/ \left| \sin \alpha \right| \right. \\
&= \frac{1}{4\pi\gamma} \iint_{\sigma} \Delta g \left. \frac{dS(\psi)}{d\psi} \right| \cos \frac{\alpha}{\sin \alpha} \\
&\quad \cdot \sin \psi d\alpha d\psi,
\end{align*}
\] (7.3)

where \( \zeta, \eta \) are the North-South and East-West direction deflections of the plumbline, and the unit-sphere surface element \( d\sigma = \sin \psi d\alpha d\psi \), where \( \sin \psi \) is the Jacobian of the \((\alpha, \psi)\) co-ordinates.

These equations were derived for the first time by the Dutch geophysicist F.A. Vening Meinesz. The angle \( \alpha \) is the azimuth or direction angle between

---

**Figure 7.2.** Integrating the Stokes equation geometrically.
Chapter 7. The Stokes equation and other integral equations

Figure 7.3. The Stokes kernel function $S(\psi)$. The argument $\psi$ in radians $[0, \pi)$. Also plotted are the three parts of the analytical expression (7.2) with their different asymptotic behaviours.

The calculation or evaluation point $(\phi, \lambda)$ and the moving integration or data point $(\phi', \lambda')$. These equations are much harder to write in spectral form, as the kernel functions are now also functions of the azimuth direction $\alpha$, in other words, they are anisotropic.

The disturbing potential, the gravity disturbance, and the gravity anomaly, are all so-called isotropic quantities: they do not depend on the direction, and therefore in the spectral representation the transformations between them are only functions of degree $n$.

7.3 The Poisson integral equation

Look at figure 7.4. The point $Q$ of the body is located at $R$, and the observation point $P$ at location $r$, the angular distance between the two location vectors, as seen from the origin, is $\psi$. The distance between points $P$ and $Q$ is $\ell$.

First we may write

$$\ell = \sqrt{r^2 + R^2 - 2rR \cos \psi}.$$

We may also write the function $1/\ell$ as the following expansion (for proof,
see Heiskanen and Moritz (1967) p. 33):

\[
\frac{1}{\ell} = \frac{1}{\sqrt{r^2 + R^2 - 2R r \cos \psi}} = \frac{1}{R} \sum_{n=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} P_n \left( \cos \psi \right) \tag{7.4}
\]

where \( r = \|r\| \) and \( R = \|R\| \) are the distances of points \( P \) and \( Q \) from the origin, the centre of the Earth.

\( 1/\ell \) is called the generating function of the Legendre polynomials.

Differentiating equation (7.4) with respect to \( r \) yields

\[
-\frac{r - R \cos \psi}{\ell^3} = -\frac{1}{R} \sum_{n=0}^{\infty} \frac{n+1}{r} \left( \frac{R}{r} \right)^{n+1} P_n \left( \cos \psi \right).
\]

This we multiply by \( 2r \):

\[
-\frac{2r^2 - 2rR \cos \psi}{\ell^3} = -\frac{1}{R} \sum_{n=0}^{\infty} (2n+2) \left( \frac{R}{r} \right)^{n+1} P_n \left( \cos \psi \right).
\]

Now we add together this equation and equation (7.4):

\[
-\frac{2r^2 + 2rR \cos \psi + \ell^2}{\ell^3} = -\frac{1}{R} \sum_{n=0}^{\infty} (2n+1) \left( \frac{R}{r} \right)^{n+1} P_n \left( \cos \psi \right).
\]

The left-hand side is simplified by using \( -r^2 + 2rR \cos \psi + \ell^2 = R^2 \):

\[
-\frac{2r^2 + 2rR \cos \psi + \ell^2}{\ell^3} = \frac{R^2 - r^2}{\ell^3},
\]

and the end result is (by multiplying by \( -R \)):

\[
\frac{R \left( r^2 - R^2 \right)}{\ell^3} = \sum_{n=0}^{\infty} (2n+1) \left( \frac{R}{r} \right)^{n+1} P_n \left( \cos \psi \right). \tag{7.5}
\]
Borrowing now the degree constituent equation (2.19) for the potential function $V$ on the spherical Earth’s surface, radius $R$:

$$V_n (\phi, \lambda) = \frac{2n + 1}{4\pi} \int_\sigma V (\phi', \lambda', R) P_n (\cos \psi) \, d\sigma',$$

as well as equation (2.12):

$$V (\phi, \lambda, r) = \sum_{n=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} V_n (\phi, \lambda),$$

we obtain

$$V (\phi, \lambda, r) = \frac{1}{4\pi} \sum_{n=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} (2n + 1) \int_\sigma V (\phi', \lambda', R) P_n (\cos \psi) \, d\sigma' =$$

$$= \frac{1}{4\pi} \int_\sigma V (\phi', \lambda', R) \left[ \sum_{n=0}^{\infty} (2n + 1) \left( \frac{R}{r} \right)^{n+1} P_n (\cos \psi) \right] \, d\sigma' =$$

$$= \frac{R}{4\pi} \int_\sigma \frac{(r^2 - R^2)}{\ell^3} V (\phi', \lambda', R) \, d\sigma'$$

by substituting directly into equation (7.5).

Thus we have obtained the Poisson equation for computing a harmonic function $V$ from values given on the Earth’s surface:

$$V_p = \frac{R}{4\pi} \int_\sigma \frac{(r^2 - R^2)}{\ell^3} V_Q d\sigma_Q,$$  \hspace{1cm} (7.6)

where $\ell$ is again the straight distance between evaluation point $P$ (where $V_p$ is being computed) and moving integration point $Q$ (on the surface of the sphere, $V_Q$ under the integral sign). In this equation we have given the points names: the co-ordinates of $P$ are $(\phi, \lambda, r)$, the co-ordinates of the integration point $Q$ are $(\phi', \lambda', R)$.

Still a third way to write the same equation, useful when the function or field $V$ isn’t actually defined between the topographic Earth surface and sea level, is

$$V = \frac{R}{4\pi} \int_\sigma \frac{(r^2 - R^2)}{\ell^3} V^* \, d\sigma,$$

where $V^*$ denotes the value of a function $V$ harmonically downward continued into the topography, all the way to sea level (or, in spherical approximation, to the surface of the sphere $r = R$), i.e., a function that above the topography is identical to $V$, that is harmonic, and that also exists between topography and sea level. The existence of such a function has been a classical theoretical nut to crack...
7.4 Gravity anomalies in the exterior space

The equation derived in the previous section 7.3, equation (7.6), applies for an arbitrary harmonic field $V$, i.e., a field for which $\Delta V = 0$. The equation may be conveniently applied to the expression $r \cdot \Delta g$, i.e., the gravity anomaly multiplied by the radius, which is also a harmonic function. This is how we can express the gravity anomaly in the external space $\Delta g(\phi, \lambda, r)$ as a function of gravity anomalies $\Delta g(\phi', \lambda', R)$ on a sphere of radius $R$. The function $r \Delta g$ is harmonic, because according to equation (4.7)

$$\Delta g = \frac{1}{r} \sum_{n=2}^{\infty} (n-1) \left( \frac{R}{r} \right)^{n+1} T_n,$$

i.e.,

$$r \Delta g = \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} (n-1) T_n = \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} S_n,$$

where $S_n(\phi, \lambda) = (n-1) T_n(\phi, \lambda)$ is a legal surface spherical harmonic just like $T_n(\phi, \lambda)$ itself. Also, the dependence on the radius $r$, the factor $(R/r)^{n+1}$, is the same as for the (harmonic) potential. So, Poisson’s integral equation (7.6) applies to function $r \Delta g$:

$$[r \Delta g(\phi, \lambda, r)] = \frac{R^2}{4\pi} \int_{\sigma} \frac{(r^2 - R^2) \Delta g(\phi', \lambda', R)}{\ell^3} d\sigma' ,$$

or

$$\Delta g(\phi, \lambda, r) = \frac{R^2}{4\pi r} \int_{\sigma} \frac{(r^2 - R^2) \Delta g(\phi', \lambda', R)}{\ell^3} d\sigma'.$$

An alternative notation:

$$\Delta^g = \frac{R^2}{4\pi r} \int_{\sigma} \frac{r^2 - R^2}{\ell^3} \Delta^g d\sigma,$$

where $\Delta^g$ denotes the gravity anomaly at sea level, again calculated by harmonic downward continuation of the exterior field, in this case the expression $r \Delta g$.

See also Heiskanen and Moritz (1967) equation 2-160. Using the approximation $r + R \approx 2r$ yields still

$$\Delta g(\phi, \lambda, r) \approx \frac{R^2}{2\pi} \int_{\sigma} \frac{(r - R) \Delta g(\phi', \lambda', R)}{\ell^3} d\sigma'.$$

Alternatively we derive the spectral form:

$$\Delta g = \frac{1}{r} \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} (n-1) T_n = \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+2} \Delta g_n.$$
The degree constituent equation (2.19) gives the functions $\Delta g_n$:

$$\Delta g_n = \frac{2n + 1}{4\pi} \int_\sigma \Delta g P_n (\cos \psi) d\sigma,$$

with the aid of which

$$\Delta g = \frac{1}{4\pi} \sum_{n=2}^\infty \left( \frac{R}{r} \right)^{n+2} (2n + 1) \int_\sigma \Delta g P_n (\cos \psi) d\sigma =$$

$$= \frac{1}{4\pi} \int_\sigma \left[ \sum_{n=2}^\infty \left( \frac{R}{r} \right)^{n+2} (2n + 1) P_n (\cos \psi) \right] \Delta g d\sigma =$$

$$= \frac{1}{4\pi} \int_\sigma K \Delta g d\sigma,$$

where

$$K(r, \psi, R) = \sum_{n=2}^\infty \left( \frac{R}{r} \right)^{n+2} (2n + 1) P_n (\cos \psi)$$

is a (modified) Poisson kernel for gravity anomalies. Its closed form can be lifted from equation (7.8):

$$K(r, \psi, R) = \frac{R^2 - R^2}{r^3}.$$

Compared to the Stokes kernel, the Poisson kernel drops off fast to zero for growing $\ell$ values. In other words, the evaluation of the integral equation may be restricted to a very local area, e.g., a cap of radius 1°. See figure 7.5. The main use of Poisson’s kernel is the harmonic continuation, upward or downward, i.e., the shifting of gravity anomalies measured and computed at various levels to the same reference level.

In the limit $r \to R$ (sea level becomes the level of evaluation) this kernel function goes asymptotically to the Dirac $\delta$ function.

The difference between equations (7.9) and (7.8) is only in their numerical behaviour.

### 7.5 The vertical gradient of the gravity anomaly

Differentiate a formula obtained from formulas (4.6), (4.7):

$$\Delta g = \sum_{n=2}^\infty \left( \frac{R}{r} \right)^{n+2} \Delta g_n \Rightarrow \frac{\partial \Delta g}{\partial r} = \frac{1}{R} \sum_{n=2}^\infty \left( \frac{R}{r} \right)^{n+3} (n + 2) \Delta g_n.$$
\( \Delta g_n \) is, as calculated according to the degree constituent equation (2.19) from the anomaly field at sea level:

\[
\Delta g_n = \frac{2n+1}{4\pi} \int_\sigma \Delta g (\phi', \lambda', R) P_n (\cos \psi) \, d\sigma',
\]

i.e.,

\[
\frac{\partial \Delta g (\phi, \lambda, r)}{\partial r} = - \frac{1}{4\pi R} \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+3} \frac{(2n+1) \, (n+2)}{2r} \cdot
\]

\[
\cdot \int_\sigma \Delta g (\phi', \lambda', R) P_n (\cos \psi) \, d\sigma' = \frac{1}{4\pi R} \int_\sigma K' (r, \psi, R) \Delta g (\phi', \lambda', R) \, d\sigma',
\]

where the kernel function is

\[
K' (r, \psi, R) = - \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+3} (2n+1) \, (n+2) \, P_n (\cos \psi).
\]

Alternatively we derive a closed formula. We start from the Poisson equation (7.8) for gravity anomalies, and differentiate with respect to \( r \):

\[
\frac{\partial \Delta g (\phi, \lambda, r)}{\partial r} = \frac{\partial}{\partial r} \left[ \frac{R^2}{4\pi r} \int_\sigma \frac{(r^2 - R^2)}{(r^2 + R^2 - 2rR \cos \psi)^{3/2}} \Delta g (\phi', \lambda', R) \, d\sigma' \right] = \frac{R^2}{4\pi} \int_\sigma \frac{1}{\ell^3} \left[ 2 - \frac{3(r^2 - R^2)}{2r^2 \ell^2} \right] \Delta g (\phi', \lambda', R) \, d\sigma' = \frac{R^2}{4\pi} \int_\sigma \frac{1}{\ell^3} \left[ 2 - \frac{3(r^2 - R^2)}{2r^2 \ell^2} \right] \Delta g (\phi', \lambda', R) \, d\sigma' = - \frac{1}{r} \Delta g (\phi, \lambda, r) = \frac{R^2}{4\pi} \int_\sigma \frac{1}{\ell^3} \left[ 2 - \frac{3(r^2 - R^2)}{2r^2 \ell^2} \right] \Delta g (\phi', \lambda', R) \, d\sigma' = - \frac{1}{r} \Delta g (\phi, \lambda, r) = \frac{R^2}{4\pi} \int_\sigma \frac{1}{\ell^3} \left[ 2 - \frac{3(r^2 - R^2)}{2r^2 \ell^2} \right] \Delta g (\phi', \lambda', R) \, d\sigma' = - \frac{1}{r} \Delta g (\phi, \lambda, r) = \frac{R^2}{4\pi} \int_\sigma \frac{1}{\ell^3} \left[ 2 - \frac{3(r^2 - R^2)}{2r^2 \ell^2} \right] \Delta g (\phi', \lambda', R) \, d\sigma' = - \frac{1}{r} \Delta g (\phi, \lambda, r).
\]

1Hint: use symbolic algebra software.
Chapter 7. The Stokes equation and other integral equations

Finally, note that, if we are integrating over the surface of the Earth sphere, radius \( R \), rather than the unit sphere \( \sigma \), radius 1, the coefficient \( R^2 \) drops out from both eq. (7.8) and eq. (7.11).

In Molodensky’s method this or similar equations can be rapidly evaluated from very local gravimetric data.

The closed expression given in Heiskanen and Moritz (1967), expression 2-217\(^2\), is the vertical gradient evaluated at sea level (on the reference sphere). That is why it differs from expression (7.11) given above. Also in the expression given here, like in expression (7.10), we need gravity anomalies at sea level. Available however are anomalies at the topographic surface level. In practice we can proceed iteratively, by first assuming that the anomaly values observed at topography level are at sea level:

\[
\Delta g^{(0)} (\phi', \lambda', R) \approx \Delta g (\phi', \lambda', r) = \Delta g (\phi', \lambda', R + H),
\]

where \( H = H (\phi', \lambda') \) is the topographic height at point \((\phi', \lambda')\). When the first, crude anomaly gradient has been calculated, e.g., using eq. (7.11), we may perform a real \textit{reduction to sea level}, initially linearly:

\[
\Delta g^{(1)} (\phi', \lambda', R) \approx \Delta g (\phi', \lambda', R + H) - \left. \frac{\partial \Delta g}{\partial r} \right|^{(0)} H,
\]

and so forth.

\(^2\)In the derivation there it is assumed that \( \Delta g \) is harmonic. Not so: \( r \Delta g \) is harmonic.
7.6 Gravity reductions in geoid determination

Use of the Stokes equation for gravimetric geoid determination presupposes that all masses are inside the geoid (and the exterior field is thus harmonic). For this reason we move the topographic masses computationally to inside the geoid. Of the many methods for achieving this, really only four come into question if one wishes to calculate a (quasi-)geoid on the Earth surface and outside. In each method, mass shifts take place that need to be specified.

- Helmert’s (second) condensation method: the masses are shifted vertically down to the geoid into a surface density layer. After this, shifting gravity down from the topographic surface to sea level is easy. The indirect effect (the effect of the mass shifts on the geoid, the “Restore” step) is small.

- Bouguer reduction: the indirect effect of this reduction is excessive and extends over a large area, which is why it is more rarely used. Here, the effect of the topographic masses is brutally removed from the observed gravity data, and, after geoid calculation, it is equally brutally restored to the result.

- The Molodensky method: the topographic masses are shifted into the geoid, to below sea level, in a way that does not change the exterior field. This is actually the same as saying that we are determining the geoid associated with the harmonically downward continued exterior field. The problem here is, that such a mass distribution below sea level, which produces the harmonically downward continued external potential in the space between topographic surface and geoid, perhaps doesn’t even precisely exist. Or, that a suitable mass distribution will contain extremely large positive and negative masses close to each other, which are physically unrealistic.

One expresses this by saying that the problem is “ill posed”. As a solution one uses regularization: we change a little – as little as possible – the exterior field, so that it corresponds precisely to some sensible field interior to the topography, and some sensible mass distribution interior to the geoid that produces it. One can start, e.g., by filtering out the short wave parts caused by the topography using a high resolution digital terrain model.

- The RTM (Residual Terrain Modelling) method. This consists of two
Figure 7.6. Residual terrain modelling (RTM). One removes from the terrain computationally the short wavelengths, i.e., the differences from the red dashed line: the masses rising above it are removed, the valleys below it are filled. After reduction, the red line (smoother than the original terrain) is the new terrain surface. The exterior potential of the new mass distribution will differ only little from the original one, but will be smoother and may therefore be harmonically downward continued to sea level.

stages:

1. First we computationally remove from the topography the short wavelengths (under 30 km) by moving the masses of the peaks into the valleys; the effect of this on the measured gravity anomalies is evaluated (“Remove” step).

2. After that, we apply Molodensky’s recipe, which now works as intended, because the first stage works as a filter: in the external field remain nearly only the long wavelengths, which can be successfully harmonically downward continued.

3. Because the mass shifts involved are so small, take place over such small distances, and are so short wavelength in nature, the indirect effect or “Restore” step – the change in geopotential due to the mass shifts that has to be applied in reverse to arrive at the final geopotential or geoid solution – is so small as to be typically negligible.

Only the first stage changes the exterior field. Therefore, e.g., the effect of unknown topographic density will remain minimal.

7.6.1 The Molodensky method in linear approximation

The method of Molodensky described above can be linearized:

\[
T(\phi, \lambda, H) = \frac{R}{4\pi} \iint \left[ \Delta g(\phi', \lambda', H') - \frac{\partial \Delta g}{\partial H'} H' \right] S(\psi) \, d\sigma' + \frac{\partial T}{\partial H} H'. \tag{7.12}
\]

So, first we reduce the \( \Delta g \) measured and calculated at the topographic surface to sea level using the gradient of the anomalies and the terrain height \( H' \) of the measurement point, with the result

\[
\Delta g^* = \Delta g - \frac{\partial \Delta g}{\partial H'} H'.
\]
7.6. Gravity reductions in geoid determination

After this, we apply, at sea level, the Stokes equation, and obtain the disturbing potential at sea level $T^*$. After this, the disturbing potential is “unreduced” back to terrain level, to the evaluation point, with the equation

$$T = T^* + \frac{\partial T}{\partial H} H.$$ 

In these equations $T$, its vertical derivative, and $\Delta g$ and its vertical derivative always belong to the external harmonic gravity field, and the connection between them is the fundamental equation of physical geodesy, equation (4.4), in spherical geometry:

$$\Delta g = - \frac{\partial T}{\partial H} - \frac{2}{r} T,$$

where $r = R + H$. Here, we need firstly the vertical derivative of the disturbing potential. This is easy: we have

$$\frac{\partial T}{\partial H} = -\Delta g - \frac{2}{r} T,$$

where the first term is directly measured, and the second term’s $T$ is obtained iteratively from the main product of the solution process.

Calculating the gradient of gravity anomalies is harder. For this, we have the integral equation (7.10):

$$\frac{\partial \Delta g}{\partial r} = -\frac{1}{4\pi R} \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+3} (2n+1)(n+2) \cdot \int_{\sigma} \Delta g (\phi', \lambda', H') P_n (\cos \psi) d\sigma'.$$

Luckily for practical calculations, this integral is very localized and one does not need gravimetric data $\Delta g$ from a very large area.

7.6.2 The evaluation point as the reference level

In the above equation (7.12) we used as the reference level the sea surface. This is arbitrary: we may use whatever reference level, e.g., $H_0$, in which case

$$T (\phi, \lambda, H) = \frac{R + H_0}{4\pi} \int_{\sigma} \left( \Delta g (\phi', \lambda', H') - \frac{\partial \Delta g}{\partial H} (H' - H_0) \right) S (\psi) d\sigma' + \frac{\partial T}{\partial H} (H - H_0).$$

If we now choose $H_0 = H$, the last term drops off, and we obtain

$$T (\phi, \lambda, H) = \frac{R + H}{4\pi} \int_{\sigma} \left( \Delta g (\phi', \lambda', H') - \frac{\partial \Delta g}{\partial H} (H' - H) \right) S (\psi) d\sigma'.$$
In this case the reduction takes place from the height of the $\Delta g$ measurement point to the height of the $T$ evaluation point, probably a shorter distance than from sea level to evaluation height, especially in the immediate surrounding of the evaluation point. This means that the linearization error will remain smaller. What is bad, on the other hand, is that the expression in parentheses is now different for each evaluation point. This complicates the use of FFT based computation techniques, see later.

Here, we were all the time discussing the determination of the disturbing potential $T(\phi, \lambda, H)$; this is in practice the same as determining the height anomaly

$$\zeta(\phi, \lambda, H) = \frac{T(\phi, \lambda, H)}{\gamma(\phi, H)}$$

Here, $\gamma$ is normal gravity calculated for point latitude $\phi (\approx \varphi)$ and height $H$.

### 7.7 The Remove-Restore method

All geoid determination methods currently in use are in one way or another “Remove-Restore” methods, even in several different ways.

1. From the observed gravity values, first the effect of the global gravity field is removed. The global model is generally given in the form of a spherical harmonic expansion. Thus, a residual gravity field is obtained

   - that has numerically smaller values (easier to work with), and
   - that is more local: the long “wavelengths”, the systematic trends of large areas, are removed from the residual field, only the local details remain.

2. From the observed gravity, the effects of all masses are removed that are outside the geoid. The purpose of this is to obtain a residual gravity field

   - on which the Stokes equation may be applied, because no masses are left outside the boundary surface; and
   - from which especially the very small “wavelengths” (details the size of which is of order a few km) caused by the topography, are gone. After this, prediction of gravity values from sparse measurement values will work better.
Gravity reduction methods, i.e., methods which computationally remove the gravity effect of the exterior masses, which have good prediction properties, are Bouguer reduction – though Bouguer anomalies contain large negative biases in the mountains – and isostatic reduction.

We may illustrate the Remove-Restore method by the following commutative diagram:

<table>
<thead>
<tr>
<th>“Remove”</th>
<th>“Restore”</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta g$</td>
<td>“Brute force” $\rightarrow$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\leftarrow$ Global grav. field model $\rightarrow$</td>
</tr>
<tr>
<td>$\Delta g_{loc}$</td>
<td>$\leftarrow$ Exterior masses $\rightarrow$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>$\leftarrow$ Stokes $\Rightarrow$</td>
</tr>
</tbody>
</table>

In this diagram the thick arrows denote calculations that are recommended, because they are easy and accurate. The arrow in parentheses, a direct computation, is troublesome and compute intensive.

### 7.8 Kernel modification in the Remove-Restore method

In the Remove-Restore method described above, the handling of reduced gravity anomalies $\Delta g_{red}$ and geoid heights $N_{red}$ happens typically within a relatively small area. E.g., when using the FFT method, the area of computation is often a rectangular area in the map projection plane, drawn generously around the country or area the geoid of which is being computed.

Also if we compute the geoid directly by integrating the Stokes equation, we will evaluate this integral, after removing the global model, only over a limited area or cap. I.e., we evaluate the equation

$$N_{red} = \frac{R}{4\pi\gamma} \int_{\sigma_0} S(\psi) \Delta g_{red} d\sigma,$$

(7.13)

where $\sigma_0$ is a cap on the unit sphere, the radius of which is, e.g., $\psi_0$.

The assumption behind this is, that, outside the cap, $\Delta g_{red}$ is both small and rapidly varying, because the longer wavelengths have been removed from it with the global-model reduction. This may however be a dangerous assumption.

Write, in the above equation (7.13),

$$S(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos \psi)$$
and
\[ \Delta g_{\text{red}}(\phi, \lambda) = \sum_{n=L+1}^{\infty} \Delta g_n(\phi, \lambda), \]
assuming that \( L \) is the largest degree number that is still along in the global spherical harmonic model that was removed from the data\(^3\).

Now, because \( \Delta g_n \) is a certain linear combination of the surface spherical harmonics
\[
Y_{nm}(\psi, \alpha) = \begin{cases} 
  P_{n|m|}(\cos \psi) \sin |m| \alpha & m = -n, \ldots, -1, \\
  P_{nm}(\cos \psi) \cos m \alpha & m = 0, \ldots, n,
\end{cases}
\]
i.e.,
\[ \Delta g_n(\psi, \alpha) = \sum_{m=-n}^{n} \Delta g_{nm} Y_{nm}(\psi, \alpha), \]
and also
\[ Y_{n0}(\psi, \alpha) = P_n(\cos \psi), \]
it follows from the orthogonality of the \( Y \) functions, that
\[ \iint_{\sigma} P_n(\cos \psi) Y_{n'm'd\sigma} = \iint_{\sigma} Y_{n0} Y_{n'm'd\sigma} = 0 \text{ if } n \neq n' \text{ or } m \neq 0. \]

Now we may write – note that the terms \( n \leq L \) drop away:
\[
S(\psi) \Delta g_{\text{red}}(\phi, \lambda) = S(\psi) \Delta g_{\text{red}}(\psi, \alpha) = \]
\[ = \left[ \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos \psi) \right] \left[ \sum_{n=L+1}^{\infty} \sum_{m=-n}^{n} \Delta g_{nm} Y_{nm}(\psi, \alpha) \right] = \]
\[ = \left[ \sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n(\cos \psi) \right] \left[ \sum_{n=L+1}^{\infty} \sum_{m=-n}^{n} \Delta g_{nm} Y_{nm}(\psi, \alpha) \right] = \]
\[ = S^L(\psi) \Delta g_{\text{red}}(\phi, \lambda), \]
where
\[ S^L(\psi) = \sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n(\cos \psi) \]
is a so-called modified Stokes kernel function. The degree number \( L \) is called the modification degree. The size of the evaluation area \( \sigma_0 \) is chosen to be compatible with this.

The modification method described here, restricting the Legendre expansion of the \( S \) function to higher degree numbers, is called the Wong-Gore\(^4\) modification (Wong and Gore, 1969).

\(^3\)...and that the model is accurate!

\(^4\)L. Wong and R.C. Gore worked at the Aerospace Corporation, a space technology research institution in California. [http://en.wikipedia.org/wiki/The_Aerospace_Corporation](http://en.wikipedia.org/wiki/The_Aerospace_Corporation).
A desirable property of the new kernel function $S^L$ is, that it would be – at least compared to the original function $S$ – small outside the cap area $\sigma_0$. In that case restricting the integral to the cap instead of the whole unit sphere (equation (7.13)) does not do much damage. It is clear that $S^L$ is much narrower than $S$, as in it, only the higher harmonic degrees are represented. This can be verified by plotting a graph of both curves. It doesn’t go totally to zero outside the cap, however, but oscillates somewhat.

The reason for this oscillation is, that in the frequency (i.e., degree number) domain, the cut-off of the modified kernel is quite sharp. Transforming such a sharp edge between space and frequency domains will invariably produce an oscillation, which is related to the so-called Gibbs\(^5\) phenomenon.

\subsection{7.9 Advanced kernel modifications}

In the literature, other kernel modification methods are found. Their general form is

$$S^L (\psi) = \sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n (\cos \psi) + \sum_{n=2}^{L} (1 - s_n) \frac{2n+1}{n-1} P_n (\cos \psi) =$$

$$= S (\psi) - \sum_{n=2}^{L} s_n \frac{2n+1}{n-1} P_n (\cos \psi), \quad (7.14)$$

\(^5\)Josiah Willard Gibbs (1839 – 1903) was an American physicist, chemist, mathematician and engineer.
where the modification coefficients $s_n$, $n = 2, \ldots, L$ can be chosen. They are chosen so as to minimize the values of $S^L$ in the area outside the cap, $\sigma - \sigma_0$. In this way one may eliminate the truncation error of equation (7.13), and the oscillation of the Wong-Gore modification, almost entirely. Molodensky et al. (1962) developed already earlier such a method.

In the above eq. (7.14) we want to minimize the function $S^L$ over the area outside a local cap, $\sigma - \sigma_0$, i.e., minimize
\[
S^L(\psi) = S(\psi) - \sum_{n=2}^{L} s_n \frac{2n + 1}{n - 1} P_n(\cos \psi).
\]

Let us multiply this expression with each of the Legendre polynomials $P_n(\cos \psi), n = 2, \ldots, L$ in turn, and integrate over the area $\sigma - \sigma_0$ outside the local cap:
\[
\int_{\sigma - \sigma_0} S(\psi) P_n(\cos \psi) \, d\sigma - \sum_{n'=2}^{L} s_{n'} \frac{2n' + 1}{n' - 1} \int_{\sigma - \sigma_0} P_{n'}(\cos \psi) P_n(\cos \psi) \, d\sigma = 0, \quad n = 2, \ldots, L,
\]
a system of $L - 1$ equations in the $L - 1$ unknowns $s_{n'}$:
\[
\sum_{n'=2}^{L} A_{nn'} s_{n'} = b_n,
\]
with
\[
b_n = \int_{\sigma - \sigma_0} S(\psi) P_n(\cos \psi) \, d\sigma
\]
and
\[
A_{nn'} = \frac{2n' + 1}{n' - 1} \int_{\sigma - \sigma_0} P_{n'}(\cos \psi) P_n(\cos \psi) \, d\sigma.
\]

From this we can solve the $s_n$ for every degree number $n$ from 2 to $L$.

This solution sets to zero the expressions
\[
\int_{\sigma - \sigma_0} S^L(\psi) P_n(\cos \psi) \, d\sigma, \quad (7.15)
\]
also for all values $n$ from 2 to $L$.

The expressions (7.15) can be understood as *inner or scalar products*, between the functions $S^L$ and $P_n$. Similarly the elements of $A_{nn'}$ contain the scalar products between the functions $P_n$ and $P_{n'}$. Note that these scalar products do not vanish: when integrating over $\sigma - \sigma_0$, unlike over $\sigma$, the Legendre

---

The choice $s_n = 1$ again gives the simply modified Stokes kernel from which the low degrees have been completely removed.
7.10 **The effect of the local zone**

In numerical gravimetric geoid computation one uses *means* of anomalies computed over standard-sized cells or *blocks*, generally $5' \times 5'$, $10' \times 10'$, $30' \times 30'$ etc. At European latitudes, often sizes $3' \times 5'$, $5' \times 10'$, $6' \times 10'$ etc., are used, which are approximately square.

The following equation applies when evaluating an integral using block means:

$$N = \sum_i c_i \overline{\Delta g}_i,$$

where $\overline{\Delta g}_i$ is the mean of block $i$, and the weight

$$c_i (\lambda, \phi) = \frac{R}{4\pi G} \int_{\sigma_i} S (\psi) \, d\sigma,$$

where $\sigma_i$ is the surface area of block $i$.

Numerical evaluation of such an integral, or *quadrature*, is done conveniently
Chapter 7. The Stokes equation and other integral equations

\[ c_i(\lambda, \phi) = \frac{R}{4\pi\gamma} \int_{\lambda_1}^{\lambda_2} \int_{\phi_1}^{\phi_2} S(\lambda, \phi, \lambda', \phi') \cos \phi \, d\lambda' \, d\phi' \approx \]
\[ \approx \frac{\Delta \lambda \Delta \phi R}{4\pi\gamma} \sum_{j=1}^{3} w_j \sum_{k=1}^{3} w_k S_{jk}(\lambda, \phi), \]

where \( \Delta \lambda \) and \( \Delta \phi \) are the block size, \( w_1 = w_3 = 1/6 \) and \( w_2 = 4/6 \). \( S_{11}, \ldots, S_{33} \) are the values of expression \( [S(\lambda, \phi, \lambda', \phi') \cos \phi] \) at the nodal points used in the evaluation, \( 3 \times 3 \) of them. See figure 7.8. Also more complicated formulas (repeated Simpson or Romberg) can be employed.

One can show that the effect of the local (inner) zone on the geoid in the evaluation point \((\phi, \lambda)\) is proportional to the gravity anomaly in the point itself, \( \Delta g \). Starting from the Stokes equation (7.1) with \( S(\psi) \approx (\sin \psi/2)^{-1} \approx 2/\psi \) we find, for a circular inner zone of radius \( \psi_0 \):

\[ N_{\text{int}} = \frac{R}{4\pi\gamma} \int_{0}^{2\pi} \int_{0}^{\psi_0} \frac{2}{\psi} \Delta g \sin \psi \, d\psi \, d\alpha \approx \frac{R}{4\pi\gamma} \cdot 2\pi \cdot \Delta g_0 \cdot 2\psi_0 = d \Delta g_0. \]

Here \( d = R\psi_0 \) is the radius of the local block or cap in metres. \( \Delta g_0 \) is the block mean of the gravity anomaly.

The local contributions to the deflections of the plumbline again are proportional to the horizontal gradients of gravity anomalies. We start from the Vening-Meinesz equations (7.3), with the above approximations for a local

---

7Thomas Simpson FRS (1710–1761) was an English mathematician and textbook writer. Actually Simpson’s rule was used already a century earlier by Johannes Kepler.
We expand $\Delta g$ into local rectangular metric co-ordinates $x, y$:

$$\Delta g \approx \Delta g_0 + x \frac{\partial \Delta g}{\partial x} + y \frac{\partial \Delta g}{\partial y} = \Delta g_0 + R \sin \psi \left( \cos \alpha \frac{\partial \Delta g}{\partial x} + \sin \alpha \frac{\partial \Delta g}{\partial y} \right),$$

and substitute:

$$\xi_{\text{int}} \approx \frac{1}{4\pi \gamma} \int_0^{\psi_0} \int_0^{2\pi} \left( -\frac{2}{\psi^2} \right) \Delta g_0 + R \sin \psi \left( \cos \alpha \frac{\partial \Delta g}{\partial x} + \sin \alpha \frac{\partial \Delta g}{\partial y} \right) \cdot \cos \alpha \sin \psi d\alpha d\psi,$$

$$\eta_{\text{int}} \approx \frac{1}{4\pi \gamma} \int_0^{\psi_0} \int_0^{2\pi} \left( -\frac{2}{\psi^2} \right) \Delta g_0 + R \sin \psi \left( \cos \alpha \frac{\partial \Delta g}{\partial x} + \sin \alpha \frac{\partial \Delta g}{\partial y} \right) \cdot \sin \alpha \sin \psi d\alpha d\psi.$$

Here, the terms in $\Delta g_0$ drop out in $\alpha$ integration (because $\int_0^{2\pi} \sin \alpha = 0$), as do the mixed terms in $\sin \alpha \cos \alpha$. What remains is

$$\xi_{\text{int}} \approx -\frac{1}{4\pi \gamma} \int_0^{\psi_0} \int_0^{2\pi} \frac{2}{\psi^2} R \sin \psi \cos \alpha \frac{\partial \Delta g}{\partial x} \cos \alpha \cdot \sin \psi d\alpha d\psi \approx -\frac{R}{2\pi \gamma} \int_0^{\psi_0} \frac{\partial \Delta g}{\partial x} \sin \alpha \cdot \sin \psi d\psi,$$

$$\eta_{\text{int}} \approx -\frac{1}{4\pi \gamma} \int_0^{\psi_0} \int_0^{2\pi} \frac{2}{\psi^2} R \sin \psi \sin \alpha \frac{\partial \Delta g}{\partial y} \sin \alpha \cdot \sin \psi d\alpha d\psi \approx -\frac{R}{2\pi \gamma} \int_0^{\psi_0} \frac{\partial \Delta g}{\partial y} \sin^2 \alpha \cdot \sin \psi d\psi.$$

Executing the final $\psi$ integration yields now, with $R\psi_0 = d$:

$$\xi_{\text{int}} = -\frac{d}{2\gamma} \frac{\partial \Delta g}{\partial x}, \quad \eta_{\text{int}} = -\frac{d}{2\gamma} \frac{\partial \Delta g}{\partial y}.$$

Sometimes these equations are useful, e.g., when estimating the errors of grid based methods. Let the mesh size of a grid be $\Delta x$; we may set in the above equations $d \approx \Delta x/2$, and for $\Delta g$ we take

$$\Delta g^\text{obs} - \Delta g^\text{grid},$$

where $\Delta g^\text{grid}$ is the gravity anomaly value interpolated from the grid file at the evaluation point. In this way one obtains a rough estimate of how much error is due to the mesh size of the grid.
8. Spectral techniques, FFT

8.1 The Stokes equation as a convolution

We start from the Stokes equation

\[ T (\phi, \lambda) = \frac{R}{4\pi} \int_{\sigma} S (\psi) \Delta g (\phi', \lambda') \, d\sigma', \]

where \((\phi', \lambda')\) is the location of the moving point (integration or data point) on the Earth surface, and \((\phi, \lambda)\) the location of the evaluation point, it too on the Earth surface. Generally the locations of both points are given in spherical co-ordinates \((\phi, \lambda)\), and correspondingly the integration is executed over the surface of the unit sphere \(\sigma\): a surface element is \(d\sigma = \cos \phi \, d\phi \, d\lambda\), where the factor \(\cos \phi\) represents the determinant of Jacobi, or Jacobian, for these spherical co-ordinates \((\phi, \lambda)\).

However locally, in a sufficiently small area, one may write the point co-ordinates also in rectangular form, and then also the integral in rectangular co-ordinates. Suitable rectangular co-ordinates are, e.g., map projection co-ordinates, see figure 8.1.

A simple example of rectangular co-ordinates in the tangent plane would be

\[ x = \psi R \sin \alpha, \]
\[ y = \psi R \cos \alpha, \tag{8.1} \]

where \(\alpha\) is the azimuth between evaluation and moving point. The centre of this projection is the point where the tangent plane touches the sphere. The location of other points is measured by the angle at the Earth’s centre, \(\psi\), i.e., the spherical distance, and by the direction angle in the tangent plane or azimuth \(\alpha\).
A more realistic example uses a popular conformal map projection, the stereographic projection:

\[
\begin{align*}
    x &= 2 \sin \left( \frac{\psi}{2} \right) R \sin \alpha, \\
    y &= 2 \sin \left( \frac{\psi}{2} \right) R \cos \alpha.
\end{align*}
\]

In the limit for small values of \( \psi \) this is the same as equations (8.1).

Taking the squares of equations (8.1), summing them, and dividing them by \( R^2 \) yields

\[
\psi^2 \approx \frac{x^2 + y^2}{R^2}.
\]

More generally \( \psi \) is the angular distance between the two points \((x, y)\) (evaluation point) and \((x', y')\) (integration or moving point) seen from the Earth’s centre, approximately

\[
\psi^2 \approx \left( \frac{x - x'}{R} \right)^2 + \left( \frac{y - y'}{R} \right)^2.
\]

Furthermore we must account for the Jacobi determinant of the projection:

\[
d\sigma = R^{-2} dxdy
\]

and the Stokes equation becomes now

\[
T(x, y) \approx \frac{1}{4\pi R} \int_{-\infty}^{\infty} S(x - x', y - y') \Delta g(x', y') \, dx' \, dy', \quad (8.2)
\]

a two-dimensional convolution integral\(^1\).

\(^1\)The integration extends from minus to plus infinity in both co-ordinates \( x \) and \( y \). This can only be kept physically realistic on a curved Earth if it is assumed that the kernel \( S \) is of bounded support. This is the case for the modified kernels discussed in section 7.8.
Convolutions have nice properties in Fourier theory. If we designate the Fourier transform with the symbol $F$, and convolution with the symbol $\circ$, we may abbreviate the above equation as follows:

$$T = \frac{1}{4\pi R} S \circ \Delta g,$$

and according to the convolution theorem ("Fourier transforms a convolution into a multiplication"):

$$F\{T\} = \frac{1}{4\pi R} F\{S\} F\{\Delta g\}.$$

This $(x,y)$ plane approximation works only, if the needed integration can be restricted to a local area where the curvature of the Earth surface may be neglected. This is possible thanks to the use of global spherical harmonic expansions, because these describe the global part of the variability of the Earth’s gravity field. After we have removed from the observed gravity anomalies $\Delta g$ the effect of the global spherical harmonic model (the “Remove” step) we may safely forget the effect of areas far removed from the evaluation point: after this removal, the anomaly field will contain only the remaining short wavelength parts, the effect of which cancels out at greater distances. Of course, once the integral has been computed and the local disturbing potential $T_{loc}$ has been obtained, we must remember to add to it again the $T_{glob}$ effect of the global spherical harmonic expansion to be calculated separately (the “Restore” step).

### 8.2 Integration by FFT

The Fourier transform needed for applying the convolution theorem is calculated as a discrete Fourier transform. For this purpose exists the highly efficient Fast Fourier Transform, FFT (e.g., Kakkuri, 1981 pp. 183–200). There are several formulations of the discrete Fourier transform to be found in the literature; it doesn’t really matter which one is chosen, as long as it is a compatible pair of a forward Fourier transform $F$ and an inverse Fourier transform $F^{-1}$.

In preparation for this we first compute a discrete grid representation of the function $\Delta g (x,y)$, a rectangular table of $\Delta g$ values on an equidistant $(x,y)$ grid of points. The values may be, e.g., the function values themselves at the
grid points:\n\[ \Delta g_{ij} = \Delta g (x_i, y_j), \]
where the co-ordinates of the grid points are:
\[ x_i = i\delta x, \quad i = -\frac{N}{2}, \ldots, \frac{N}{2} - 1, \]
\[ y_j = j\delta y, \quad j = -\frac{N}{2}, \ldots, \frac{N}{2} - 1, \]
for suitably chosen grid spacings \((\delta x, \delta y)\). The sequence of values of the subscripts \(i\) and \(j\) has been chosen so, that the centre of the area \((x = y = 0)\) is also in the centre of the constructed data table \(\Delta g_{ij}, (i = j = 0)\). The integer \(N\), assumed even, is here the grid size, assumed the same in both directions (this is not a necessary assumption).

Next, we do the same for the kernel function
\[ S (\psi) = S (x - x', y - y') = S (\Delta x, \Delta y), \]
i.e., we write
\[ S_{ij} = S (\Delta x_i, \Delta y_j), \]
where again \((N\) being the grid size):
\[ \Delta x_i = i\delta x, \quad i = -\frac{N}{2}, \ldots, \frac{N}{2} - 1, \]
\[ \Delta y_j = j\delta y, \quad j = -\frac{N}{2}, \ldots, \frac{N}{2} - 1. \]
Also here, the choice of the value sequences for the \(i\) and \(j\) subscripts is based in the wish to get the central peak at the origin of the \(S\) function – \(S (\psi) \to \infty\) when \((\Delta x, \Delta y) \to (0, 0)\) – placed at the centre of the grid of values \(S_{ij}\).\n
Next:
1. the grid representations \(\Delta g_{ij}\) and \(S_{ij}\) thus obtained of the functions \(S\) and \(\Delta g\) are transformed to the frequency domain – they become functions \(S_{uv}\) and \(G_{uv}\) of the two “frequencies”, the wave numbers \(u\) and \(v\) in the \(x\) and \(y\) directions. The spatial frequencies are \(\omega_u = u/L, \omega_v = v/L\), where \(L\) is the size of the area (assumed square).

\[ ^2\text{There exist alternatives to this. E.g., one could calculate for every grid point the average over a square cell surrounding the point.} \]
\[ ^3\text{Without this measure the result of the calculation would be correct, but in the wrong place...} \]
2. They are multiplied with each other “one frequency pair at a time”, i.e., we calculate

\[ T_{uv} = S_{uv} G_{uv}, \quad u, v = 0 \ldots N - 1, \]

3. and we transform the result, \( T = \mathcal{F}\{T\}\), back to the space domain:

\[ T = \mathcal{F}^{-1}\{T\}, \]

i.e., to a grid \( T_{ij} = T(x_i, y_j) \) describing the disturbing potential \( T \). The disturbing potential of an arbitrary point can be obtained from this grid by interpolation. The co-ordinates \( x_i, y_j \) run as functions of \( i,j \) in the same way as described above for \( \Delta g^4 \).

This method is good for computing the disturbing potential \( T \) – and similarly the geoid height \( N = T/\gamma \) – from gravity anomalies using the Stokes equation. It is just as good for evaluating other quantities, like, e.g., the vertical gradient of gravity by the Poisson equation. The only requirement is, that the equation can be expressed as a convolution.

Also the reverse computation is easy: in the Fourier or spectral domain it is just a simple division.

Using the discrete FFT transform requires that the input data in a field to be integrated – in the example, gravity anomalies – is given on a regular grid covering the area of computation, or can be converted to one. The result – e.g., the disturbing potential – is obtained on a regular grid in the same geometry. Values can then be interpolated to chosen locations.

The FFT method may again be depicted as a commutative diagram:

<table>
<thead>
<tr>
<th>Free observation point selection</th>
<th>( \Rightarrow ) Interpolation ( \Rightarrow ) Point grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \downarrow )</td>
<td>( \downarrow ) FFT ( \downarrow )</td>
</tr>
<tr>
<td>(Direct solution) ( \downarrow )</td>
<td>Solution points in their own places ( \Leftrightarrow ) Interpolation ( \Leftrightarrow ) Point grid</td>
</tr>
</tbody>
</table>

Appendix C offers a short explanation of why FFT works and what makes it as efficient as it is.

4In fact, for both \( \Delta g \) and \( T \) we could choose the simpler grid geometry where the subscript sequences are \( i,j = 0, \cdots, N - 1 \); however, for \( S \) it is mandatory to have the origin in the middle of the grid.
8.3 Solution in rectangular co-ordinates

In the above equation (8.2), the co-ordinates $x$ and $y$ are rectangular. In practice, often latitude and longitude $(\phi, \lambda)$ are taken, causing additional errors due to meridian convergence – as a latitude and longitude co-ordinate system isn’t actually rectangular. Slightly more suitable would be the pair $(\phi, \lambda \cos \phi)$.

The problem has also been addressed on a more conceptual level.

8.3.1 The Strang van Hees method

The Stokes kernel function $S(\psi)$ depends only on the angular distance $\psi$ between evaluation point $(\phi, \lambda)$ and the data point $(\phi', \lambda')$. The angular distance may be written as follows (spherical approximation):

$$\cos \psi = \sin \phi' \sin \phi + \cos \phi' \cos \phi \cos (\lambda' - \lambda).$$

Substitute

$$\cos (\lambda' - \lambda) = 1 - 2 \sin^2 \left(\frac{\lambda' - \lambda}{2}\right),$$

$$\cos \psi = 1 - 2 \sin^2 \frac{\psi}{2},$$

$$\cos (\phi' - \phi) = 1 - 2 \sin^2 \frac{\phi' - \phi}{2},$$

and obtain

$$\cos \psi = \cos (\phi' - \phi) - 2 \cos \phi' \cos \phi \sin^2 \frac{\lambda' - \lambda}{2} \Rightarrow$$

$$\sin^2 \frac{\psi}{2} = \sin^2 \frac{\phi' - \phi}{2} + \cos \phi' \cos \phi \sin^2 \frac{\lambda' - \lambda}{2}. $$

Here we may use the following approximation:

$$\cos \phi' \approx \cos \phi \equiv \cos \phi_0,$$

where $\phi_0$ is a reference value in the middle of the calculation area. Now the above formula becomes:

$$\sin^2 \frac{\psi}{2} \approx \sin^2 \frac{\phi' - \phi}{2} + \cos^2 \phi_0 \sin^2 \frac{\lambda' - \lambda}{2}, \quad (8.3)$$

which depends only on the differences $\Delta \phi \equiv \phi' - \phi$ and $\Delta \lambda \equiv \lambda' - \lambda$; a requirement for convolution.
After this, the FFT method may be applied by using co-ordinates $\phi, \lambda$ and the modified Stokes function
\[
S^*(\Delta\phi, \Delta\lambda) = S \left( 2 \sqrt{\arcsin \left( \frac{\sin^2 \frac{\Delta\phi}{2} + \cos^2 \phi_0 \sin^2 \frac{\Delta\lambda}{2}}{2} \right)} \right).
\]

This clever way of using FFT in geographic co-ordinates was invented by the Dutchman G. Strang van Hees in 1990.

### 8.3.2 “Spherical FFT”, multi-band model

We divide the area in several narrow latitude bands. In each band we apply the Strang van Hees method using its own optimal central latitude.

Write the Stokes equation as follows:
\[
N(\phi, \lambda) = \frac{R}{4\pi \gamma} \int \int S(\phi - \phi', \lambda - \lambda'; \phi) \left[ \Delta g(\phi', \lambda') \cos \phi' \right] d\phi' d\lambda', \quad (8.4)
\]
where we have expressed $S(\cdot)$ as a function of latitude difference, longitude difference and evaluation latitude. Now, choose two support latitudes: $\phi_i$ ja $\phi_{i+1}$. Assume furthermore that $S$ is between these a linear function of $\phi$. In that case we may write:
\[
S(\Delta\phi, \Delta\lambda, \phi) = \frac{(\phi - \phi_i) S_{i+1}(\Delta\phi, \Delta\lambda) + (\phi_{i+1} - \phi) S_i(\Delta\phi, \Delta\lambda)}{\phi_{i+1} - \phi_i},
\]
where $\Delta\phi = \phi - \phi'$, $\Delta\lambda = \lambda - \lambda'$ and
\[
S_i(\Delta\phi, \Delta\lambda) = S(\phi - \phi', \lambda - \lambda'; \phi_i), \\
S_{i+1}(\Delta\phi, \Delta\lambda) = S(\phi - \phi', \lambda - \lambda'; \phi_{i+1}).
\]

We obtain by substitution into integral equation (8.4):
\[
N(\phi, \lambda) = \frac{R}{4\pi \gamma} \left\{ \left[ \frac{\phi - \phi_i}{\phi_{i+1} - \phi_i} \right] \int \int S_{i+1}(\Delta\phi, \Delta\lambda) \left[ \Delta g(\phi', \lambda') \cos \phi' \right] d\phi' d\lambda' + \right. \\
+ \left. \left[ \frac{\phi_{i+1} - \phi}{\phi_{i+1} - \phi_i} \right] \int \int S_i(\Delta\phi, \Delta\lambda) \left[ \Delta g(\phi', \lambda') \cos \phi' \right] d\phi' d\lambda' \right\}. \quad (8.5)
\]

This equation is the sum of two convolutions. Both are evaluated by FFT and from the solutions obtained one forms the weighted mean according to equation (8.5).

---

5 In practice one uses the geodetic/geographic latitude $\phi$ instead of $\phi$ without significant error.

6 Govert L. Strang van Hees (1932–2012) was a Dutch gravimetric geodesist.
In this method we may use, instead of the approximative equation (8.3), an exact equation, in which \( \phi' \) is expressed into \( \phi \) and \( \Delta \phi \):

\[
\sin^2 \frac{\psi}{2} = \sin^2 \frac{\phi' - \phi}{2} + \cos \phi' \cos \phi \sin^2 \frac{\lambda' - \lambda}{2} = \\
= \sin^2 \frac{\Delta \phi}{2} + \cos (\phi - \Delta \phi) \cos \phi \sin^2 \frac{\Delta \lambda}{2}.
\]

Here again, we calculate \( S_i \) and \( S_{i+1} \) for the values \( \phi = \phi_i \) and \( \phi = \phi_{i+1} \), we evaluate the integrals with the aid of the convolution theorem, and interpolate \( N(\phi, \lambda) \) according to equation (8.5) when \( \phi_i \leq \phi < \phi_{i+1} \). After this, the solution isn’t entirely exact, because inside every band we still use linear interpolation. However by making the bands narrower, we can keep the error arbitrarily small.

### 8.3.3 “Spherical FFT”, Taylor expansion model

This somewhat more complicated but also more versatile approach expands the Stokes kernel into a Taylor expansion with respect to latitude about a reference latitude located in the middle of the computation area\(^7\). Each term in the expansion depends only on the difference in latitude. The integral to be calculated similarly expands into terms, of which each contains a pure convolution.

Let us write the general problem as follows:

\[
\ell (\varphi, \lambda) = \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} C (\varphi, q', \Delta \lambda) \left[ m (q', \lambda') \cos q' \right] d\varphi' d\lambda',
\]

where \( \ell \) contains values to be computed, \( m \) contains values given, and \( C \) is the coefficient or kernel function. Here is assumed only rotational symmetry for the geometry, i.e., the kernel function depends only on the difference between longitudes \( \Delta \lambda \) rather than the absolute longitudes \( \lambda, \lambda' \).

In a concrete case \( m \) contains for example \( \Delta g \) values in various points \((\varphi', \lambda')\), \( \ell \) contains geoid heights \( N \) in various points \((\varphi, \lambda)\), and \( C \) contains coefficients calculated using the Stokes kernel function.

We first change the dependence upon \( \varphi, q' \) into a dependence upon \( \varphi, \Delta \varphi \):

\[
C = C (\varphi, q', \Delta \lambda) = C (\Delta \varphi, \Delta \lambda, \varphi).
\]

Linearize:

\[
C = C_0 (\Delta \varphi, \Delta \lambda) + (\varphi - \varphi_0) C_{\varphi} (\Delta \varphi, \Delta \lambda) + \ldots
\]

\(^7\)In the literature the method has been generalized by expanding the kernel also with respect to height.
where we define for a suitable reference latitude \( \varphi_0 \):

\[
C_0 (\Delta \varphi, \Delta \lambda) \doteq C (\Delta \varphi, \Delta \lambda, \varphi_0), \\
C_\varphi (\Delta \varphi, \Delta \lambda) \doteq \frac{\partial}{\partial \varphi} C (\Delta \varphi, \Delta \lambda, \varphi) \bigg|_{\varphi = \varphi_0}.
\]

Substitution yields

\[
\ell = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C m' \cos \varphi' d\varphi' d\lambda' = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \left[ C_0 + (\varphi - \varphi') C_\varphi \right] m' \cos \varphi' d\varphi' d\lambda' = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C_0 m' \cos \varphi' d\varphi' d\lambda' + (\varphi - \varphi_0) \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C_\varphi m' \cos \varphi' d\varphi' d\lambda'.
\]

(8.6)

**Important** here is now, that the integrals in the first and second terms, i.e.,

\[
\int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C_0 m' \cos \varphi' d\varphi' d\lambda' = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C_0 (\Delta \varphi, \Delta \lambda) \left[ m' \cos \varphi' \right] d\varphi' d\lambda' \doteq \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C_0 \circ [m \cos \varphi]
\]

ja

\[
\int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C_\varphi \left[ m' \cos \varphi' \right] d\varphi' d\lambda' = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C_\varphi (\Delta \varphi, \Delta \lambda) \left[ m' \cos \varphi' \right] d\varphi' d\lambda' \doteq \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} C_\varphi \circ [m \cos \varphi]
\]

are both convolutions: both \( C \) functions depend only on \( \Delta \varphi \) and \( \Delta \lambda \). Both integrals can be calculated only if the corresponding \( \Delta \varphi = \varphi - \varphi' \) and \( \Delta \lambda = \lambda - \lambda' \), and the corresponding coefficient grids \( C_0, C_\varphi \), are calculated first. After this (in principle expensive, but, thanks to FFT and the convolution theorem, a lot cheaper) integration, computing the compound (8.6) is cheap: one multiplication and one addition for each evaluation point \((\varphi, \lambda)\).

**Example:** let the evaluation area at latitude \( 60^\circ \) be \( 10^\circ \times 20^\circ \) in size. If the grid mesh size is \( 5' \times 10' \), the number of cells is \( 120 \times 120 \). Let us choose, e.g., a \( 256 \times 256 \) grid (i.e., \( N = 256 \)) and fill the missing values with extrapolated values.

Also the values of the kernel functions \( C_0 \) and \( C_\varphi \) are calculated on a \( 256 \times 256 \) size \((\Delta \varphi, \Delta \lambda)\) grid. The number of these is thus also \( 65,536 \). Calculating the convolutions \( C_0 \circ [m \cos \varphi] \) and \( C_\varphi \circ [m \cos \varphi] \) by means of FFT – i.e.,

\[
\int \int C_0 m' \cos \varphi' d\varphi' d\lambda' = C_0 \circ [m \cos \varphi] = \mathcal{F}^{-1} \{ \mathcal{F} \{ C_0 \} \mathcal{F} \{ m \cos \varphi \} \},
\]

\[
\int \int C_\varphi m' \cos \varphi' d\varphi' d\lambda' = C_\varphi \circ [m \cos \varphi] = \mathcal{F}^{-1} \{ \mathcal{F} \{ C_\varphi \} \mathcal{F} \{ m \cos \varphi \} \},
\]
requires \((N^2) \times 2 \log (N^2) = 65,536 \times 16\) = more than a million operations, multiplication with \((\phi - \phi_0)\) and adding together, each again 65,536 operations.

The grid matrices corresponding to functions \(C_0\) and \(C_\phi\) are obtained as follows: for three reference latitudes \(\phi_-, \phi_0, \phi_+\) we compute numerically the grids

\[
C_\phi = C(\Delta \phi, \Delta \lambda, \phi_0),
\]

\[
C_m = C(\Delta \phi, \Delta \lambda, \phi_0),
\]

\[
C_\phi = C(\Delta \phi, \Delta \lambda, \phi_+),
\]

where \(C_0\) is directly available, and

\[
C_\phi \approx \frac{C_+ - C_-}{\phi_+ - \phi_-}.
\]

Also inversion is thus directly feasible. Let \(\ell\) be given in suitable point grid form. We compute the first approximation to \(m\) as follows:

\[
\mathcal{F}\{C_0\} \mathcal{F}\{m \cos \phi\} = \mathcal{F}\{\ell\} \Rightarrow [m \cos \phi]^{(0)} = \mathcal{F}^{-1}\left\{ \frac{\mathcal{F}\{\ell\}}{\mathcal{F}\{C_0\}} \right\}.
\]

The second approximation is obtained by first calculating

\[
\ell^{(0)} = C_0 \circ [m \cos \phi]^{(0)} + (\phi - \phi_0) \cdot C_\phi \circ [m \cos \phi]^{(0)},
\]

after which we make the improvement:

\[
[m \cos \phi]^{(1)} = [m \cos \phi]^{(0)} + \mathcal{F}^{-1}\left\{ \mathcal{F}\left\{ \ell - \ell^{(0)} \right\} \mathcal{F}\{C_0\} \right\},
\]

and so on, iteratively. Two, three iterations are usually enough. This method has been used to compute underground mass points to represent gravity anomalies in the exterior gravity field of the Earth. More is explained in Forsberg and Vermeer (1992).

8.3.4 “1D-FFT”

This is a limiting case of the previous ones, where FFT is used only in the longitude direction. In other words, a zones method where the zones are only one point narrow. This method is exact if all longitudes \((-360^\circ - 360^\circ)\) are along in the calculation. It requires a bit more computing time compared to the previous methods. In fact, it is identical to a Fourier transform in variable \(\lambda\), longitude. Details are found in Haagmans et al. (1993).
8.4 Complications; bordering, tapering

The discrete Fourier transform presupposes the data to be *periodically continuous*. In practice it is not. Therefore, always when using FFT with the convolution theorem (8.2), we continue the data by adding a border area to the data area, so-called *bordering*. Often the border area is 25% of the size of the data area; then, the size of the whole calculation area will become four times that of the data area itself. The border is often filled with zeroes, although predicted values – or even measured values, if those exist – are a better choice.

Also the calculation area for the *kernel function* is made similarly four times larger; in this case, as the function is *symmetric*, the border area is filled with real (computable) values, making it automatically periodically continuous.

Because the discrete Fourier transform assumes periodicity, one should make sure that the data really is periodic. If the values at the borders are not zero, they may be forced to zero by multiplying the whole data area by a so-called *tapering* function, which goes smoothly to zero towards the edges. Such a function can easily be built, e.g., a cubic spline polynomial or cosine. See figure 8.2, showing a 25% tapering function, as well as example images 8.3, where one sees how non-periodicity (differing left and right, and upper and lower, edges) causes horizontal and vertical artefacts in the Fourier transform. These artefacts are called the *Gibbs phenomenon*, already mentioned in section 7.8: a sharp cut-off or edge in the space domain will produce signal in all frequencies up to the highest ones.

Many journal articles have appeared on these technicalities. Groups that have been involved in early development of FFT geoid computation (in the 1980s) are Forsberg’s group in Copenhagen, the group of Schwarz and Sideris in Calgary, Canada, the Delft group (Strang van Hees, Haagmans, De Min, Van Gelderen), the Milanese (Sansò, Barzaghi, Brovelli), the Hannover
Chapter 8. Spectral techniques, FFT

Figure 8.3. Example images for FFT transform without (above) and with (below) tapering. Used on-line FFT [http://www.ejectamenta.com/Imaging-Experiments/fourierimagefiltering.html].

“Institut für Erdmessung”, and many others.

8.5 Computing a geoid model with FFT

Nowadays computing a geoid or quasi-geoid model is easy thanks to increased computing power, especially using FFT. On the other hand the spread of precise geodetic satellite positioning has made the availability of precise geoid models an important issue, so that one can use GNSS (Global Navigation Satellite Systems) technology for rapid and inexpensive height determination.

8.5.1 The GRAVSOFT software

The GRAVSOFT geoid computation software has been mainly produced in Denmark. Authors include Carl Christian Tscherning, René Forsberg, Per Knudsen, and the Greek Dimitris Arabelos.

GRAVSOFT manual.
This package is in widespread use and offers, in addition to variants of FFT geoid computation, also, e.g., least-squares collocation, routines for evaluation of various terrain effects, etc. Its spread can be partly explained by it being free for scientific use, and being distributed as source code. It is also well documented. Therefore it has also found commercial use, e.g., in the petroleum extraction industry.

GRAVSOFT has been used a lot also for teaching, e.g., at many research schools organized by the IAG (International Association of Geodesy) in various countries. [http://www.isgeoid.polimi.it/Schools/schools.html](http://www.isgeoid.polimi.it/Schools/schools.html).

### 8.5.2 The Finnish FIN2000 geoid

Currently two geoid models are in use in Finland: FIN2000 (figure 8.4) and FIN2005Noo (Bilker-Koivula and Ollikainen, 2009). The first model is a ref-
Chapter 8. Spectral techniques, FFT

The geoid heights given by the model are above the GRS80 reference ellipsoid. The second model is similarly a reference surface for the new N2000 height system. It too gives heights from the GRS80 reference ellipsoid. The precisions (mean errors) of FIN2000 and FIN2005N00 are on the level of $\pm 2 - 3$ cm.

8.6 Use of FFT computation in other contexts

8.6.1 Altimetry

Per Knudsen and Ole Balthasar Andersen have computed a gravity map of the world ocean by starting from satellite altimetry derived “geoid heights” and inverting them to gravity anomalies (Andersen et al., 2010). A pioneer of the method has been David Sandwell.

8.6.2 Satellite gravity missions; airborne gravimetry

Also the data from satellite gravity missions (like CHAMP, GRACE and GOCE) can be regionally processed using the FFT method: in the case of GOCE, the inversion of gradiometric measurements, i.e., calculating geoid heights on the Earth’s surface from measurements at satellite level. Also airborne gravity measurements are processed in this way. The problem is called “downward continuation” and is in principle unstable.

Airborne gravimetry is a practical method for gravimetric mapping of large areas; in the pioneering days, the gravity field over Greenland was mapped, as well as many areas around the Arctic and Antarctic. Later, areas were measured like the Brazilian Amazonas, Mongolia and Ethiopia, where no full-coverage terrestrial gravimetric data existed. The advantage of this method is that one measures rapidly large areas in a homogeneous way. Also for the processing of airborne gravimetry data, FFT is suitable.

8.7 Computing terrain corrections with FFT

The terrain correction is a very localized phenomenon, the calculation of which requires high-resolution terrain data from a relatively small area sur-
Computing terrain corrections with FFT

8.7. Computing terrain corrections with FFT

rounding the computation point. Thus, calculating the terrain correction is ideally suited to applying the FFT method.

We show how, with FFT, we can simply and efficiently evaluate the terrain correction. We make the following simplifying assumptions:

1. terrain slopes are relatively small;
2. the density $\rho$ of the Earth’s crust is constant;
3. the Earth is flat.

These assumptions are not mandatory. The general case however leads us into a jungle of formulas without aiding the conceptual picture.

The terrain correction, the joint effect of all the topographic masses, or lacking topographic masses, above and below the height level $H$ of the evaluation point, can be calculated under these assumptions using the following rectangular equation, which describes the attraction of rock columns projected onto the vertical direction (figure 8.5):

\[
TC(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{G\rho (H' - H)}{\ell^2} \cos \theta \ dx'\,dy'
= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{G\rho (H' - H)}{\ell^2} \cdot \frac{1}{2} \frac{H' - H}{\ell} \ dx'\,dy'
= \frac{1}{2} G\rho \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{(H' - H)^2}{\ell^3} \ dx'\,dy'.
\]  

(8.7)

Here $G\rho (H' - H) \ell^{-2}$ is the attraction of the column and $\frac{1}{2} (H' - H) \ell^{-1}$ is the cosine of the angle $\theta$ between the force vector (assumed coming from the midpoint of the rock column) and the vertical direction.

We will make a linear approximation, wherein $\ell$, the slant distance between the evaluation point $(x, y)$ and the moving data point $(x', y')$, is also the...
horizontal distance:
\[ \ell^2 \approx (x' - x)^2 + (y' - y)^2. \]

Equation (8.7) is easy to check straight from Newton’s law of gravitation. When it is assumed that the terrain is relatively free of steep slopes, then \( \ell \) is large compared to \( H' - H \).

From the above we obtain by development into terms:
\[ TC (x, y) = \frac{1}{2} G \rho H^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\ell^3} dx' dy' - G \rho H \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{H'}{\ell^3} dx' dy' + \]
\[ \frac{1}{2} G \rho \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{(H')^2}{\ell^3} dx' dy', \] (8.8)

where every integral is a convolution with kernel \( \ell^{-3} \), and functions to be integrated 1, \( H' \) and \( (H')^2 \). Unfortunately the function \( \ell^{-3} \) as defined above has no Fourier transform, wherefore we change the above definition a tiny bit by adding a small term:
\[ \ell^2 = (x' - x)^2 + (y' - y)^2 + \delta^2. \] (8.9)

Then, the terms in the above sum are large numbers that almost cancel each other, giving a nearly correct result. However, numerically this is an unpleasant situation.

If \( \ell \) is defined according to equation (8.9), then the Fourier transform of kernel \( \ell^{-3} \) is (Harrison and Dickinson, 1989):
\[ \mathcal{F}\{\ell^{-3}\} = \frac{2\pi}{\delta} \exp(-2\pi\delta q) = \frac{2\pi}{\delta} \left[ 1 - 2\pi\delta q + \frac{4\pi^2q^2\delta^2}{1 \cdot 2} - \ldots \right], \]

where \( q = \sqrt{u^2 + v^2} \), and \( u, v \) are wave numbers (i.e., “frequencies”) in the \( x \) and \( y \) directions in the \( (x, y) \) plane. If we substitute this into equation (8.8), we notice that the terms containing \( \delta^{-1} \) sum to zero, and of course also the terms containing positive powers of \( \delta \) vanish when \( \delta \to 0 \). As follows (Harrison and Dickinson, 1989):
\[ \mathcal{F}\{TC\} \approx \frac{1}{2} G \rho H^2 \mathcal{F}\{1\} \left[ \frac{2\pi}{\delta} (1 - 2\pi\delta q) \right] - \]
\[ G \rho H \mathcal{F}\{H'\} \left[ \frac{2\pi}{\delta} (1 - 2\pi\delta q) \right] + \]
\[ \frac{1}{2} G \rho \mathcal{F}\{(H')^2\} \left[ \frac{2\pi}{\delta} (1 - 2\pi\delta q) \right] \]

where we leave off all terms in higher powers of \( \delta \).

Re-order the terms:
\[ \mathcal{F}\{TC\} = \frac{\pi}{\delta} G \rho \left[ H^2 \mathcal{F}\{1\} - 2HF \mathcal{F}\{H'\} + \mathcal{F}\{(H')^2\} \right] + \]
\[ 4\pi^2 q \left[ -\frac{1}{2} G \rho H^2 \mathcal{F}\{1\} + G \rho H \mathcal{F}\{H'\} - \frac{1}{2} G \rho \mathcal{F}\{(H')^2\} \right]. \]
Because $F\{1\} = 0$ if $q \neq 0$, the first term inside the second term will always vanish. We obtain (remember that $H$ is a constant, the height of the evaluation point):

$$F\{TC\} = \frac{\pi}{\delta}G\rho \left[F\left\{H^2 - 2HH' + (H')^2 \right\} + 4\pi^2 q \left[G\rho H F\left\{H'\right\} - \frac{1}{2} \left\{F\left\{(H')^2\right\} \right\} \right]\right]$$

and the inverse Fourier transform yields:

$$TC = \frac{2\pi G\rho}{\delta} \left[\frac{1}{2}H^2 - H'H + \frac{1}{2} (H')^2 \right] + \left[G\rho H F^{-1}\left\{F\left\{H'\right\} \cdot 4\pi^2 q \right\} - \frac{1}{2} \left\{F\left\{(H')^2\right\} \cdot 4\pi^2 q \right\} \right].$$

In the first term

$$\frac{1}{2}H^2 - H'H + \frac{1}{2} (H')^2 = \frac{1}{2} (H' - H)^2 = 0$$

in point $(x, y)$ where $H' = H$, and we obtain:

$$TC = 4\pi^2 G\rho F^{-1}\left\{q \cdot \left[H F\left\{H'\right\} - \frac{1}{2} \left\{F\left\{(H')^2\right\}\right\} \right]\right\},$$

from which now the troublesome $\delta^{-1}$ has vanished.

A condition for this “regularization” or “renormalization” is, that at point $(x, y)$ $H' = H$, i.e., the evaluation happens at the Earth’s surface. The convolutions above are evaluated by the FFT technique; a more detailed account is found, e.g., in the article Vermeer (1992).

For calculating the terrain effect $TC$ in the exterior space – airborne gravimetry, but also the effect of the sea floor at the sea surface, or the effect of the Mohorovičić discontinuity at the Earth’s surface – there are techniques that express $TC$ as a sum of convolutions, as a Taylor expansion. An early paper on this is Parker (1972).
Chapter 8. Spectral techniques, FFT
9. Statistical methods

9.1 The role of uncertainty in geophysics

In geophysics, we often act, or obtain results, based on uncertain or deficient observational data. This applies also in the study of the Earth’s gravity field: e.g., the density of gravity observations on the Earth surface varies a lot, and large areas on the oceans and polar regions are covered only by a very sparse network of measurements.

Measurement technologies that work from space provide coverage of the whole globe, oceans, poles and all. They however don’t again measure at a very high resolution. Either the resolution of the instrument is limited (e.g., the gravity field parameters calculated from satellite orbit perturbations) or the instruments measure only straight underneath the satellite’s path (e.g., satellite altimetry).

Another often relevant uncertainty factor is, that one can do precise measurements on the Earth surface, but inside the Earth the uncertainty is much larger and the data is obtained much more indirectly.

In previous chapters we described techniques by which we could calculate desired values or parameters for the Earth’s gravity field, assuming that, e.g., gravity anomalies are available everywhere on the Earth surface, and with an unlimitedly high resolution. In this chapter we look at mathematical means to handle real-life situations where this is not the case.
9.2 Linear functionals

In mathematics, an operator that associates with every function in a given function space a certain numerical value is called a functional. One such is, e.g., a (partial) derivative in a certain point:

\[ f \mapsto \frac{\partial}{\partial x} f(x) \bigg|_{x=x_0}. \]

Other functionals are, e.g., the integral over a given area:

\[ f \mapsto \int_{\sigma} f(x) \, dx, \]

and so on.

We may write symbolically:

\[ L = \frac{\partial}{\partial x} \bigg|_{x=x_0}, \text{ i.e., } L(f) = \frac{\partial f}{\partial x} \bigg|_{x=x_0}. \]

A functional or operator is linear if

\[ L(\alpha f + \beta g) = \alpha L(f) + \beta L(g), \forall \alpha, \beta \in \mathbb{R}. \]

Note that all partial derivatives, as also the Laplace operator \( \Delta \), are linear.

In physical geodesy, all interesting functionals are functionals of the function \( T(\phi, \lambda, R) = T(\phi, \lambda, r)|_{r=R}, \text{ i.e., the disturbing potential on the Earth's surface. The theory thus uses the spherical approximation}, \text{ and the surface of the sphere, radius } R, \text{ corresponds to mean sea level. E.g., the disturbing potential } T_P(\phi, \lambda, R) \text{ in a point } P \text{ at sea-level location } (\phi, \lambda) \text{ is such a functional:} \]

\[ T(\cdot, \cdot, R) \mapsto T(\phi, \lambda, R). \]

Even if point \( P \) were not at sea level:

\[ T(\cdot, \cdot, R) \mapsto T(\phi, \lambda, r). \]

Even if the quantity were not the disturbing potential, but, e.g., the gravity anomaly or the deflection of the plumbline:

\[ T(\cdot, \cdot, R) \mapsto \Delta g(\phi, \lambda, r), \]

\[ T(\cdot, \cdot, R) \mapsto \zeta(\phi, \lambda, r), \]

\[ T(\cdot, \cdot, R) \mapsto \eta(\phi, \lambda, r). \]

\[ ^1 \text{This is not mandatory, but the error of approximation is small.} \]
All these are also linear functionals. In fact, if we write

\[ T = \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sum_{m=0}^{n} P_n \left( \sin \phi \right) (a_{nm} \cos m\lambda + b_{nm} \sin \lambda), \]

then even the spherical harmonic coefficients \(a_{nm}, b_{nm}\) are all linear functionals of the disturbing potential \(T\):

\[ T(\cdot, \cdot, R) \mapsto a_{nm}, \]
\[ T(\cdot, \cdot, R) \mapsto b_{nm}. \]

Here, \(T(\cdot, \cdot, R)\) is shorthand for the whole function

\[ T(\phi, \lambda, R), \phi \in [-\pi/2, +\pi/2], \lambda \in [0, 2\pi), \]

for which we’ll also use the notation \(T_R\).

### 9.3 Statistics on the Earth surface

In statistics we define a stochastic process as a stochastic quantity (random variable) the value space or domain of which is a function space, i.e., the realization values of which are functions. A stochastic process may be a quantity developing over time, the precise behaviour of which is uncertain, e.g., a satellite orbit. In the same way as for a (real-valued) stochastic quantity \(x\) we may calculate an expected value or expectancy \(E\{x\}\) and a variance \(C_{xx} = \text{Var}\{x\} = E\left\{\left[x - E\{x\}\right]^2\right\}\), we may also do so for a stochastic process. The only difference is, that by doing so we obtain a function.

Let, e.g., the stochastic process \(x(t)\) be a function of time. Then we may compute its variance function as follows:

\[ C_{xx}(t) = \text{Var}\{x(t)\}. \]

For a stochastic process however, much more can be computed: e.g., the covariance of values of the same function taken at different points in time, the so-called autocovariance:

\[ A_{xx}(t_1, t_2) = \text{Cov}\{x(t_1), x(t_2)\} = E \left\{ [x(t_1) - E\{x(t_1)\}] [x(t_2) - E\{x(t_2)\}] \right\}. \]

Similarly if we have two different functions, we may compute between them the so-called cross covariance, etc.
Chapter 9. Statistical methods

The argument of a stochastic process is commonly time, $t$. However in geophysics we study stochastic processes the arguments of which are the location on the Earth surface, i.e., we talk of processes of the form $x(\phi, \lambda)$. The definition of auto- and cross-covariances works otherwise in the same way, but in case of the Earth we have a special problem. A stochastic quantity is generally defined as a quantity $x$, from which realizations $x_1, x_2, x_3, \ldots$ are obtained, which have certain statistical properties. The classical example is the dice throw. A die can be thrown again and again, and one can practice the art of statistics on the results of the throws. Another classic example is measurement. Measurement of the same quantity can be repeated, and is repeated, in order to improve precision.

For a stochastic process defined on the Earth’s surface, the situation is different.

We have only one Earth.

For this reason, statistics must be done in a somewhat different fashion.

Given a stochastic process on the surface of the Earth, $x(\phi, \lambda)$, we define a quantity similar to the statistical expectancy $E\{\cdot\}$, the geographic mean

$$M \{x\} = \frac{1}{4\pi} \int_{0}^{2\pi} x(\phi, \lambda) d\phi = \frac{1}{4\pi} \int_{-\pi/2}^{\pi/2} x(\phi, \lambda) \cos \phi d\phi d\lambda. \quad (9.1)$$

Here $x(\phi, \lambda)$ is the one and only realization of process $x$ that we have available on this Earth.

Clearly this definition makes sense only in the case where the statistical behaviour of the process $x(\phi, \lambda)$ is the same everywhere on Earth, independently of the value of $(\phi, \lambda)$. This is called the assumption of homogeneity. It is in fact the assumption that the spherical symmetry of the Earth extends to the statistical behaviour of her gravity field.

Similarly to the expectancy, we may define the (geographic) variance:

$$C_{xx}(\phi, \lambda) = \text{Var}\{x(\phi, \lambda)\} = M \{[x - M \{x\}]^2\}. \quad (9.2)$$

The global average of gravity anomalies $\Delta g(\phi, \lambda)$ vanishes based on their definition, i.e.

$$M \{\Delta g\} = 0.$$

In that case, equation (9.2) is simplified as follows:

$$C_{\Delta g \Delta g}(\phi, \lambda) = \text{Var}\{\Delta g(\phi, \lambda)\} = M \{\Delta g^2\} = \frac{1}{4\pi} \int_{\sigma} [\Delta g(\phi, \lambda)]^2 d\sigma.$$
The definition given here of the geographic mean $M \{ \cdot \}$ is based on integration over the possible states of a system. As has been seen, in statistics, the mean is defined in a slightly different way, as the expectancy of a stochastic process. For gravity anomalies in this case $E \{ \Delta g \}$, where $\Delta g$ is the anomaly considered as a stochastic process, i.e., the series of values of $\Delta g$ that results if we look at an infinitely long series of randomly formed Earths. Not very practical!

If the expectancy of a stochastic process is the same as the mean of one realization computed by integration, we speak of an ergodic process. Establishing empirically that a process is ergodic is in geophysics typically difficult to impossible.

### 9.4 The covariance function of the gravity field

Defining a covariance function between points $P$ and $Q$ is more complicated. Something like equations (9.1), (9.2) cannot be used directly, because both $\Delta g_P$ and $\Delta g_Q$ can move over the whole Earth’s surface. We have

\[
\Delta g_P = \Delta g (\phi_P, \lambda_P), \\
\Delta g_Q = \Delta g (\phi_Q, \lambda_Q).
\]

In the following we assume that the covariance to be calculated will only depend on the relative location of points $P$ and $Q$. In a homogeneous gravity field, the covariance function will not depend on the absolute location of the points, but only on the difference in location between points $P$ and $Q$.

Write

\[
\phi_Q = \phi_Q (\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}), \\
\lambda_Q = \lambda_Q (\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}).
\]

$(\phi_Q, \lambda_Q)$ can be computed\(^2\), if we know $(\phi_P, \lambda_P)$ and both the angular distance $\psi_{PQ}$ and the azimuth angle $\alpha_{PQ}$. See figure 9.1.

Now we may write

\[
\Delta g_Q = \Delta g_Q (\phi_Q (\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}), \lambda_Q (\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ})) = \\
= \Delta g_Q (\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}),
\]

---

\(^2\)This is called the geodetic forward problem on the sphere.
Figure 9.1. Definition of geocentric angular distance and azimuth.

and we may define as the covariance function:

$$ C_{\Delta g \Delta g} (\psi_{PQ}, \alpha_{PQ}) = M \{ \Delta g_P (\phi_P, \lambda_P) \Delta g_Q (\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}) \} = 
\frac{1}{4\pi} \int_\sigma \Delta g_P (\phi_P, \lambda_P) \Delta g_Q (\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}) \, d\sigma_P. $$

Also here, $M$ is a geographical averaging operator. First we fix point $Q$ in relation to point $P$: both azimuth $\alpha_{PQ}$ and distance $\psi_{PQ}$ are held fixed. The point $P$ – and with it, point $Q$ – are moved over the whole Earth’s surface. We compute the corresponding integral over the unit sphere and divide by $4\pi$:

$$ C_{\Delta g \Delta g} (\psi_{PQ}, \alpha_{PQ}) = M \{ \Delta g_P \Delta g_Q (P) \} = \frac{1}{4\pi} \int_\sigma \Delta g_P \Delta g_Q (P) \, d\sigma_P = 
\frac{1}{4\pi} \int_{-\pi/2}^{+\pi/2} \int_0^{2\pi} \Delta g_P \Delta g_Q (P) \, d\lambda_P \cos \phi_P d\phi_P. $$

In addition to the assumption of homogeneity, we may make still the assumption of isotropy: the covariance function – or more generally, the statistical behaviour of the gravity field – does not depend on the relative direction $\alpha_{PQ}$ of point pair $(P, Q)$, but only on the angular distance $\psi_{PQ}$ between them. (This too is, like homogeneity, one of the forms in which the Earth’s rotational symmetry is expressed.) In this case we may compute the geographic mean in a slightly different way, by also averaging over all azimuth angles $\alpha_{PQ} \in [0, 2\pi)$:

$$ C_{\Delta g \Delta g} (\psi_{PQ}) = M' \{ \Delta g_P \Delta g_Q (P) \} = \frac{1}{2\pi} \int_0^{2\pi} M \{ \Delta g_P \Delta g_Q (P) \} \, d\alpha_{PQ} = 
\frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} \Delta g_P \Delta g_Q (P) \cos \phi_P d\phi_P d\lambda_P d\alpha_{PQ}. \quad (9.3) $$

Remark. The true gravity field of the Earth isn’t terribly homogeneous or isotropic, but in spite of this, both hypotheses are widely used.
9.5 Least-squares collocation

9.5.1 Stochastic processes in one dimension

Collocation is a statistical estimation technique used to estimate the values of a stochastic process, and calculate the uncertainties (e.g., mean errors) of the estimates.

Let \( \xi(t) \) be a stochastic process, the autocovariance function of which is \( C(t_i, t_j) \). Let the process furthermore be stationary, i.e., for any two points in time \( t_i, t_j \) we have \( C(t_i, t_j) = C(t_j - t_i) = C(\Delta t) \). The argument \( t \) is generally time, but could be any parameter, e.g., distance of a journey.

Of this process, we have observations made at times \( t_1, t_2, \ldots, t_N \); the process values for those times are \( \xi(t_1), \xi(t_2), \ldots, \xi(t_N) \). Let us assume, for the moment, that these values are error free observations. Then the observations are function values of process \( \xi \), stochastic quantities, the variance-covariance matrix of which we may write as follows (signal variance matrix):

\[
\text{Var} \{ \xi \} = \begin{bmatrix}
C(t_1, t_1) & C(t_1, t_2) & \cdots & C(t_1, t_N) \\
C(t_2, t_1) & C(t_2, t_2) & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
C(t_N, t_1) & C(t_N, t_2) & \cdots & C(t_N, t_N)
\end{bmatrix}.
\]

We use for this the symbol \( C_{ij} \), both for one element \( C_{ij} = C(t_i, t_j) \) of the matrix, and for the whole matrix, \( C_{ij} = [C(t_i, t_j), i, j = 1, \ldots, N] \). The symbol \( \xi_i \) again denotes a vector \( \xi(t_i), i = 1, \ldots, N \) consisting of process values – or one of its elements \( \xi(t_i) \).

Note that, if the function \( C(t_i, t_j) \) or \( C(\Delta t) \) is known, then the whole matrix and all of its elements can be calculated if only all parameter values \( t_i \) are known.

Let the shape of the problem now be, that one should estimate, i.e., predict, the value of \( \xi \) at the moment in time \( T \), i.e., \( \xi(T) \), using the above described observations \( \xi(t_i), i = 1, \ldots, N \).

In the same way as we calculated above the covariances between \( \xi(t_i) \) and \( \xi(t_j) \) (elements of the variance matrix \( C_{ij} \)), we also compute the variances between \( \xi(T) \) and all \( \xi(t_i), i = 1, \ldots, N \). We obtain

\[
\text{Cov} \{ \xi(T), \xi(t_i) \} = \begin{bmatrix}
C(T, t_1) & C(T, t_2) & \cdots & C(T, t_N)
\end{bmatrix}.
\]
For this we may again use the notation $C_{Tj}$. It is assumed here, that there is only one point in time $T$ for which estimation is done. Generalization to the case where there are several $T_p, p = 1, \ldots, M$, is straightforward. In that case, the covariance matrix will be of size $M \times N$:

$$
\text{Cov} \{ \xi(T_p), \xi(t_i) \} = 
\begin{bmatrix}
C(T_1, t_1) & C(T_1, t_2) & \cdots & C(T_1, t_N) \\
C(T_2, t_1) & C(T_2, t_2) & \cdots & C(T_2, t_N) \\
\vdots & \vdots & \ddots & \vdots \\
C(T_M, t_1) & C(T_M, t_2) & \cdots & C(T_M, t_N)
\end{bmatrix}.
$$

For this we may use the more general notation $C_{pj}$.

### 9.5.2 Signal and noise

The process $\xi(t)$ is called the signal. It is a physical phenomenon that we are interested in. There exist also physical phenomena that are otherwise similar, but that we are not interested in: to the contrary, we wish to remove their influence. Such stochastic processes are called noise.

When we make an observation, the purpose of which is to obtain a value for the quantity $\xi(t_i)$, we obtain in reality a value that is not absolutely precise. The real observation thus is

$$
\ell_i = \xi(t_i) + n_i.
$$

Here, $n_i$ is a stochastic quantity: observational error or noise. Let its variance – or more precisely, the joint variance-covariance of multiple observations – be $D_{ij}$. This is a very similar matrix to the above $C_{ij}$, and also symmetric and positive definite. The only difference is that $D$ designates noise, which we are not interested in. Often it may be assumed, that the errors $n_i, n_j$ of two different observations $\ell_i, \ell_j$ do not correlate, in which case $D_{ij}$ is a diagonal matrix.

### 9.5.3 The estimator and its error variance

Now we construct an estimator

$$
\hat{s}(T_p) \equiv \sum_i \Lambda_{pi} \ell_i
$$

a linear combination of the observations available $\ell_i$. The purpose in life of this estimator is to get as closely as possible to $\xi(T_p)$. So, the quantity to be minimized is the difference

$$
\hat{s}(T_p) - \xi(T_p) = \Lambda_{pi} \ell_i - \xi(T_p) = \Lambda_{pi} [\xi(t_i) + n_i] - \xi(T_p).
$$
Here, for the sake of writing convenience, we left the summation sign \( \sum \) away (Einstein summation convention): We always sum over adjacent, identical indices, in this case \( i \).

Study the variance of this difference, i.e.,

\[
\Sigma_{pp} \equiv \text{Var} \{ \hat{s}(T_p) - \hat{s}(T_p) \}.
\]

We exploit propagation of variances, the notations introduced above, and our knowledge that surely there is no physical relationship, or correlation, between observation process noise \( n \) and signal \( s \). Like this:

\[
\Sigma_{pq} = \Lambda_{pi} (C_{ij} + D_{ij}) \Lambda_{jq}^T + C_{pq} - \Lambda_{pi} C_{iq}^T - C_{pj} \Lambda_{jq}^T. \tag{9.5}
\]

The variances \( \Sigma_{pp} \) are now obtained by setting \( q = p \).

### 9.5.4 Showing optimality

Here we show that the optimal estimator is indeed the one producing the minimum possible variances.

Choose

\[
\Lambda_{pj} = C_{pi} (C_{ij} + D_{ij})^{-1}.
\]

Then, from equation (9.5), and exploiting the symmetry of the \( C \) and \( D \) matrices:

\[
\Sigma_{pp} = C_{pi} (C_{ij} + D_{ij})^{-1} C_{ip}^T + C_{pp} - C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp}^T -
\]

\[
- C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp}^T =
\]

\[
= C_{pp} - C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp}^T. \tag{9.6}
\]

Let us study next the alternative choice

\[
\Lambda_{pj} = C_{pi} (C_{ij} + D_{ij})^{-1} + \delta \Lambda_{pj}.
\]

In this case we obtain

\[
\Sigma'_{pp} = C_{pp} - C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp}^T +
\]

\[
+ \delta \Lambda_{pi} \left[ (C_{ij} + D_{ij}) \Lambda_{jp}^T \right] + \left[ \Lambda_{pi} (C_{ij} + D_{ij}) \right] \delta \Lambda_{jp}^T +
\]

\[
+ \delta \Lambda_{pi} (C_{ij} + D_{ij}) \delta \Lambda_{jp}^T - \delta \Lambda_{pi} C_{ip}^T - C_{pj} \delta \Lambda_{jp}^T =
\]

\[
= C_{pp} - C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp}^T +
\]

\[
+ \delta \Lambda_{pi} C_{ip}^T + C_{pj} \delta \Lambda_{jp}^T - \delta \Lambda_{pi} C_{ip}^T - C_{pj} \delta \Lambda_{jp}^T +
\]

\[
C_{pp} = C_{pp} - C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp}^T + \delta \Lambda_{pi} (C_{ij} + D_{ij}) \delta \Lambda_{jp}^T.
\]
Chapter 9. Statistical methods

Here, the last term is positive, because the matrices $C_{ij}$ and $D_{ij}$ are positive definite: $\Sigma'_{pp} > \Sigma_{pp}$, except when $\delta \Lambda_{pi} = 0$. In other words, the solution given above,

$$\Lambda_{pj} = C_{pi} (C_{ij} + D_{ij})^{-1} \Rightarrow \tilde{s} (T_p) = C_{pi} (C_{ij} + D_{ij})^{-1} L_{j},$$

is optimal in the sense of least squares – more precisely, in the sense of minimizing the error variance $\Sigma_{pp}, p = 1, \ldots, M$.

9.5.5 The covariance function of gravity anomalies

Least-squares collocation is used much to optimally estimate gravity values and other functionals of the gravity field on the Earth’s surface.

If we have two points, $P$ and $Q$, with measured gravity anomalies $\Delta g_P$ and $\Delta g_Q$, we would like to have the covariance between these two anomalies,

$$\text{Cov} \{ \Delta g_P, \Delta g_Q \}.$$

As argued in section 9.4, we can only empirically derive such a covariance by looking at all point pairs $(P, Q)$ in the same relative position around the globe, and averaging over them using the $M$ or $M'$ operator.

Normally the covariance is assumed to depend only on the distance $\psi$ between points $P, Q$. Then, we speak of an isotropic process $\Delta g (\phi, \lambda)$. Then also, the covariance will be

$$\text{Cov} \{ \Delta g_P, \Delta g_Q \} = M' \{ \Delta g_P, \Delta g_Q(P) \} = C (\psi_{PQ}).$$

A popular covariance function for gravity anomalies is Hirvonen’s formula:

$$C (\phi) = \frac{C_0}{1 + (\psi/\psi_0)^2}, \quad (9.7)$$

where $C_0 = C (0)$ and $\psi_0$ are parameters describing the behaviour of the gravity field. $C_0$ is called the signal variance, $\psi_0$ the correlation length. $\psi_0$ gives the distance at which the correlation between the gravity anomalies in two points is still 50%.

In local applications, instead of the angular distance $\psi$ one uses the metric distance

$$s = \psi R,$$

3Reino Antero Hirvonen (1908–1989) was a Finnish physical and mathematical geodesist.
where $R$ is the mean Earth radius. Then

$$C(s) = \frac{C_0}{1 + (s/d)^2}.$$ 

This equation was derived from gravimetric data for Ohio state, U.S.A., but it has broader validity. $C(0) = C_0$, the signal variance when $s = 0$. Also the variable $d$ is called the correlation length. It is the distance $d$ for which $C(d) = \frac{1}{4}C_0$, as seen from the equation.

The quantity $C_0$ varies considerably between areas, from hundreds to thousands of mGal$^2$, and is largest in mountainous areas. The quantity $d$ is generally order of magnitude tens of km.

**Warning:** The Hirvonen covariance formula is meant for use with (free-air) gravity anomalies, i.e., quantities obtained by subtracting normal gravity from the measured gravity. Nowadays anomalies are often obtained by subtracting from the observations a high degree “normal field”, i.e., a spherical harmonic expansion. Then one uses Hirvonen’s formula at one’s own risk!

Alternative functions that are also often used in local applications are the covariance functions of first and second order Gauss-Markov processes:

$$C(\psi) = C_0 e^{-\psi/\psi_0} 	ext{ or } C(\psi) = C_0 e^{-\psi^2/\psi_0^2}.$$
9.5.6 Least-squares collocation for gravity anomalies

If given are $N$ points $P_i, i = 1, \ldots, N$, where were measured gravity values (anomalies) $\Delta g_i$, we may, like above, construct a variance matrix

$$\text{Var} \left\{ \Delta g \right\} = \begin{bmatrix} C_0 & C(\psi_{21}) & \cdots & C(\psi_{N1}) \\ C(\psi_{12}) & C_0 & \cdots & C(\psi_{N2}) \\ \vdots & \vdots & \ddots & \vdots \\ C(\psi_{1N}) & C(\psi_{2N}) & \cdots & C_0 \end{bmatrix} = C_{ij},$$

where all elements $C(\psi_{ij})$ are calculated using the covariance function (9.7) given above.

If we also compute for the point $P$ in which gravity is unknown:

$$\text{Cov} \left\{ \Delta g_P, \Delta g_i \right\} = \begin{bmatrix} C(\psi_{P1}) & C(\psi_{P2}) & \cdots & C(\psi_{PN}) \end{bmatrix} = C_{Pi},$$

we obtain, in fully the same way as before, for the least-squares collocation solution:

$$\Delta g_P = C_{Pi} \left( C_{ij} + D_{ij} \right)^{-1} \Delta g_j \approx C_{Pi} C_{ij}^{-1} \Delta g_j,$$

where the $\Delta g_j$ are gravity anomaly observations made in points $j = 1, \ldots, N$.

The matrix $D_{ij}$ (which we ignore here) again describes the random observa-
tion error (imprecision, uncertainty) associated with making those measurements. Usually $D_{ij}$ is a diagonal matrix, i.e., the observations are statistically independent and don’t correlate with each other.

We may also compute a precision assessment of the above solution, i.e., the variance of prediction, (equation (9.11)):

$$\Sigma_{PQ} \approx C_{PQ} - C_{Pi}C_{ij}^{-1}C_{jQ}.$$  

In the case of one unknown prediction point $P, Q = P$ and

$$\text{Var} \{ \Delta g_P \} = \Sigma_{PP} = C_0 - C_{Pi}C_{ij}^{-1}C_{jP}.$$  

Its square root

$$\sigma_{\Delta g_P} = \sqrt{\Sigma_{PP}}$$  

is the mean error of estimator $\Delta g_P$.

### 9.5.7 Calculation example

![Graph](image)

*Given* two points where gravity has been measured and gravity anomalies calculated: $\Delta g_1 = 15$ mGal, $\Delta g_2 = 20$ mGal. The co-ordinates in the $x$ and $y$ directions are in kilometres. Assumed is that between the gravity anomalies of different points, Hirvonen’s covariance formula applies:

$$C(s) = \frac{C_0}{1 + \frac{s^2}{d^2}} \quad (9.8)$$

where $d = 20$ km and $C_0 = \pm 1000$ mGal$^2$. Additionally it is assumed that the gravity measurements done (including height determination of the gravity points!) were errorless. So, $D_{ij} = 0, i, j = 1, 2$.

*Calculate* an estimate of the gravity anomaly $\Delta g_P$ at point $P$ and its mean error $\sigma_{PP} = \sqrt{\Sigma_{PP}}$. 
Calculate first the distances \( s \) and the corresponding covariances \( C \).

\[
\begin{align*}
    s_{12}^2 &= \left( (30 - 20)^2 + (20 - 30)^2 \right) \text{ km}^2 = 200 \text{ km}^2 \\
    C_{12} &= C_{21} = \frac{1000 \text{ mGal}^2}{1 + 200/400} = 666.66 \ldots \text{ mGal}^2 \\
    s_{1P}^2 &= \left( (30 - 10)^2 + (20 - 10)^2 \right) \text{ km}^2 = 500 \text{ km}^2 \\
    C_{1P} &= \frac{1000 \text{ mGal}^2}{1 + 500/400} = 444.44 \ldots \text{ mGal}^2 \\
    s_{2P}^2 &= \left( (20 - 10)^2 + (30 - 10)^2 \right) \text{ km}^2 = 500 \text{ km}^2 \\
    C_{2P} &= \frac{1000 \text{ mGal}^2}{1 + 500/400} = 444.44 \ldots \text{ mGal}^2
\end{align*}
\]

From this follows

\[
C_{ij} + D_{ij} = C_{ij} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} 1000 & 666.66 \\ 666.66 & 1000 \end{bmatrix} \text{ mGal}^2,
\]

and its inverse matrix

\[
(C_{ij} + D_{ij})^{-1} = \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \text{ mGal}^{-2}.
\]

We also have

\[
C_{Pi} = \begin{bmatrix} C_{P1} & C_{P2} \end{bmatrix} = \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \text{ mGal}^2.
\]

As the vector of observations is

\[
\Delta g_j = \begin{bmatrix} \Delta g_1 \\ \Delta g_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 20 \end{bmatrix} \text{ mGal},
\]

we get the result

\[
\hat{\Delta g}_P = \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \begin{bmatrix} 444.44 \\ 444.44 \end{bmatrix} \text{ mGal} =
\]

\[
= 9.3333 \text{ mGal}.
\]

Precision:

\[
\Sigma_{PP} = C_{PP} - C_{Pi} (C_{ij} + D_{ij})^{-1} C_{Pj} =
\]

\[
= C_0 - \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \begin{bmatrix} 444.44 \\ 444.44 \end{bmatrix} \text{ mGal}^2 =
\]

\[
= 762.96 \text{ mGal}^2,
\]

i.e.,

\[
\sigma_{\Delta g_P} = \pm 27.622 \text{ mGal}.
\]

Summarizing the result:

\[
\hat{\Delta g}_P = 9.3333 \pm 27.622 \text{ mGal}.
\]
We may observe here, that the gravity anomaly estimate found is much smaller than its own uncertainty, and thus does not differ significantly from zero. In fact, not using the observational data at all would leave us with the a priori estimate
\[ \Delta g_P = 0 \pm \sqrt{1000} \text{ mGal} = 0 \pm 31.623 \text{ mGal}, \]
almost as good.

If, instead, we would choose to locate point \( P' \) in between points 1 and 2, at location \((25 \text{ km}, 25 \text{ km})\), then \( C_{P'1} = C_{P'2} = \frac{1000 \text{ mGal}^2}{1 + \frac{30}{400}} = 888.89 \text{ mGal}^2 \) and \( \Delta g_{P'} = 18.667 \pm 7.201 \text{ mGal} \), which is clearly better than the prior estimate of zero.

And if we had chosen instead the first order Gauss-Markov covariance formula
\[ C = C_0 e^{-s/d}, \]
we would have obtained the results \( \Delta g_P = 7.663 \pm 29.272 \text{ mGal} \) for the original point location, and \( \Delta g_{P'} = 16.460 \pm 18.426 \text{ mGal} \) for the shifted point location.

### 9.5.8 Theory of least-squares collocation

Above we presented one popular application of least-squares collocation. Here we look at the method more generally. The basic equation is:
\[ \hat{f} = C_{fg} \left[ C_{gg} + D_{gg} \right]^{-1} \left[ g + n \right]. \]  
(9.9)
The vector \( g \) contains observed quantities \( g_i \) (and is itself a stochastic quantity) and \( \hat{f} \) is a vector of quantities \( \hat{f}_i \) to be predicted. The hat is a commonly used symbol for an estimator.

Both vectors \( g \) and \( \hat{f} \) can, e.g., be gravity anomalies, in which case we have homogeneous prediction, a type of interpolation. More generally \( \hat{f} \) and \( g \) are of different type, e.g., \( \hat{f} \) consists of geoid heights \( N \), and \( g \) of gravity anomalies \( \Delta g \). In the latter case, the Stokes equation is “covertly” along in the structure of the \( C \) matrices.

These matrices are built from covariance functions. Their elements can be expressed as follows:
\[
\begin{align*}
[C_{fg}]_{ij} &= M \{ f_i g_j \} = \text{Cov} \{ f_i, g_j \}, \\
[C_{gg}]_{jk} &= M \{ g_j g_k \} = \text{Cov} \{ g_j, g_k \}, \\
[D_{gg}]_{jk} &= E \{ n_j n_k \} = \text{Cov} \{ n_j, n_k \}.
\end{align*}
\]
where \( n \), an element of vector \( \mathbf{n} \), represents the uncertainty of the observation process appearing in the observation equation (9.4):

\[
\ell_i = g_i + n_i \iff \ell = g + \mathbf{n}.
\]

\( \ell \) is the vector of observation values themselves, including measurement uncertainty \( \mathbf{n} \).

The \( D \) is the covariance matrix of observational errors, describing the observational process and not a property of the gravity field. While the values of \( M \left\{ \Delta g_i, \Delta g_j \right\} \) can be of order 1200 mGal\(^2\), the values of \( E \left\{ n_i, n_j \right\} \) can be much smaller, depending on measurement technique used, e.g., as small as 0.01 mGal\(^2\). Not however in the case of block means – e.g., averages over blocks of 1° \( \times \) 1°, computed from scattered measurements – which often are very imprecise.

The great advantage of least-squares collocation is its flexibility. Different observation types may be handled with a single unified theory and method, the locations of observation points (or blocks) are totally free, and the result is obtained directly as freely choosable quantities in locations where one wants them.

### 9.6 Prediction of gravity anomalies

If the quantity to be calculated or estimated, \( \hat{f} \), is of the same type as the observed quantity, \( g \), we often speak of prediction. E.g., the prediction equation for gravity anomalies already presented in subsection 9.5.6 is obtained from equation (9.9) by substitution:

\[
\Delta \hat{g}_P = C_{Pi} (C_{ij} + D_{ij})^{-1} \Delta g_j. \tag{9.10}
\]

Here are several points \( j \) where gravity is given: let us say, \( N \) observations \( \Delta g_j, j = 1, \ldots, N \). Points to be predicted there may be one, \( P \), or also many. The matrices \( C_{ij} \) and \( D_{ij} \) are square, and the inverse of their sum exists. \( C_{Pi} \) is a rectangular matrix; if there is only one point \( P \), it is a size 1 \( \times \) \( N \) row matrix.

The prediction error is now the difference quantity \( \Delta \hat{g}_P - \Delta g_P \), and its variance (“variance of prediction”) is

\[
\Sigma_{PP} \equiv \text{Var} \left\{ \Delta \hat{g}_P - \Delta g_P \right\} = \text{Var} \left\{ \Delta \hat{g}_P \right\} + \text{Var} \left\{ \Delta g_P \right\} - 2 \text{Cov} \left\{ \Delta \hat{g}_P, \Delta g_P \right\}.
\]
Here (propagation of variances applied to equation (9.10)):

\[ \text{Var} \left\{ \hat{\Delta}g_p \right\} = C_{pi} (C_{ij} + D_{ij})^{-1} C_{j\ell} (C_{k\ell} + D_{k\ell})^{-1} C_{\ell p} = \]
\[ = C_{pi} (C_{ij} + D_{ij})^{-1} \left[ (C_{jk} + D_{jk}) - D_{jk} \right] (C_{k\ell} + D_{k\ell})^{-1} C_{\ell p} = \]
\[ = C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp} - C_{pi} (C_{ij} + D_{ij})^{-1} D_{jk} (C_{k\ell} + D_{k\ell})^{-1} C_{\ell p}, \]
and

\[ \text{Cov} \left\{ \hat{\Delta}g_p, \hat{\Delta}g_p \right\} = C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp}. \]

Here, \( C_{iP} \) (or \( C_{jP}, \) or \( C_{\ell P} \)) is the transpose of \( C_{pi} \). The matrix \( \left( C_{ij} + D_{ij} \right)^{-1} \) is symmetric and its own transpose.

The end result is (remember that \( \text{Var} \left\{ \Delta g_p \right\} = C_{pp} \)):

\[ \Sigma_{pp} = C_{pp} - C_{pi} (C_{ij} + D_{ij})^{-1} C_{jp} - C_{pi} (C_{ij} + D_{ij})^{-1} D_{jk} (C_{k\ell} + D_{k\ell})^{-1} C_{\ell p}. \]

In case \( D_{ij} \ll C_{ij} \), we obtain a simpler, often used result:

\[ \Sigma_{pp} \approx C_{pp} - C_{pi} C_{ij}^{-1} C_{jp}. \tag{9.11} \]

Borderline cases:

1. Point \( P \) is far from all points \( i \). Then \( C_{pi} \approx 0 \) ja \( \Sigma_{pp} \approx C_{pp} \), i.e., prediction is impossible in practice (and the prediction formula will yield zero). The error of prediction is the same as the size of the signal (gravity anomaly) in the prediction point.

2. Point \( P \) is identical with one of the points \( i \). Then, if we use only that point \( i \), we obtain

\[ \Sigma_{pp} = C_{pp} - C_{pp} C_{pp}^{-1} C_{pp} = 0, \]
i.e., no prediction error whatsoever (as the value at the prediction point was already known!).

(However, if \( D_{pp} \neq 0 \) (but small), the result is \( \Sigma_{pp} = D_{pp} \). Show.)

### 9.7 Covariance function and degree variances

#### 9.7.1 The covariance function of the disturbing potential

In theoretical work we use, instead of gravity anomalies, rather the covariance function of the *disturbing potential* \( T \) on the Earth’s surface:

\[ K(P, Q) = M \left\{ T_P T_Q(P) \right\}. \]
We write this in the following form using the definition of $M' \{ \cdot \}$, equation (9.3):

$$K (\psi_{PQ}) = M' \left\{ T_P T_Q (P) \right\} = \frac{1}{8\pi^2} \int_0^{2\pi} \int_{\psi_{PQ} + \pi/2}^{\pi} \int_{-\pi/2}^{\pi/2} T_P T_Q (P) d\lambda_P d\phi_P d\alpha_P Q.$$

(9.12)

Here it is assumed that the potential is isotropic: $K$ does not depend on $\alpha$ but only on $\psi$.

We choose on the unit sphere a co-ordinate system where point $P$ is a “pole”. In this system the parameters $\alpha_{PQ}$ and $\psi_{PQ}$ are the spherical co-ordinates of point $Q$. The covariance function is expanded into the following sum:

$$K (\psi) = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} k_{nm} Y_{nm} (\alpha, \psi)$$

with $Y_{nm}$ defined as in equation (2.13).

Based on isotropy, all coefficients vanish\(^4\) for which $m \neq 0$:

$$K (\psi) = \sum_{n=2}^{\infty} k_{n0} Y_{n0} (\psi) = \sum_{n=2}^{\infty} k_{n} P_{n} (\cos \psi).$$

The coefficients $k_{n}$ are called the degree variances (of the disturbing potential). For isotropic covariance functions $K (\psi)$ the information content of the degree variances $k_{n}, n = 2, 3, \ldots$ is the same as that of the function itself, and is in fact its spectral representation.

### 9.7.2 Degree variances and spherical harmonic coefficients

We can in a simple way specialize equation (B.7):

$$f_n = \frac{2n + 1}{4\pi} \int_{0}^{\pi} f (\psi) P_n (\cos \psi) d\sigma = \frac{2n + 1}{2} \int_{0}^{\pi} f (\psi) P_n (\cos \psi) \sin \psi d\psi$$

if the expansion of function $f$ is

$$f (\psi) = \sum_{n=2}^{\infty} f_{n} P_{n} (\cos \psi).$$

Comparison with the previous yields

$$k_{n} = \frac{2n + 1}{2} \int_{0}^{\pi} K (\psi) P_{n} (\cos \psi) \sin \psi d\psi,$$

i.e., if $K (\psi)$ is given, we can calculate all $k_{n}$.

\(^4\)because

$$Y_{nm} (\alpha, \psi) = \begin{cases} P_{nm} (\cos \psi) \cos m\alpha & m \geq 0, \\ P_{n||m||} (\cos \psi) \sin ||m|| \alpha & m < 0, \end{cases}$$

an expression that can only be independent of $\alpha$ if $m = 0$.\)
Substituting $K(\psi_{PQ})$ from equation (9.12) yields

$$k_n = \frac{2n + 1}{16\pi^2} \int_{-\pi/2}^{+\pi/2} \int_0^{2\pi} T_P \left\{ \int_{\psi_{PQ}}^{\pi} T_{Q(P)} d\alpha_{PQ} P_n(\cos \psi_{PQ}) \sin \psi_{PQ} d\psi_{PQ} \right\} \cdot d\lambda_P \cos \phi_P d\phi_P.$$  

Here we have already interchanged the order of the integrals, as is allowed, and moved $T_P$ to another place.

The expression inside the curly braces is a surface integral over the unit sphere

$$\int_{\psi_{PQ}}^{\pi} T_{Q(P)} P_n(\cos \psi_{PQ}) \sin \psi_{PQ} d\psi_{PQ} = \int_{\sigma} T_P(\cos \psi_{PQ}) d\sigma = 4\pi T_n(\sigma),$$  

the constituent of $T$ for the harmonic degree number $n$, compare the degree constituent equation (2.19). Substitution yields

$$k_n = \frac{1}{4\pi} \int_{-\pi/2}^{+\pi/2} \int_0^{2\pi} TT_n(\cos \phi) d\lambda d\phi = 4\pi T_n = M[T^n],$$  

according to the definition of operator $M$, and considering the orthogonality of the functions $T_n$.

If we now write (with the familiar definitions)

$$T(\phi, \lambda) = \sum_{n=2}^{\infty} T_n(\phi, \lambda) =$$

$$\sum_{n=2}^{\infty} \sum_{m=0}^{n} \left[ a_{nm} P_n(\cos \psi) \cos m\lambda + b_{nm} \bar{P}_n(\sin \psi) \sin m\lambda \right] =$$

$$\sum_{n=2}^{\infty} \sum_{m=-n}^{n} \bar{f}_{nm} Y_{nm}(\phi, \lambda)$$

we obtain

$$K(\psi) = \sum_{n=2}^{\infty} k_n P_n(\cos \psi) = \sum_{n=2}^{\infty} \frac{1}{4\pi} \int_{\sigma} T_n^2 d\sigma \cdot P_n(\cos \psi) =$$

$$\sum_{n=2}^{\infty} \left( \sum_{m=0}^{n} \bar{a}_{nm}^2 + \bar{b}_{nm}^2 \right) P_n(\cos \psi) = \sum_{n=2}^{\infty} \left( \sum_{m=-n}^{n} \bar{f}_{nm}^2 \right) P_n(\cos \psi).$$  

Here, we have exploited the orthonormality of the fully normalized base functions $Y_{nm}$ (or $\bar{P}_{nm}(\sin \psi)$). One sees from the equation, that

$$k_n = \sum_{m=0}^{n} \bar{a}_{nm}^2 + \bar{b}_{nm}^2 = \sum_{m=-n}^{n} \bar{f}_{nm}^2,$$  

(9.13)
The degree variances \( k_n \) of the potential can be calculated directly from the spherical harmonic coefficients.

The literature offers also many alternative notations for the degree variances, e.g.:

\[
k_n \equiv \sigma_n^2 \equiv \sigma_i^{TT}.
\]

### 9.8 Propagation of covariances

The covariance function \( K \) derived above can be used to also derive the covariance functions of other quantities. This works in principle for quantities that can be expressed as linear functionals of the disturbing potential \( T(\cdot,\cdot,R) \) on the Earth’s surface, as explained in section 9.2.

#### 9.8.1 Example: upward continuation of the potential

Let us write the disturbing potential in space \( T(\phi,\lambda,r) \) as a functional of the surface disturbing potential \( T(\phi,\lambda,R) = T(\cdot,\cdot,R) = T_R \). We know that

\[
T(\phi,\lambda,r) = \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} T_n(\phi,\lambda).
\]

We may express this symbolically:

\[
T(\phi,\lambda,r) = L\{T_R\},
\]

Here, \( L \) is the linear operator (functional)

\[
L\{f\} = \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} f_n,
\]

where the \( f_n \) are defined according to the degree constituent equation (2.19), so that on the surface of the sphere

\[
f = \sum_{n=2}^{\infty} f_n.
\]

Symbolically we may write

\[
L\{f\} = \sum_{n=2}^{\infty} L^n f_n,
\]

where

\[
L^n = \left( \frac{R}{r} \right)^{n+1}
\]
is the spectral representation of the operator $L$.

We may still write in a certain point $P (\phi, \lambda, r)$ in space:

$$L_P \{ f \} = \sum_{n=2}^{\infty} L^n_P f_n; \quad L^n_P = \left( \frac{R}{r_P} \right)^{n+1}.$$ 

Now the covariance function in space of $T$ is obtained:

$$K (r_P, r_Q, \psi_{PQ}) = M \{ T (\phi, \lambda, r_P) T (\phi_Q, \lambda_Q, r_Q) \} =$$

$$= M \{ L_P \{ T_R \} L_Q \{ T_R \} \} =$$

$$= M \left\{ \sum_{n=2}^{\infty} [L^n_P T_n] \sum_{n'=2}^{\infty} [L^n_Q T_{n'}] \right\} =$$

$$= \sum_{n=2}^{\infty} \sum_{n'=2}^{\infty} L^n_P L^n_Q M \{ T_n T_{n'} \}.$$ 

Based on the orthogonality of the functions $T_n$ the following holds

$$M \{ T_n T_{n'} \} = \begin{cases} k_n P_n (\cos \psi_{PQ}) \quad &\text{if} \quad n = n' \\ 0 \quad &\text{if} \quad n \neq n' \end{cases}$$

i.e., the harmonic components of the surface covariance function

$$K (\psi_{PQ}) = \sum_{n=2}^{\infty} k_n P_n (\cos \psi_{PQ}).$$

Thus we obtain\(^5\)

$$K (r_P, r_Q, \psi_{PQ}) = \sum_{n=2}^{\infty} \left( \frac{R^2}{r_P r_Q} \right)^{n+1} k_n P_n (\cos \psi_{PQ}). \quad (9.14)$$

Here we have expressed the covariance function of the potential in space $T (\phi, \lambda, r)$ into an expansion into the degree variances $k_n$ of the corresponding Earth surface potential $T (\phi, \lambda, R) = T_R$. Thus we have obtained the three-dimensional covariance function for the disturbing potential, needed, e.g., in mountainous countries and in air and space applications.

### 9.8.2 Example: the covariance function of gravity anomalies

We know that there exists the following relationship between gravity anomalies and the disturbing potential:

$$\Delta g = \frac{1}{R} \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} (n-1) T_n,$$

\(^5\) This works only this cleanly because in this case the operator $L^n$ is of multiplier type, $(R/r)^{n+1}$. 

---

\(^{189}\)
symbolically: \( \Delta g = L_g \{ T \} \) for a suitable \( L_g \) functional:

\[
L_g \{ f \} = \sum_{n=2}^{\infty} L^n_g f_n,
\]

where now

\[
L^n_g = \frac{n-1}{R} \left( \frac{R}{r} \right)^{n+1}.
\]

Now we can show in the same way as above, that

\[
M \{ \Delta g_P \Delta g_Q \} = \sum_{n=2}^{\infty} (n-1)^2 \left( \frac{R^2}{r_P r_Q} \right)^{n+1} k_n P_n (\cos \psi_{PQ}).
\]

Often we write

\[
C (\psi_{PQ}) = \sum_{n=2}^{\infty} \left( \frac{R^2}{r_P r_Q} \right)^{n+1} c_n P_n (\cos \psi_{PQ}),
\]

where the degree variances of gravity anomalies are

\[
c_n = \left( \frac{n-1}{R} \right)^2 k_n.
\]

Similarly we calculate also the “mixed covariances” between disturbing potential and gravity anomaly:

\[
\text{Cov} \left\{ T_P, \Delta g_Q \right\} = M \{ T_P \Delta g_Q \} = -\sum_{n=2}^{\infty} \frac{n-1}{R} \left( \frac{R^2}{r_P r_Q} \right)^{n+1} k_n P_n (\cos \psi_{PQ}).
\]

All these are examples of propagation of covariances, when applied to a series expansion:

\[
\text{Cov} \left\{ L_1 \{ T_P \}, L_2 \{ T_Q \} \right\} = \sum_n L^n_{1,P} L^n_{2,Q} M \{ T_n T_n \} = \sum_n L^n_{1,P} L^n_{2,Q} k_n P_n (\cos \psi_{PQ}),
\]

for arbitrary functionals

\[
L_1 \{ T_P \} = \sum_{n=2}^{\infty} L^n_{1,P} T_n, \quad L_2 \{ T_Q \} = \sum_{n=2}^{\infty} L^n_{2,Q} T_n,
\]

where the \( T_n = T_n (\phi, \lambda) \) are the degree constituents of the disturbing potential on the Earth’s surface.

### 9.9 Global covariance functions

Empirical covariance functions have been calculated a lot. Empirical covariance functions for the whole Earth there have been only a few. Typically
they are given in the form of a degree variance formula. The best known is the one observed by William Kaula:

\[ k_n = \alpha n^{-4}. \]

By writing

\[ c_n = \left( \frac{n-1}{R} \right)^2 k_n, \]

where \( c_n \) are the degree variances of gravity anomalies, we obtain

\[ c_n = \frac{\alpha}{R^2} \frac{(n-1)^2}{n^4} \approx \frac{\alpha}{R^2} n^{-2}. \]

Here, \( \alpha/R^2 \) is a planet specific constant, value about 1200 mGal\(^2\) for the Earth. The Kaula rule does not hold very precisely for very high degree numbers. It applies, by the way, fairly well for the gravity field of Mars, of course with a different constant (Yuan et al., 2001).

Another well known rule is the Tscherning-Rapp formula (Tscherning and Rapp, 1974):

\[ c_n = \frac{A (n-1)}{(n-2) (n+B)} = \left( \frac{n-1}{R} \right)^2 k_n. \]

The constants are, according to the authors, \( A = 425.28 \) mGal\(^2\) and \( B = 24 \) (exactly). As a technical detail, one usually chooses \( R = R_B = 0.999 \overline{R} \), the radius of a Bjerhammar sphere inside the Earth (\( \overline{R} \) is the Earth mean radius). The form of the above equation is chosen so the covariance functions of various quantities will be closed expressions.

### 9.10 Collocation and the spectral viewpoint

Also the calculations in least-squares collocation can be executed efficiently by way of FFT. For this one should study the symmetries present in the geometry, especially the rotational symmetry, which exists, e.g., in the direction of longitude on the whole Earth: nothing changes when we turn the whole Earth by a certain angle \( \theta \) around its rotation axis: for all longitudes \( \lambda \to \lambda + \theta \). In the following we discuss a simplified example.

---

6William M. Kaula (1926–2000) was an American geophysicist and space geodesist who studied the determination of the Earth’s gravity field by means of satellite geodesy.

7Carl Christian Tscherning (1942–2014) was a Danish geodesist and expert in the mathematics of geoid determination.

8Arne Bjerhammar (1917–2011) was an eminent Swedish geodesist.
Chapter 9. Statistical methods

Figure 9.4. Global covariance functions as degree variances. The GOCE model cuts off at degree 280.

Let observations $\ell_i = g_i + n_i$ of a field $g(\psi), \psi \in [0, 2\pi)$ be given on the edge of a circle, in points $\psi_i \equiv 2\pi i/N, i = 0, 1, 2, \ldots, N - 1$. Let us assume that also the results of the calculation, i.e., estimates $\hat{f}_i$ of the result function $f(\psi)$ are desired in the same points. Then equation (9.9) yields

$$\hat{f} = C_{fg} \left[ C_{gg} + D_{gg} \right]^{-1} \left( g + n \right)$$  \hspace{1cm} (9.15)

with

$$[C_{fg}]_{ij} = C_{fg} (f(\psi_i), g(\psi_j)) = C_{fg} (\psi_i, \psi_j),$$

$$[C_{gg}]_{ij} = C_{gg} (g(\psi_i), g(\psi_j)) = C_{gg} (\psi_i, \psi_j),$$

$$[D_{gg}]_{ij} = D_{gg} (g(\psi_i), g(\psi_j)) = D_{gg} (\psi_i, \psi_j).$$

If the physics of the whole situation is rotationally symmetric, we must have

$$C_{fg} [f(\psi_i), g(\psi_j)] = C_{fg} [(\psi_i - \psi_j) \mod 2\pi] = C_{fg} [(i - j) \mod N],$$

and similarly

$$C_{gg} [g(\psi_i), g(\psi_j)] = C_{gg} [(i - j) \mod N].$$
Also because generally the observations do not correlate with each other, we have
\[ D_{gg} = \sigma^2 I_N, \]
\( \sigma^2 \) (the variance of observations, assumed equal for all) times the \( N \times N \) unit matrix.

Matrices of this form are called Toeplitz circulant\(^9\). Thanks to this property, equation (9.15) is a string of convolutions.

Without proof we present that the spectral version of equation (9.15) looks like this:
\[
\hat{f} = \mathcal{F} \{ C_{fg} \} \mathcal{F} \{ C_{gg} \} + \sigma^2 \mathcal{F} \{ \hat{g} + \hat{n} \}. \tag{9.16}
\]

This is an easy and rapid way to calculate the solution using FFT. In the limit in which the observations are exact, i.e., \( \sigma^2 = 0 \), by equation (9.16) \( \hat{f} \) follows straight from \( \hat{g} + \hat{n} = \hat{g} \). If for a suitable functional \( L \) we have \( f = L \{ g \} \), the equation simplifies as follows:
\[
\mathcal{F} \{ \hat{f} \} = \frac{\mathcal{F} \{ L \}}{\mathcal{F} \{ C_{gg} \} + \sigma^2} \mathcal{F} \{ \hat{g} + \hat{n} \},
\]

\( \sigma^2 \) is the variance of observations, assumed equal for all.

---

\(^9\)In fact, the unit or identity matrix is also known as the Kronecker delta, and as a Toeplitz matrix may be interpreted as a discrete version of the Dirac delta function. Its discrete Fourier transform is \( \mathcal{F} \{ I \} = 1 \).

\(^{10}\)Otto Toeplitz (1881–1940) was a German Jewish mathematician who contributed to functional analysis.
and if also \( \sigma^2 = 0 \), then

\[
\mathcal{F}\{\hat{f}\} = \mathcal{F}\{L\} \mathcal{F}\{g\} \iff \hat{f} = L \{g\}.
\]

E.g., if the \( \Delta g_i = g_i + n_i \) are gravity anomalies and the \( f \) values of the disturbing potential, then\(^{11}\)

\[
\mathcal{F}\{L\} = \frac{R}{(n-1)}.
\]

The approach is called Fast Collocation, e.g., Bottoni and Barzaghi (1993). Of course it is used in two dimensions on the Earth surface, though our example is one-dimensional. As always, it requires that the observations are given on a grid, and in this case also, that the precision of the material is homogeneous (the same everywhere) over the area. This requirement is hardly ever precisely fulfilled.

### 9.11 Exercises

#### 9.11.1 Hirvonen’s covariance formula and prediction

Hirvonen’s covariance equation is

\[
C(s) = \frac{C_0}{1 + (s/d)^2},
\]

and with the Ohio parameters, \( C_0 = 337 \text{ mGal}^2 \) and \( d = 40 \text{ km} \),

1. What is the “variance of prediction” of gravity anomalies in point \( Q \) which is at a distance of 5 km from a (precisely!) given anomaly in point \( P \)?
2. And what if the distance is 25 km?

#### 9.11.2 Propagation of covariances

Given the covariance function (9.14) of the disturbing potential

\[
\operatorname{Cov}(T_P, T_Q) = \sum_{n=2}^{\infty} \left( \frac{R^2}{r_{PQ}} \right)^{n+1} k_n P_n(\cos \psi),
\]

\(^{11}\)In real computation it is not so simple... the degree number \( n \), which refers to global spherical geometry, must first be converted to the Fourier wave number expressed on the computational grid used.
1. calculate the covariance function of the gravity disturbance $\delta g$ (equation (4.3)).

2. Compute the covariance function of the gravity gradient $\partial^2 T/\partial z^2$ (i.e., the vertical gradient of the gravity disturbance!).

### 9.11.3 Underground mass points

1. If a mass point is placed inside the Earth at a depth $D$ beneath an observation point $P$, what then is the correlation length $s$ of the gravitational field it causes on the Earth surface, for which $C(s) = \frac{1}{2} C_0$?

2. Thus, if we wish to construct a model made of mass points, where under each observation point $\Delta g_P$ there is one mass point, how deep should we place them if the correlation length $d$ is given?
10. Gravimetric measurement devices

10.1 History

The first device ever built based on a pendulum was a clock. The pendulum equation,

\[ T = 2\pi \sqrt{\frac{\ell}{g}} \]

tells that the swinging time \( T \) of a pendulum of a given length is a constant that depends only on the length \( \ell \) and local gravity \( g \), on condition that the swings are small. The Dutch Christiaan Huygens\(^1\) built in 1657 the first useable pendulum clock based on this (http://en.wikipedia.org/wiki/Pendulum_clock).

When the young French researcher Jean Richer\(^2\) visited French Guyana in 1671 with a pendulum clock, he noticed that the clock ran clearly slower. The matter was corrected simply by shortening the pendulum. The cause of the effect could not be the climatic conditions in the tropics, i.e., the thermal expansion of the pendulum; the right explanation was that in the tropics, gravity \( g \) is weaker than in Europe. After return to France, Richer had again to make his pendulum longer. The observation is described in just one paragraph on pages 87–88 in his report “Observations astronomiques et physiques faites en l’île de Caïenne”, Richer (1731).

This is how the pendulum gravimeter was invented. Later, much more pre-

\(^1\)Christiaan Huygens (1629–1695) was a leading Dutch natural scientist and mathematician. Besides inventing the pendulum clock, he also was the first to realize that the planet Saturn has a ring.

\(^2\)Jean Richer (1630–1696) was a French astronomer. He is really only remembered for his pendulum finding.
cise special devices were built, e.g., Kater’s reversion pendulum, and the four-pendulum Von Sterneck device, which was also used in Finland in the 1920’s and 1930’s. We must mention also the submarine measurements, e.g., in the Java Sea by the Dutch F.A. Vening Meinesz in which it was observed that above the trenches in the ocean floor there is a notable shortage of gravity, and that they thus are in a state of strong isostatic disequilibrium.

For production gravimetric observations, pendulum gravimeters are however too hard to operate and too slow. For that purpose the spring gravimeter has been developed, see section 10.2.

Pendulum gravimeters are in principle absolute measurement devices, i.e., gravity is obtained directly as an acceleration. There are, however, systematic effects associated with the suspension of the pendulum that make that one cannot trust in the absoluteness of measurement after all. One tried out solution is the very long wire pendulum, e.g., Hytönen (1972). However, nowadays absolute measurements are made with ballistic gravimeters, cf. section 10.3. It has been observed that the older measurements made with pendulum apparatus in the so-called Potsdam system are systematically 14 mGal too large...
In a linear spring balance the equation of motion of the test mass is

$$m \left( \frac{d^2 \ell}{dt^2} - g \right) = -k (\ell - \ell_0),$$

where $m$ is the test mass, $g$ the local (to be measured) gravity, $k$ the spring constant. The quantity $\ell_0$ is the “rest length” of the spring, its length if there were no external forces acting on it; $\ell$ is the true, instantaneous length of the string.

The equilibrium between the spring force and gravity is

$$\frac{d^2 \ell}{dt^2} = 0 \Rightarrow mg = k (\ell - \ell_0) = k (\bar{\ell} - \ell_0), \quad (10.1)$$

where $\bar{\ell}$ is the mean length of the spring during the oscillation, and also the equilibrium length in the absence of oscillations.

When the test mass is disturbed, it starts oscillating about its equilibrium position. The oscillation equation, obtained by summing the above two equations, is

$$\frac{d^2 \ell}{dt^2} = -\frac{k}{m} (\ell - \bar{\ell}).$$

The period is

$$T = 2\pi \sqrt{m/k} = 2\pi \sqrt{(\bar{\ell} - \ell_0)/g} = 2\pi \sqrt{\delta \ell/g}, \quad (10.2)$$

where $\delta \ell = \bar{\ell} - \ell_0$ is the difference between the equilibrium length and the length in the state of rest, i.e., the lengthening of the spring by gravity.
Chapter 10. Gravimetric measurement devices

The sensitivity of the instrument is obtained by differentiating equation (10.1) in the form

\[ mg = k \left( \ell - \ell_0 \right) = k \delta \ell \]

with the result

\[ \frac{d\ell}{dg} = \frac{d (\delta \ell)}{dg} = \frac{m}{k} = \frac{T^2}{4\pi^2}. \tag{10.3} \]

Substitution, e.g., of \( \delta \ell = 5 \text{ cm} \) and \( g = 10 \text{ m/s}^2 \) into equation (10.2) yields \( T = 0.44 \text{ s} \). One milligal of change in gravity \( g \) produces according to equation (10.3) a lengthening of only \( 5 \cdot 10^{-8} \text{ m} \) (check)! Clearly then, the sensor observing or compensating this displacement must be extremely sensitive!

### 10.2.1 Astatization

An astatized gravimeter offers a different measurement geometry. We use as our example the LaCoste-Romberg gravimeter which long enjoyed great popularity. In it, the test mass is at the end of a cantilever beam, see figure 10.3. Two torques are operating on the beam, which are in equilibrium. The torque by the spring is

\[ \tau_s = k \left( \ell - \ell_0 \right) b \sin \beta, \]

where \( \ell \) is the spring’s true (stretched) and \( \ell_0 \) the theoretical or state-of-rest length without loading.

According to the sine rule

\[ \ell \sin \beta = c \sin (90^\circ + \epsilon) = c \cos \epsilon, \]

from which upon substitution in the previous:

\[ \tau_s = k \left( \ell - \ell_0 \right) \frac{bc}{\ell} \cos \epsilon. \]

Gravity pulling at the mass again is \( mg \), and the corresponding torque

\[ \tau_g = mgp \cos \epsilon. \]

Between these there has to be equilibrium:

\[ \tau_g - \tau_s = mgp \cos \epsilon - k \left( \ell - \ell_0 \right) \frac{bc}{\ell} \cos \epsilon = 0, \tag{10.4} \]

or

\[ mgp \ell - kbc \left( \ell - \ell_0 \right) = 0. \tag{10.5} \]
10.2. The relative (spring) gravimeter

Figure 10.3. Operating principle of spring gravimeter. On the right, how to build a “zero-length spring”.

By differentiation

\[ mp \ell \frac{dg}{dg} + mg \frac{d\ell}{d\ell} - kbc \frac{d\ell}{d\ell} = 0 \]

from which we obtain, by substituting equation (10.5), a sensitivity equation:

\[ \frac{d\ell}{dg} = - \frac{mp\ell}{mgp - kbc} = - \frac{mp\ell}{mgp - mgp \ell - \ell_0} = \frac{\ell - \ell_0}{g} \ell_0. \]

From this we see that the sensitivity can be driven up arbitrarily by choosing \( \ell_0 \) as short as possible, almost zero – a so-called zero-length spring solution\(^6\).

Of course, levelling the instrument, with its bull’s eye level and three foot-screws, is critical.

E.g., assuming \( \ell = 5 \) cm, \( \ell_0 = 0.1 \) cm, \( g = 10 \) m/s\(^2\) gives

\[ \frac{d\ell}{dg} = 2.5 \cdot 10^{-6} \text{ m/mGal}, \]

a 50 times\(^7\) better result than earlier! The “improvement ratio” is precisely \( (\ell - \ell_0)/\ell_0 \).

This is the operating principle of the astatized gravimeter, e.g., the LaCoste-Romberg\(^8\).

---


\(^7\)For comparability we should still multiply by \( p/b \sin \beta \), if we measure the position of the test mass.

\(^8\)Lucien LaCoste (1908–1995) was an American physicist and metrologist, who,
10.2.2 Period of oscillation

There is another way to look at this: if the instrument is not in equilibrium, the beam will slowly oscillate about the equilibrium position. We start from equation (10.5):
\[
mg \ell - kbc (\ell - \ell_0) = 0,
\] (10.6)
but for a state of disequilibrium; then, the test mass will be undergoing an acceleration \(a\), and we have
\[
m (g - a) \ell - kbc (\ell - \ell_0) = 0,
\]
where, instead of the equilibrium spring length \(\ell\), we have the instantaneous length \(\ell\). Subtracting the above two equations yields
\[
mg \ell - map \ell - kbc (\ell - \ell) = 0.
\]
We use equation (10.6) again to eliminate \(kbc\), yielding:
\[
mg \ell - map \ell - mgp \frac{\ell}{\ell - \ell_0} (\ell - \ell) = 0.
\]
Rearranging terms gives
\[
map \ell = mgp \frac{\ell_0}{\ell - \ell_0} (\ell - \ell),
\]
or
\[
a = \frac{g \ell_0}{\ell - \ell_0} (\ell - \ell).
\]
Here we see again the “astatizing ratio” \(\ell_0/(\ell - \ell_0)\) appear, which for a zero-length spring \((\ell_0 \approx 0)\) is very small.

Now the string length disequilibrium \(\ell - \ell\) is connected with the vertical displacement \(z\) (reckoned upward) of the test mass, as follows:
\[
z = - \left( \ell - \ell \right) \frac{p}{b \sin \beta}.
\]
With this we get
\[
a = \frac{d^2 z}{dt^2} = \frac{g \ell_0 \ell}{\ell - \ell_0} \frac{b \sin \beta}{p} z.
\]
This is again an oscillation equation in \(z\), with a period of
\[
T = 2\pi \sqrt{\frac{\ell}{g \frac{p}{b \sin \beta} \frac{\ell - \ell_0}{\ell_0}}}
\]

as an undergraduate, together with his physics professor Arnold Romberg (1882–1974) discovered the principle of the astatized gravimeter and zero-length spring.
10.2. The relative (spring) gravimeter

The relative (spring) gravimeter

The elastic force of an ordinary spring grows steeply with extension (left), whereas the weight of the test mass is constant. The cantilever, diagonal arrangement (right) causes the part of the force of the spring in the direction of motion of the cantilever (red) to diminish with extension, while the spring force itself grows similarly with extension. This near-cancellation boosts sensitivity. The spring used is a zero-length spring.

For the same values as above, $\ell_0 = 0.1$ cm, $\ell = 5$ cm $\approx \ell$, $g = 10$ m/s$^2$, and $p/b \sin \beta = 2$, we find

$$T = 4.4 \text{ s}.$$

What this long oscillation period also means is, that the instrument is insensitive to high-frequency vibrations by passing traffic, microseismicity, etc. A significant operational advantage.

10.2.3 Practicalities of measurement

An ordinary spring gravimeter is based on elasticity. Because there is no material that is perfectly elastic, but always also plastic (viscous)$^9$, the gravimeter itself changes during the measurement process. This change is called drift. The drift is managed in practical measurements by the following measures:

- we measure along lines starting from a known point and ending on a known point, producing a closing error. The line is traversed as rapidly as possible. The closing error is eliminated by adjusting the values obtained from the measurement in proportion to their times of measurement.
- The gravimeter is transported carefully without bumping it, and
- we remember always to arrest (clamp down the cantilever) during transport!

$^9$Plastic deformation in a metal crystal is mediated by crystal-lattice defects called dislocations. As dislocations travel through the crystal lattice under load, the properties of the metal change, and may eventually result in metal fatigue, a known problem, e.g., in aviation. https://en.wikipedia.org/wiki/Dislocation.
Because the elastic properties of the spring and the instrument geometry both depend on temperature, precision gravimeters are always thermostated.

A sea gravimeter differs from an ordinary (land) gravimeter in having a powerful damping. This applies also for an airborne gravimeter. Both types are mounted on a stabilized platform, keeping the measurement axis along the local vertical in spite of vehicle motions.

### 10.3 The absolute (ballistic) gravimeter

The absolute or ballistic gravimeter is a return to roots, the definition of gravity: it measures directly the acceleration of free fall. The instrument comprises a vacuum tube, inside of which an object, a prism reflecting light, falls freely.

Here we describe shortly the JILA gravimeter, built at the University of Colorado at Boulder by Jim Faller\(^{10}\) of which the Finnish Geodetic Institute has acquired two. Figure 10.6 shows the newer model, FG5, built by the same group. In Finland this instrument, serial number 221, has served as the national standard for the acceleration of free fall. It was upgraded to a model FG5X in 2012.

During the fall of the prism, a “cage” with a window in the bottom moves along with the prism inside it without touching it. The purpose of the cage is to prevent the last remaining traces of air from affecting the motion of the prism. Approaching the bottom, the cage, which moves along a rail under computer control, decelerates, and the prism lands relatively softly on its bottom. After that the cage moves back to the top of the tube and a new measurement cycle starts.

A laser interferometer measures the locations of the prism during its fall; the measurements are repeated thousands of times to get a good precision through averaging. Another prism, the reference prism, is suspended in another tube from a very soft spring (actually an electronically simulated “superspring”) to protect it from microseismicity. The instrument is designed to

---

\(^{10}\)James E. Faller (1934 – ) is an American physicist, metrologist, geodesist and student of gravitation. He proposed the installation of laser retroreflectors on the Lunar surface in the context of the Apollo project, in order to measure the distance to the Moon – LLR, Lunar Laser Ranging.
achieve the greatest precision possible; e.g., the vibration caused by the drop is controlled by a well-designed mount. Precisions are of order several $\mu$Gal, similar to what ordinary LaCoste-Romberg relative gravimeters are capable of. The instrument is however large and, though transportable, it cannot be called a field instrument. Of late, development has gone in the direction of smaller devices, which are essentially better portable.

The motion of a freely falling mass is described by the equation

$$\frac{d^2}{dt^2}z = g(z),$$

where it is assumed – realistically – that gravity $g$ depends on the location $z$ within the drop tube. If we nevertheless take $g$ to be constant, we obtain by integration

$$\frac{dz}{dt} = v_0 + gt,$$
$$z = z_0 + v_0t + \frac{1}{2}gt^2,$$

from which we obtain the observation equations of the measurement process

$$z_i = \begin{bmatrix} 1 & t_i & \frac{1}{2}t_i^2 \end{bmatrix} \cdot \begin{bmatrix} z_0 \\ v_0 \\ g \end{bmatrix}. \quad (10.7)$$
Here, the unknowns\footnote{It would be easy (exercise!) to add an unknown representing the vertical gradient of gravity to this.} are $z_0, v_0$ and $g$. The quantities $z_i$ are the interferometrically measured vertical locations of the falling prism. Determining precisely the corresponding measurement time or epoch $t_i$ is of course essential. The volume of measurements obtained from each drop is large. We write the observation equations in matric form:

$$\ell = Ax,$$

where

$$\ell = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_i \\ \vdots \\ z_n \end{bmatrix}, \quad A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_i & t_i^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} z_0 \\ v_0 \\ g \end{bmatrix}. $$

From this, the solution follows according to the method of least squares, from the normal equations

$$A^T A \hat{x} = A^T \ell$$

giving the solution

$$\hat{x} = \left[ A^T A \right]^{-1} A^T \ell.$$  

\footnote{It would be easy (exercise!) to add an unknown representing the vertical gradient of gravity to this.}
An alternative type of absolute gravimeter throws the prism up (inside the tube), after which it moves along a symmetric parabolic path. Such a “rise-and-fall” instrument is, e.g., the Italian IMGC-02. Theoretically this method would give more precise results; the technical challenges are however larger than in case of the dropping method. Intercomparisons between instruments of these two types have helped to identify error sources.

Recently also so-called atomic or quantum gravimeters have been built, in which interferometrically the falling of individual atoms is measured (de Angelis et al., 2009). The idea of the device is, that it measures the effect of gravity on the phase angle of the matter wave of falling atoms. Firstly an extremely cold, so-called Bose–Einstein condensate is prepared; perhaps a million atoms in identical quantum states, with the same phase angle like marching soldiers. The condensate is dropped, and the first laser pulse splits it into two. Half of the atoms\(^{12}\) fall first slowly, then faster; the other half fast at first and then slower. In order to achieve this, a second laser pulse pair is used that acts like a mirror, or perhaps a tennis racket. The third and last laser pulse is for reading out interferometrically the phase difference between the two merging atomic beams. The interaction between light and atoms is based on the Raman effect.

As the atoms travel two different paths through space-time where the gravity potential is different\(^{13}\), a phase difference is formed between these which can be measured. Without gravity (dashed lines) this phase difference would be zero. See figure 10.7, where the horizontal axis is time.

### 10.4 Network hierarchy in gravimetry

In gravimetry, network hierarchy is just as important as in measurements of location or height. The procedure has typically been, that the highest measurement order consisted of points measured by absolute gravimeters; in the old days this meant pendulum measurements. Stepwise densification of this network, i.e., measurement of the base network, was then done with relative or spring gravimeters, like also the lowest-order measurements, gravity mapping surveys. In base network measurement, fast transportation

\(^{12}\)This is a quantum theoretically erroneous statement. The matter wave of each individual atom splits into two!

\(^{13}\)In fact, the spinning of the atom’s phase angle acts like a clock, and the speed at which time elapses depends on the local geopotential (Vermeer, 1983).
was used, such as aircraft: national or regional reference points often were located at airports.

Because pendulum instruments were not genuinely absolute, the old, so-called Potsdam system collected a 14 mGal error: all values were that much too high. Nowadays we use instead ballistic gravimeters, the possible systematics of which are much smaller – but not nonexistent, order of magnitude microgals. As there are no better – more absolute – instruments than these, the issue cannot really be resolved. Nevertheless, international instrument intercomparisons (e.g., the International Comparison of Absolute Gravimeters) are organized regularly and are valuable.

In Finland, regular absolute gravimetric measurements have been made besides in Metsähovi, also in Vaasa (two points), Joensuu (two points), Kuusamo, Sodankylä, Kevo and Eurajoki.

### 10.5 The superconducting gravimeter

This gravimeter type is based on a superconducting metal sphere levitating on a magnetic field, the precise place of which is measured electronically. Because a superconducting material is impenetrable by a magnetic field, the sphere will remain forever in the same spot inside the field. Of course the field itself must be constant; it is generated by superconducting solenoids inside a vessel made of mu-metal which keeps out the Earth’s magnetic
The superconducting gravimeter


Superconductivity in these applications still demands working at the temperature of liquid helium (He). For this reason the device is not only expensive, but requires also an expensive laboratory in an environment where societal infrastructure works.

The number of these instruments in the world is over twenty. One GWR 20 type instrument has worked from 1994 in Kirkkonummi at the Metsähovi research station of the then Finnish Geodetic Institute, now the National

![Figure 10.8. International intercomparison of absolute gravimeters (© 2003 EGCS, Luxembourg).](image)

![Figure 10.9. Principle of operation of a superconducting gravimeter. Reading out the sphere position is done capacitively.](image)

The most important property of a superconducting gravimeter is, in addition to its precision\(^\text{14}\), its stability, i.e., its small drift. For this reason it is extremely suited for monitoring long period phenomena, like the free oscillations of the solid Earth after large earthquakes\(^\text{15}\). Thus it is suitable for measurements that are unsuitable for an ordinary gravimeter because of its larger drift and poorer sensitivity, and measurements for which a seismometer is unsuited because the frequencies are too low.

A recent trend is the development of lightweight, “portable” and remotely controllable superconducting gravimeters, e.g., the GWR iGrav, which weighs 30 kg and doesn’t consume any liquid helium at all. On the other hand it needs over a kilowatt in grid power for its refrigeration system\(^\text{16}\). Perhaps this will lead to improvement over the current situation where the bulk of instruments is located in Europe and North America.

10.6 Atmospheric influence on gravity measurement

The atmosphere has the following two effects on gravity:

1. **Instrumental effects.** These are due to the way the gravimeter is constructed. By putting the instrument in a pressure chamber, one could make these effects go away. In practice it is easier to calibrate the instrument (in the laboratory) and calculate a correction term according to the calibration certificate to be applied to the field measurements.

2. **Attraction of the atmosphere.** This is real gravitation. However, it causes an irregular local variation of gravity that we would rather be without. The effect of the atmosphere can be evaluated with the aid of the Bouguer plate approximation: if air pressure is \(p\), then the surface

\(^{14}\)Virtanen (2006) reports how the instrument at Metsähovi detected the change in gravity as workmen cleared snow from its laboratory roof, including tea break! “Weighing” visitors to the lab by their gravitational attraction is also standard fare.

\(^{15}\)Their periods range from under an hour to over twenty hours, and they are of considerable geophysical interest.

\(^{16}\)http://catalog.gwrinstruments.com/item/gravity-meters/superconducting-gravimeter-for-portable-operation/item-1001?
mass density of the atmosphere is

\[ \kappa = \frac{p}{g}, \]

where \( g \) is a representative gravity value inside the atmosphere. We don’t make a very large error by assuming

\[ g \approx 9.8 \text{ m/s}^2, \]

giving us on sea level \( \kappa \approx 10,000 \text{ kg/m}^2 \). The effect of the Bouguer plate is

\[ -2\pi G\kappa = -0.43 \text{ mGal}. \]

Variations in air pressure affect proportionally. If the air pressure disturbance is \( \Delta p = p - p_0 \), where \( p_0 \) is mean air pressure, 1015 hPa, its effect on gravity measurement will be

\[ \delta g_A = -0.43 \frac{\Delta p}{p_0} \text{ mGal}. \]

During the passage of a storm or weather front, this beautiful theory collapses, and simple equations give misleading results. Then it is best to just not do any gravity measurements!

3. Including the atmosphere into the mass of the Earth. This is not a correction to be applied to gravity measurements. It is a correction which is used in the calculation of gravity anomalies, if we want anomalies from which the effect of the atmosphere has been removed.

Remember that the reference gravity field of GRS80 is defined in such a way, that the parameter \( GM \) contains the whole mass of the Earth including atmosphere; i.e., the gravitational field as satellites are observing it (Heikkinen, 1981). Therefore, also when calculating gravity anomalies \( \Delta g \), one should reduce gravity by computationally moving the whole atmosphere above the point of measurement to below the measurement point, e.g., to sea level.

The total mass of the atmosphere is

\[ M_A = 4\pi\kappa R^2 = 4\pi \frac{p}{g} R^2. \]

According to Newton its attraction is

\[ \frac{GM_A}{R^2} = 4\pi G \frac{p}{g}, \]

twice the atmospheric reduction given above. At sea level, the effect is 0.86 mGal. At height, the effect is

\[ 0.86 \frac{p}{p_0} \text{ mGal}, \]
where \( p \) and \( p_0 \) are the air pressures at height and at sea level, respectively.

### 10.7 Airborne gravimetry and GNSS

In the early years of the 1990s GPS, the Global Positioning System, has changed airborne gravimetry from a difficult technology to something completely operational. To understand this, one must know the principle of operation of airborne gravimetry.

An aircraft carries an airborne gravimeter, an instrument that, in the same way as a sea gravimeter, is strongly damped. The measurement is done automatically, generally using electrostatic compensation. The instrument is mounted on a stabilized platform that follows the local vertical.

During flight, the gravimeter measures total gravity on board the aircraft. This consists of two parts:

1. gravity proper – i.e., gravity as felt in a reference frame connected to the Earth’s surface –, and
2. The pseudo-forces caused by the inevitable accelerations of the aircraft even in cruise flight.

Attached to the aircraft are a number of GNSS antennas; with these, and a geodetic GNSS instrument, the motions of the aircraft can be monitored with centimetre accuracy. From these can then be calculated the pseudo-forces mentioned above under item 2.

If we measure the position of the plane (or instrument) \( x_i \) at moment \( t_i \), \( \Delta t = t_{i+1} - t_i \), we obtain estimated acceleration values as follows:

\[
\mathbf{a}_i \approx \frac{(x_{i+1} + x_{i-1} - 2x_i)}{(\Delta t)^2}.
\] (10.8)

When the measured acceleration is \( \Gamma_i \) and the direction of the local plumbline \( \mathbf{n}_i \), local gravity \( g_i \) follows:

\[
g_i = \Gamma_i - (\mathbf{a}_i \cdot \mathbf{n}_i).
\]

\(^{17}\)GNSS, Global Navigation Satellite Systems, comprise, besides GPS, also the satellite positioning systems of other countries, like the Russian GLONASS.
A critical matter in the whole method is the choice of the time constant $\Delta t$. It is best to choose it as long as possible; then, the precision of the calculated GNSS accelerations $a_i$ is as good as possible. Also the damping of the gravimeter is chosen in accordance with $\Delta t$, and the observations are filtered digitally: all frequencies above the bound $\Delta t^{-1}$ are removed, because they are largely caused by the motions of the aircraft.

In practice, often the high-frequency part removed from the signal is 10,000 times larger than the gravity signal we are after!

If the uncertainty (mean error) of one GNSS position co-ordinate measurement is $\sigma_x$ (and they don’t correlate!), then according to equation (10.8) the uncertainty of the vertical acceleration is

$$\sigma_a = \frac{\sigma_x \sqrt{6}}{(\Delta t)^2}.$$  

Making the time interval $\Delta t$ as long as possible without resolution suffering, requires a low flight speed. Generally a propeller aircraft or even a helicopter is used. Of course the price of the measurement grows with the duration of the flight – a helicopter rotor hour is expensive!

For the flight height $H$ we choose in accordance with resolution $\Delta x$:

$$H \sim \Delta x = v\Delta t,$$

where $v$ is the flight speed. The separation between adjacent flight lines is chosen similarly.

The first major airborne gravimetry project was probably the Greenland Aerogeophysics Project (Brozena, 1992). In this ambitious American-Danish project in the summers of 1991 and 1992, over 200,000 km was flown, all the time measuring gravity and the magnetic field, and the height of the ice surface using a radar altimeter (Ekholm et al., 1995).

After that, also other large uninhabited areas in the Arctic and Antarctic regions have been mapped, see Brozena et al. (1996), Brozena and Peters (1994). Already in subsection 8.6.2 we made mention of other large surveys. Activity continues, see Coakley et al. (2013), Kenyon et al. (2012). The method is well suitable for large, uninhabited areas, but also, e.g., for sea areas close to the coast where ship gravimeters would have difficulty navigating long straight tracks. In 1999 an airborne gravimetry campaign was undertaken over the Baltic, including the Gulf of Finland (J. Kääriäinen, personal comm.).
In addition to the economic viewpoint, an important advantage of airborne gravimetry is, that a homogeneous coverage by gravimetric data is obtained from a large area. The homogeneity of surface gravimetric data collected over many decades is difficult to guarantee in the same way. Also the effect of the very local terrain, which for surface measurements is a hard to remove systematic error source especially in mountainous terrain, (see section 5.3 on page 89), does not come into play for airborne gravimetry.

The operating principle of satellite gravimetry, e.g., GOCE (Geopotential and Steady-state Ocean Circulation Explorer) is similar. An essential difference is however, that the instrumentation on the satellite is in a state of weightlessness: \( \Gamma = 0 \) (in a high orbit, or when using an air drag compensation mechanism); or \( \Gamma \) is small and is measured using a sensitive accelerometer (in a low orbit, where air drag is noticable).

The greatest challenge in planning a satellite mission is choosing the flight height. The lowest possible height is some 150 km; at that height, already a tankload of fuel is needed, or the flight will not last long. However, the resolution of the measurements on the Earth surface is limited; e.g., the smallest details in the gravity field “seen” by the GOCE satellite are 50–100 km in diameter.

### 10.8 Measuring the gravity gradient

The acceleration of gravity \( g \) is the gradient of the geopotential \( W \). It varies with place, especially close to masses. We speak of the gravity-gradient tensor or Marussi\(^{18}\) tensor:

\[
M \equiv \begin{bmatrix}
\frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial x \partial y} & \frac{\partial^2}{\partial x \partial z} \\
\frac{\partial^2}{\partial y \partial x} & \frac{\partial^2}{\partial y^2} & \frac{\partial^2}{\partial y \partial z} \\
\frac{\partial^2}{\partial z \partial x} & \frac{\partial^2}{\partial z \partial y} & \frac{\partial^2}{\partial z^2}
\end{bmatrix}
W.
\]

We know that gravity increases going down, at least in free air. Going up, gravity diminishes, about 0.3 mGal for every metre of height.

In topocentric co-ordinates \((x, y, z)\), where \(z\) points to the zenith, this matrix

\(^{18}\)Antonio Marussi (1908–1984) was an Italian geodesist and mathematician.
is approximately

\[
M \approx \begin{bmatrix}
-0.15 & 0 & 0 \\
0 & -0.15 & 0 \\
0 & 0 & 0.3
\end{bmatrix} \text{ mGal/m},
\]

where \( \frac{\partial^2 W}{\partial z^2} = \frac{\partial}{\partial z} g_z \approx 0.3 \text{ mGal/m} \) is the standard value for the free-air gradient of gravity: Newton’s law gives for a spherical Earth (the minus sign is because \( g \) points downward while the \( z \) co-ordinate increases going up):

\[
g_z = -\frac{GM}{(R + z)^2}.
\]

Derivation gives

\[
\frac{\partial}{\partial z} g_z = 2 \frac{GM}{(R + z)^3} \frac{\partial (R + z)}{\partial z} = -\frac{2g_z}{(R + z)} \approx 3 \cdot 10^{-6} \text{ m/s}^2 / \text{m} = 0.3 \text{ mGal/m}.
\]

The quantities \( \frac{\partial^2 W}{\partial x^2} \) and \( \frac{\partial^2 W}{\partial y^2} \) again describe the curvatures of the equipotential surfaces in the \( x \) and \( y \) directions, equations (3.3), (3.4):

\[
\frac{\partial^2 W}{\partial x^2} = -\frac{g}{\rho_1} \text{ and } \frac{\partial^2 W}{\partial y^2} = -\frac{g}{\rho_2},
\]

where \( \rho_1 \) and \( \rho_2 \) are the radii of curvature in the \( x \) and \( y \) directions. Substitution \( \rho_1 = \rho_2 = R \approx 6378 \text{ km} \) yields

\[
\frac{\partial^2 W}{\partial x^2} = \frac{\partial^2 W}{\partial y^2} \approx -1.5 \cdot 10^{-6} \text{ m/s}^2 / \text{m} = -0.15 \text{ mGal/m}.
\]

The Hungarian researcher Loránd Eötvös did a number of clever experiments (Eötvös, 1998) in order to measure components of the gravity-gradient tensor with torsion balances built by him. The method continues to be in use in geophysical research, as the gravity gradient as a measured quantity is very sensitive to local variations in matter density in the Earth’s crust.

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 above tensor is now

\[
M \approx \begin{bmatrix}
-1500 & 0 & 0 \\
0 & -1500 & 0 \\
0 & 0 & 3000
\end{bmatrix} \text{ E}.
\]

Note that

\[
\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 W}{\partial z^2} \approx 0,
\]
the familiar Laplace equation. However, the equation is not exact: in a co-
ordinate system co-rotating with the Earth, the term for the centrifugal force,
$2\omega^2$, must be added, equation (3.1).

The gravity-gradient field of Sun and Moon is known on the Earth’s surface
as the tidal field, see section 13.1.
11. The geoid, mean sea level, sea-surface topography

11.1 Basic concepts

On the ocean, the geoid is on average at the same level as mean sea level, the surface obtained by removing from the instantaneous sea surface all periodic and quasi-periodic variations. These variations are, for example:

- tidal phenomena (caused by Sun and Moon); order of magnitude ±1 m, locally even more
- variations caused by air pressure variations (“inverted barometer effect”); typically of order decimetres, but up to metres under the cores of tropical cyclones
- wind pile-up, water being pushed by winds
- littoral seas: variation in the volume of sweet water flowing out from rivers into the sea
- eddies that are formed in the oceans in connection with, e.g., the Gulf Stream and the Agulhas Stream (“mesoscale eddies”) that may live for months, and inside of which the sea surface may be even decimetres above or below that of the surroundings
- the continual shifting of ocean currents from one place to another
- ENSO, El Niño Southern Oscillation, is a very long time scale, quasi-periodic weather phenomenon happening in the waters of the Pacific Ocean and the air above it, but affecting weather phenomena worldwide. The time scale of variability ranges from two to seven years.

If we remove all these periodic and quasi-periodic variations, we are left
with *mean sea level*. If the water of the seas was in a state of equilibrium, then this mean sea surface would be an equipotential surface, the *geoid*.

This is however not how things really are. Also mean sea level differs from an equipotential surface due to the following phenomena:

- permanent ocean currents cause, though the Coriolis force, permanent differences in mean water level
- also permanent differences in temperature and salinity cause permanent differences in mean water level, the latter, e.g., in front of the mouths of rivers.

These physical phenomena, among others, cause the so-called *sea-surface topography*, a permanent separation between sea surface and geoid.

A classical definition of the geoid is

> “the equipotential surface agreeing most closely with mean sea level.”

The practical problem with this definition is, that determining the geoid requires knowledge of mean sea level everywhere on the ocean. This is why many “geoid” models in practice don’t coincide with global mean sea level, but with some local (mean) sea level – and often only approximately.

*Mean sea level* in its turn is also a problematic concept. It is sea level from which has been computationally removed all periodic effects – but who can know if a so-called secular effect in reality is perhaps long period? A sensible compromise is the average sea level over 18 years – an important periodicity in the orbital motion of the Moon.

The *sea-surface topography* again is defined as that part of the difference between mean sea level and the geoid, which is permanent. Also here, the measure of permanency is the time series that are available; tide gauges have been widely operating already for about a century, when again many satellite time series (TOPEX/Poseidon and its successors) are just about two decades long. See figure 12.1.

### 11.2 Geoids and national height datums

A locally determined geoid model is generally *relative*. Locally, at the current state of the art, one has no access to global mean sea level at an acceptable
precision. This may change with technology development.

In general, a local geoid model is tied to a national height system, and the difference from the definition is thus the same as the difference of the national height system from global mean sea level. In the case of Finland, the difference is about one metre, mostly caused by the sea-surface topography in the Baltic (about 30 cm) and the North Atlantic.

In Finland, heights were determined for a long time in the N60 height system, which is tied to mean sea level in Helsinki harbour at the start of 1960. The reference bench mark however is located in nearby Kaivopuisto. Precise levelling disseminated heights from here all over Finland. Today, the Finnish height system is N2000, which is in principle tied to sea level in Amsterdam, but the reference bench mark in Finland is similarly located at the Metsähovi research station in the Kirkkonummi municipality, West of Helsinki.

At the beginning of 1960, the reference surface of the Finnish height system N60 was an equipotential surface of the Earth’s gravity field. However, due to post-glacial land uplift, that is no longer the case: the post-glacial land uplift varies from some four millimetres per year in the Helsinki area to some ten millimetres per year in the area of maximum land uplift near Ostrobothnia. This is the main reason why in Fennoscandia height systems have a “best before” date and must be modernized a couple of times per century.

Generally geoid maps for practical use, like the Finnish geoid model FIN2000 (figure 8.4) are constructed so, that they transform heights in the national height system, e.g., N60 heights (Helmert heights) above “mean sea level” to heights above the GRS80 reference ellipsoid. As, however, land uplift is a continuous process, it must be tied to a certain epoch, a point in time at which the GNSS measurements were done to which the original gravimetric geoid solution has been fitted. In the case of FIN2000 this was 1997.0 (Matti Ollikainen, several sources).

Strictly speaking then, FIN2000 is not a model of the geoid. A better name might be “transformation surface”. This holds true, in fact, for all national or regional geoid models that are built primarily for the purpose of enabling the use of GNSS in height determination ("GNSS levelling"). These “geoid-like surfaces” are constructed generally in this way, that

1. we calculate a gravimetric geoid model by using the Stokes method and Remove-Restore, e.g., by the FFT method;
2. we fit this geoid surface solution to a number of comparison points, where both the height from levelling (“above sea level”) and from the GNSS method (above the reference ellipsoid) are known. The fit takes place, e.g., by describing the differences by a polynomial function:

\[ \delta N = a + b (\lambda - \lambda_0) + c (\varphi - \varphi_0) + \ldots \]

or something more complicated, and solving the coefficients \( a, b, c \) from the geoid differences in the known points using the least-squares method.

11.3 The geoid and post-glacial land uplift

Global mean sea level is not constant. It rises slowly by an amount that, over the past century, has slowly grown. Over the whole 20th century, the rate has been 1.5 – 2.0 mm/\( a \), e.g., 1.6 mm/\( a \) (Wöppelmann et al., 2009). Over the last couple of decades, the rate has accelerated and is now some 3 mm/\( a \), see figure 12.1.

This value is called the eustatic rise of mean sea level. It is caused partly by the melting of glaciers, ice caps and continental ice sheets, partly by thermal expansion of sea water. A precise value for the eustatic rise is very hard to determine: almost all tide gauges used for monitoring sea level have their own vertical motions, and distinguishing these from the rise of sea level requires a highly representative geographic distribution of measurement locations. Especially the ongoing response of the solid Earth to the latest deglaciation, the so-called GIA (Glacial Isostatic Adjustment) is a global phenomenon that it only in the latest decades has been possible to observe by satellite positioning.

Because of eustatic sea-level rise, a distinction must be made between absolute and relative land uplift:

**Absolute land uplift** is the motion of the Earth’s crust relative to the centre of mass of the Earth. This land uplift is measured when using satellites, the orbits of which are determined in a reference system tied to the Earth’s centre of mass. E.g., satellite radar altimetry, GNSS positioning of tide gauges.

**Relative land uplift** is the motion of the Earth’s crust relative to mean sea level. This motion is measured by tide gauges, also called mareographs.
11.3. The geoid and post-glacial land uplift

**Geoid rise:** as the post-glacial land uplift is the shifting of masses internal to the Earth from one place to another, it is clear that also the geoid must change. The geoid rise is however small compared to the land uplift, only a few percent of it.

Equation (the point above a quantity denotes the time derivative \( \frac{d}{dt} \)):

\[
\dot{h} = \dot{H}_r + \dot{H}_e + \dot{H}_t + \dot{N},
\]

where

- \( \dot{h} \) is the absolute land uplift,
- \( \dot{H}_r \) is the relative land uplift,
- \( \dot{H}_e \) is the eustatic (mean sea level) rise,
- \( \dot{H}_t \) is the change over time of the sea-surface topography (probably small),
- \( \dot{N} \) is the geoid rise.

The change in the geoid as a result of land uplift can be simply calculated with the Stokes equation:

\[
\frac{dN}{dt} = \frac{R}{4\pi\gamma} \int_S S(\psi) \left( \frac{d}{dt} \Delta g \right) d\sigma.
\]

Here, \( \frac{d}{dt} \Delta g \) is the change of gravity anomalies over time due to land uplift. Unfortunately we do not precisely know the mechanism by which mass flows into the land uplift area in the Earth’s mantle; we may write

\[
\frac{d}{dt} \Delta g = c \frac{dh}{dt},
\]

where the constant \( c \) may range from \(-0.16\) to \(-0.31\) mGal/m.

- The value \(-0.16\) mGal/m is called the “Bouguer hypothesis”: it corresponds to the situation where upper mantle material flows into the hole forming underneath the rising Earth’s crust, in order to fill it.
- The value \(-0.31\) mGal/m is the opposite extreme, the “free-air hypothesis”. By this hypothesis, the ice load during the last ice age has only compressed the Earth’s mantle, and now it is slowly expanding again into its former volume (“rising dough model”).
Up until fairly recently, the most likely value was about $-0.2 \text{ mGal/m}$, with substantial uncertainty. The latest results (Mäkinen et al., 2010) give $-0.16 \pm 0.02 \text{ mGal/m}$ (one standard deviation), which would seem to settle the issue. It looks like the Bouguer hypothesis is closer to physical reality. The flow of mass happens probably within the asthenosphere.

This problem has been studied much in the Nordic countries. The method has been gravimetric measurement along the 63°N parallel (“Blue Road Geotraverse” project). The measurement stations extend from the Norwegian coast to the Russian border, and have been chosen so, that gravity along them varies within a narrow range. In this way, the effect of the scale error of the gravimeters is avoided. Clearly, absolute gravity is of no interest here, only the change in gravity differences over time between the stations.

These measurements have been made over many years using high precision spring or relative gravimeters. In recent years, there has been a shift to using absolute gravimeters, obviating the need for measurement lines.
11.4 Methods for determining the sea-surface topography

In principle three geodetic methods exist:

1. satellite radar altimetry and gravimetric geoid determination
2. GNSS positioning along the coast (tide gauges) and gravimetric geoid determination
3. precise levelling along the coast.

In addition to this, we still have the oceanographic method, i.e., physical modelling. The method is termed steric levelling if temperature and salinity measurements along vertical profiles are used on the open ocean; geostrophic levelling if ocean current measurements are used to determine the Coriolis effect, generally close to the coast.

All methods should give the same results. The Baltic Sea is a textbook example, where all three methods have been used. A result was that the whole Baltic Sea surface is tilted: relative to an equipotential surface, the sea surface goes up from the Danish straits to the bottoms of the Gulf of Finland and the Bothnian Bay by some $25 - 30$ cm.

Oceanographic model calculations show, that this tilt is mainly due to a salinity gradient: on the Atlantic, salinity is $30 - 35$ o/o, when in the Baltic it drops to $5 - 10$ o/o due to the massive production of sweet water (Ekman, 1992). Of course on top of this come temporal variations, like oscillations like in a bathtub, the amplitude of which can be over a metre.
Chapter 11. The geoid, mean sea level, sea-surface topography

In Ekman (1992) more is said about the sea-surface topography of the Baltic and its determination.

11.5 Global sea-surface topography and heat transport

One important reason why researchers are interested in the global sea-surface topography, is that it offers an opportunity to study more precisely the currents in the oceans and thus the transport of the Sun’s heat energy from the equator to higher latitudes. There are actually many things the study of which is helped by precise knowledge of ocean currents: carbon dioxide dissolved into the water, chlorophyll (phytoplankton), salinity, etc.

The Coriolis force (acceleration) caused by the Earth’s rotation is:

\[ \mathbf{a} = 2 \langle \mathbf{v} \times \mathbf{\omega} \rangle, \quad (11.1) \]

where \( \mathbf{v} \) is the vector of motion in a system attached to the rotating Earth, and \( \mathbf{\omega} \) is the rotation vector of the Earth.

If a fluid flows on the Earth’s surface, then, in the above equation (11.1) only the part of \( \mathbf{\omega} \) normal to the surface will have an effect: this part has a length of \( \langle \mathbf{\omega} \cdot \mathbf{n} \rangle = \omega \sin \varphi \), and the vector equation (11.1) may be replaced by a simpler scalar equation:

\[ a = 2v \omega \sin \varphi, \]

where \( a \doteq \| \mathbf{a} - \langle \mathbf{a} \cdot \mathbf{n} \rangle \| \), i.e., the length of the projection of \( \mathbf{a} \) onto the tangent plane to the Earth, and \( v \doteq \| \mathbf{v} \|, \omega \doteq \| \mathbf{\omega} \| \) etc. in the familiar way. The direction of the Coriolis acceleration is always perpendicular to the flow velocity: when watching along the flow direction, to the right on the Northern hemisphere, to the left on the Southern hemisphere.

As a result of the Coriolis force, the sea surface in the area of an ocean current will be tilted by an angle

\[ \frac{a}{\gamma} = \frac{2v \omega}{\gamma} \sin \varphi. \]

This equilibrium between Coriolis force and the horizontal gradient of pressure is called the geostrophic equilibrium. On the equator it can be seen that the tilt is zero, but everywhere else, ocean currents are tilted. E.g., in case of the Gulf Stream, the height variation caused by this effect is several decimetres. If we define a local \((x, y)\) co-ordinate system where \( x(\varphi, \lambda) \) is pointing
11.5. Global sea-surface topography and heat transport

North and y (φ, λ) East, we may write for the sea-surface topography $H$:

$$\frac{\partial H}{\partial x} = -2v_y \frac{\omega}{\gamma} \sin \varphi, \quad \frac{\partial H}{\partial y} = +2v_x \frac{\omega}{\gamma} \sin \varphi,$$

(11.2)

As we will see in chapter 12, we can measure the location in space of the sea surface at this precision using satellite radar altimetry. If we furthermore have a precise geoid map, we may calculate the sea-surface topography, and with the aid of equations (11.2) solve for the flow velocity field

$$\begin{bmatrix} v_x (x, y) \\ v_y (x, y) \end{bmatrix}^T = \begin{bmatrix} v_x (\varphi, \lambda) \\ v_y (\varphi, \lambda) \end{bmatrix}^T.$$  

An elegant property of these equations is, that we don’t even have to know the absolute level of the field $H (x, y) = H (\varphi, \lambda)$ because that vanishes in differentiation.

The described method requires a sufficiently precise geoid map of the world’s oceans. To this need, the GOCE satellite fits like a glove, see section 12.7. One objective of the mission was, as the name indicates, to get a full picture of ocean currents and especially their capacity for heat transport. This knowledge helps understand how the Earth’s climate functions and how it is changing, also as a result of human activity. This is for Europe, and Finland, a survival issue, as these areas are habitable only thanks to the thermal energy brought in by the Gulf Stream.

Even without a geoid model, we can study, using satellite altimetry, the variations of ocean currents. It has been known for long that in the North

\footnote{A popular, though unofficial, unit for ocean current is the Sverdrup (\url{http://en.wikipedia.org/wiki/Sverdrup}), a million cubic metres per second. All the rivers of the world together make about one Sverdrup, while the Gulf Stream is 30 – 150Sv.}
Atlantic, *mesoscale eddies* have been moving alongside the Gulf Stream, eddies of size $10 - 100$ km which show up in altimetric imagery. Interesting is that the eddies also show up in maps of the ocean surface temperature, and biologists have observed that life inside the eddies differs from that outside (Godø et al., 2012). The life span of the eddies can be weeks, even months.

A good, though somewhat dated, introduction into “geodetic oceanography” is given by Rummel and Sansó (1992).

### 11.6 The global behaviour of sea level

Water exists on the Earth in three phases: liquid, ice, and vapour. During geological history, especially the ratio between liquid water and ice has varied substantially. Also today, a large amount of ice is tied up in continental ice sheets, specifically Antarctica and Greenland. Of these, Antarctica is the overwhelmingly largest.

When the amount of water tied up in continental ice sheets varies, so does sea level. The end of the last ice age has raised sea level by as much as $120$ m, a process that ran to completion some $6000$ years ago$^2$ (http://en.wikipedia.org/wiki/Current_sea_level_rise). Not until the last

---

$^2$6000 years “before present”, 6 ka BP. BP conventionally means: before 1950. Nowadays is also used $b2k$, before the year 2000.
century or two has sea level again started rising, and the rise accelerating, as a consequence of global warming.

We still live in the aftermath of the last glaciation; there were large continental ice sheets which have since molten away, like in Fennoscandia and Canada (the so-called Laurentide ice sheet), the land is still rising at an even pace, in places even 10 millimetres per year. Around the land uplift areas, in central Europe and the United States, again takes place a subsidence of the land at an annual pace of 1−1.5 mm, as directly underneath the hard crust of the Earth or lithosphere, in the upper mantle layer called asthenosphere, material is flowing slowly inward under the rising Earth’s crust.

In order to complicate the picture, the sea-level rise caused by the melting of continental ice sheets also presses the ocean floor down – by as much as 0.3 mm per year, the so-called Peltier effect (Peltier, 2009). Therefore the measured sea-level rise – whether on the coast by tide gauges, or from space using satellite altimetry – does not represent the whole change in total ocean water volume. If that is what interests us, as it always does in climate research, this Peltier correction must still be added to the observation values.

The subsidence of the sea floor hasn’t even been globally uniform: at the edges of the continents happens a “cantilever motion” when the sea floor subsides but dry land doesn’t. And in the tropics in the Indian and Pacific Oceans, sea level reached 6000 years ago its maximum level, the so-called mid-Holocene highstand, relative to the Earth’s crust; after this, local sea level has subsided and the coral formations from that age have remained, dead, some 2−3 m above modern sea level. This is how, e.g., Tuvalu and the Maldives were formed, which are now again being threatened by modern sea-level rise.

11.7 The sea-level equation

Scientifically the variations in sea level are studied using the sea-level equation (http://samizdat.mines.edu/sle/sle.pdf). A pioneer in this field has been W. Richard Peltier (http://www.atmosp.physics.utoronto.ca/~peltier/data.php), who has constructed physics based models of how both the solid Earth and sea level respond when the total mass of the continental ice sheets changes.
The sea-level equation is (http://samizdat.mines.edu/sle/sle.pdf):

\[ S = S_E + \frac{\rho_i}{\gamma_0} \left[ G_s \otimes I - G_s \otimes I \right] + \frac{\rho_o}{\gamma_0} \left[ G_s \otimes_o S - G_s \otimes_o S \right], \]  

(11.3)

where

- \( S = S(\omega, t) = S(\varphi, \lambda, t) \) describes the variations of sea level as a function of place \( \omega = (\varphi, \lambda) \) and time \( t \),

- \( I = I(\omega, t) \) is similarly a function of place and time describing the geometry of ice sheets and glaciers,

- \( S_E \) is the eustatic term, i.e., the variation in ice volume converted to “equivalent global sea-level variation”, in an equation

\[ S_E(t) = \frac{m_i(t)}{\rho_o A_o}, \]

where \( m_i(t) \) is the variation in total ice mass as a function of time, \( \rho_o \) the density of sea water, and \( A_o \) the total surface area of the oceans,

- \( \rho \) is the density of matter: \( \rho_i \) that of ice, and \( \rho_o \) that of the ocean,

- \( \otimes \) is the symbol of a convolution on the surface of the Earth and the time axis, \( \otimes_i \) over land ice, \( \otimes_o \) over the oceans – i.e., Green’s function is multiplied with the ice and sea functions and integrated over the domain in question. These integrals are by the way very similar to the ones discussed in section 7.1, e.g.:

\[ \{ G_s \otimes_o S \}(\omega, t) = \int_{-\infty}^{t} \int_{\text{ocean}} C_s \{ \psi(\omega, \omega'), (t - t') \} S(\omega', t') \, d\omega' \, dt', \]

where \( \psi(\omega, \omega') \) is the geocentric angular distance between evaluation point \( \omega = (\varphi, \lambda) \) and integration point \( \omega' = (\varphi', \lambda') \). The surface integral is \( d\omega = N(\varphi) M(\varphi) \cos \varphi \, d\varphi \, d\lambda \), where \( N, M \approx R \) are the principal
The sea-level equation

The radii of curvature of the Earth ellipsoid. As can be seen, we have here a convolution applied both over the Earth’s surface \( \omega \) and over the time axis \( t \).

- The overbar designates taking the average over the whole ocean surface,
- \( \gamma_0 \) is an average acceleration of gravity,
- \( G_s \) is the so-called Green’s function of sea level:

\[
G_s = G_V - \gamma_0 G_u,
\]

where the Green’s function of the geopotential is

\[
G_V (\psi, t) = G^r_V (\psi, t) + G^e_V (\psi, t) + G^v_V (\psi, t)
\]

where again \( \psi \) is the distance of evaluation point from integration point, and \( G^r_V, G^e_V \) ja \( G^v_V \) are the rigid, elastic and plastic (“viscous”) partial Green’s functions of deformation. These thus describe the rheological behaviour of the Earth, and their theoretical calculation requires the internal viscosity distribution \( \eta (r) \) of the Earth, assuming it is isotropic, i.e., only dependent upon \( r \).

\[
G_u (\psi, t) = G^u_r (\psi, t) + G^u_v (\psi, t)
\]

is again similarly Green’s function function of vertical displacement, in the same way split into elastic and plastic parts.

The behaviour of sea level can now be computed in this way, that one first tries to construct an “ice-load history”, i.e., \( I (\omega, t) \); Then, from this one tries to calculate iteratively, using the sea-level equation (11.3), \( S (\omega, t) \). Note that \( S \) describes relative sea-level variation, i.e., changes in the relative positions of sea level and the Earth’s solid body or Earth’s crust. It is a function of place: one may not assume that it would be the same everywhere. In the article Mitrovica et al. (2001) it is shown how, e.g., the melting water from Greenland flees to the Southern hemisphere, when the melting water from Antarctica again comes similarly to the North. This is a consequence from the change in the Earth’s gravity field and the geoid, when large volumes of ice melt. Another factor is that also the figure of the Earth changes, when the ice load changes: so-called Glacial Isostatic Adjustment, GIA.

This also complicates the monitoring of global mean sea level from local measurements: the problem is familiar in Fennoscandia, as the Earth’s crust, for now, moves up faster than global sea-level rise...
The geoid, mean sea level, sea-surface topography

Green’s functions in the sea-level equation are functions of both $\psi$ and time $t$; this tells us that GIA is a function both of place and time. On a spherically symmetric Earth, the functions may be written as expansions, e.g.\(^3\)

\[
G_V(\psi, t) = H(t) \frac{R \gamma_0}{M} \sum_{\ell=1}^{\infty} \left( \sum_{i=1}^{l} k_{\ell i} e^{-s_{\ell i} t} \right) P_\ell(\cos \psi),
\]

where $H(t)$ is a step function (the “Heaviside function”). The index $i$ counts the so-called viscous relaxation modes for every degree number $\ell$; $k_{\ell i}$ are “viscous loading deformation coefficients” and $\tau_{\ell i} = 1/s_{\ell i}$ correspond to the relaxation times in which the mode in question will decay over time. Generally the modes that are of large spatial extent – i.e., low $\ell$ values – decay slower, when again the local modes – high $\ell$ values – tend to decay faster, and the local modes of the last deglaciation have today already vanished.

E.g., the geographic pattern of the Fennoscandian land uplift is already very smooth, and the seismicity accompanying the deglaciation is pretty much over. Back then, immediately after the retreat of the ice sheet at its edge, there were strong earthquakes, the traces of which are visible in the landscape (Kuivamäki et al., 1998). The now dominant viscoelastic modes are many hundreds of kilometres in geographic extent, and correspondingly of time scales thousands of years.

\(^3\)We consider here only the plastic or viscous deformation.
12. Satellite altimetry and satellite gravity missions

12.1 Satellite altimetry

*Satellite altimetry* measures, using microwave radar, the distance from a satellite straight downward to the sea surface. Over time, there have been many satellites carrying an altimetry radar, see table 12.1.

- The GEOS-3 (1975-027A) and Seasat satellites were American testing

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Launch year</th>
<th>Orbital inclination (°)</th>
<th>Orbit height (km)</th>
<th>Repeat periods (days)</th>
<th>Precision (m)</th>
<th>Positioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEOS-3</td>
<td>1975</td>
<td>115.0</td>
<td>843</td>
<td>–</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Seasat</td>
<td>1978</td>
<td>108.0</td>
<td>780</td>
<td>3 (17)</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Geosat</td>
<td>1985</td>
<td>108.0</td>
<td>780</td>
<td>3, 17</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>ERS-1</td>
<td>1991</td>
<td>98.5</td>
<td>780</td>
<td>3, 35, 2 × 168</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>TOPEX-Poseidon</td>
<td>1992</td>
<td>66.0</td>
<td>1337</td>
<td>10</td>
<td>0.033</td>
<td>GPS</td>
</tr>
<tr>
<td>ERS-2</td>
<td>1995</td>
<td>98.5</td>
<td>780</td>
<td>35</td>
<td>0.03</td>
<td>PRARE</td>
</tr>
<tr>
<td>Geosat follow-on</td>
<td>1998</td>
<td>108</td>
<td>800</td>
<td>17</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>Envisat</td>
<td>2001</td>
<td>98.5</td>
<td>784</td>
<td>35</td>
<td>0.045</td>
<td>GPS</td>
</tr>
<tr>
<td>Jason-1</td>
<td>2001</td>
<td>66.1</td>
<td>1336</td>
<td>9.9156</td>
<td>0.025</td>
<td>GPS</td>
</tr>
<tr>
<td>Jason-2</td>
<td>2008</td>
<td>66</td>
<td>1336</td>
<td>9.9156</td>
<td>0.025</td>
<td>GPS</td>
</tr>
<tr>
<td>Cryosat-2</td>
<td>2010</td>
<td>92.0</td>
<td>725</td>
<td>369, 30</td>
<td>0.085</td>
<td>DORIS, GPS</td>
</tr>
<tr>
<td>HY-2A</td>
<td>2011</td>
<td>99.3</td>
<td>970</td>
<td>14, 168</td>
<td>0.085</td>
<td>DORIS, GPS</td>
</tr>
<tr>
<td>SARAL/AltiKa</td>
<td>2013</td>
<td>98.5</td>
<td>781</td>
<td>35</td>
<td></td>
<td>DORIS</td>
</tr>
</tbody>
</table>
Chapter 12. Satellite altimetry and satellite gravity missions

satellites aimed at developing the altimetric technique. The measurement precision of GEOS-3 was still rather poor. Before that, altimetry was also tested with a device on board Skylab (1973-027A).

- Seasat (1978-064A) broke down only three months after launch. However, the data from Seasat was the first large satellite altimetry data set used for determining the mean sea surface, also of the Baltic Sea.

- Geosat (1985-021A) was a satellite launched by the U.S. Navy, intended to map the gravity field on the world’s oceans, more precisely the deflections of the plumbline, which are needed to impart the correct departure direction to ballistic missiles launched from submarines. The 17-day repeat data from the geodetic mission was initially classified; then, the Southern hemisphere data was published for scientists to use, and currently the whole data set is public.

- The ERS-1/2 (1991-050A, 1995-021A) and Envisat (2002-009A) satellites were launched by ESA, the European Space Agency. The altimeter was just one among many packages. On the ERS satellites, a German positioning device called PRARE was along, but only on ERS-2 it functioned after launch.

- TOPEX/Poseidon (1992-052A) was an American-French collaboration, one goal of which was to precisely determine the sea-surface topography. A special feature was the on-board precise GPS positioning device, which allowed the determination of the location of the sea surface geocentrically. Together with its successors Jason-1 and 2, this satellite mission has also produced, and continues to produce, valuable information on the global rise of sea level over the last 20 years, about 3 mm per year. See figure 12.1.

The famous oceanographer Walter Munk described TOPEX/Poseidon in 2002 as “the most successful ocean experiment of all time” (http://en.wikipedia.org/wiki/TOPEX/Poseidon).

- HY-2A (2011-043A) is a Chinese satellite also launched by China.

- SARAL/AltiKa (2013-009A) is a satellite launched by India. The altimeter and DORIS are French contributions.

- Cryosat-2 (2010-013A) is a satellite launched by the European Space Agency (ESA) to study polar sea ice. Especially of interest is the so-called freeboard, the amount by which the ice sticks out of the water.
From this the thickness, and with surface area, the total volume may be calculated. The launch of Cryosat-1 failed. In-orbit positioning is done with the French DORIS system.

The measurement method of satellite radar altimetry is depicted in figure 12.2. Here we see all the quantities that are along in altimetry: the measured range \( \ell \) is the height \( h \) of the satellite from the reference ellipsoid, corrected for the geoid height \( N \), sea-surface topography \( H \), and variations of the sea surface, like tides, eddies, annual periods etc.

Furthermore, if the satellite does not contain a precise positioning device, the true orbit of the satellite will differ from the calculated (even after the fact!) one. Therefore

\[
h_{\text{sat}} = h_{0,\text{sat}} + \Delta h,
\]

where \( h_{0,\text{sat}} \) is the calculated orbit, and \( \Delta h \) the orbit-error correction.

The measurements are performed by sending 10 – 20 pulses per second down; the travel times of the reflected return pulses are measured, the largest and smallest values are thrown away (as possibly erroneous), and from the remainder, a mean value is calculated for the central epoch of the pulse train using linear regression. The value thus obtained from the regression line is the actual “measurement”; one every second, making the measurement frequency 1 Hz.

The details will vary from satellite to satellite. The pulse shape is never quite crisp; the place of the reflection on the ocean surface, or footprint, has a diameter of several kilometres. Especially if the ocean has wave motion (significant wave height, SWH), then in the processing phase one should make...
careful instrument corrections so no biases are created: if the SWH is large, also the altimeter footprint (the area on the sea from which radio energy returns to the receiver) will be larger, and the distance travelled by the radio waves longer on average.

Of all the corrections related to instrumentation, atmosphere, ocean, and solid Earth, we mention:

- the height of sea waves (SWH)
- solid-Earth tides
- ocean tides
- the “wet” tropospheric propagation delay, best computed from water vapour content measured with a water vapour radiometer on the satellite, otherwise from an atmospheric model
- the “dry” tropospheric propagation delay
- the ionospheric delay, only for the part of the ionosphere below the satellite, depending on flight height
- the altimeter’s own calibration correction. Nowadays one always strives
for “in-flight” calibration.

The measurements and all corrections to be made to them are collected into a “geophysical data record” (GDR), one per observation epoch. The files built this way are distributed to researchers. This allows all kind of experimentation, e.g., the replacement of a correction by one calculated from improved models, etc.

### 12.2 Crossover adjustment

When a satellite orbits the Earth over months or even years, thousands of points are formed where the tracks cross each other. If we assume that sea level is the same for both satellite overflights, then this forms a condition that can be used to adjust orbit errors.

Observation equations:

\[ h_a = N + H + \Delta h + \epsilon + n, \]

where \( h_a \) is the altimetric measurement of the height of the sea surface, \( N \) is the geoid height, \( H \) is the sea-surface topography (the permanent deviation of the sea surface from an equipotential surface), \( \Delta h \) is the orbit-error correction, \( \epsilon \) is the variability of the sea surface due to, e.g., the tides, and \( n \) is the noise in the radar observations.

From this we obtain in the crossing point of tracks \( i \) and \( j \):

\[ \ell_k \equiv h^i_a - h^j_a = (\Delta h_i - \Delta h_j) + (\epsilon_i - \epsilon_j) + (n_i - n_j). \]

Here we see the complication that in crossover adjustment, both sea-surface variability and orbit errors are along in the same equation.

If we forget for now the sea-surface variability (or assume that it behaves randomly, i.e., it is part of the noise \( n \)), we may write

\[ \ell_k = \Delta h_i - \Delta h_j + n_k, \]

the observation equation of crossover adjustment. The index \( k \) counts crossover points, the indices \( i,j \) count tracks. Next, we choose a suitable model for the satellite orbit error. The simplest choice, sufficient for a small area, is the assumption that the orbit error is constant. See a simple example, figure 12.3.
In the figure we have three tracks and two crossing points. The observation equations, which describe the discrepancies in the known crossover points as functions of the orbit errors, are

\[ \ell_1 = \Delta h_2 - \Delta h_3 + n_1 \]
\[ \ell_2 = \Delta h_1 - \Delta h_3 + n_2 \]

or in matrix form\(^1\)

\[
\begin{bmatrix}
\ell_1 \\
\ell_2 \\
\end{bmatrix} =
\begin{bmatrix}
0 & 1 & -1 \\
1 & 0 & -1 \\
\end{bmatrix}
\begin{bmatrix}
\Delta h_1 \\
\Delta h_2 \\
\Delta h_3 \\
\end{bmatrix} +
\begin{bmatrix}
n_1 \\
n_2 \\
\end{bmatrix}.
\]

Symbolically

\[ \ell = A \hat{x} + \mathbf{n}. \]

If you now try to calculate the solution with ordinary least squares,

\[ \hat{x} = \left( A^T A \right)^{-1} A^T \ell, \]

\(^1\)Note the similarity with the observation equations for levelling! Instead of bench marks, we have tracks, instead of levelling lines, crossover points.
you will notice that you cannot. The matrix $A^T A$ is singular (check!). This makes sense, as one can move the whole track network up or down without the observations $\ell_k$ changing. No unique solution can be found for such a system.

Finding a solution requires that something must be fixed. E.g., one track, or, more democratically, the mean level of all tracks. This fixing is achieved by adding the following “observation equation”:

$$\ell_3 = 0 = \begin{bmatrix} c & c & c \end{bmatrix} \cdot x,$$

where $c$ is some suitable constant. Then the matrix $A$ becomes

$$A = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ c & c & c \end{bmatrix},$$

and

$$\begin{bmatrix} \Delta \hat{h}_1 \\ \Delta \hat{h}_2 \\ \Delta \hat{h}_3 \end{bmatrix} = (A^T A)^{-1} A^T \ell = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{3^2} \end{bmatrix} \begin{bmatrix} 0 & 1 & c \\ -1 & -1 & c \end{bmatrix} \begin{bmatrix} \ell_1 \\ \ell_2 \\ 0 \end{bmatrix} =$$

$$= \frac{1}{3} \begin{bmatrix} -1 & 2 \\ 2 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix},$$

a unique solution from which $c$ has vanished.

An alternative representation of orbit errors more suitable for use in a larger area, is a linear function:

$$\Delta h = a + b \tau,$$

where the parameter $\tau$ is the place along the track reckoned from its starting point. The dimension of this place can be time (seconds) or angular distance (degrees). Now the set of observation equations for the situation described above is (note the notation: $\tau_k$, $k$ the number of the observation, i.e., the crossover, $i$ the number of the unknown, i.e., the track):

$$\begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & \tau_1^2 & -1 & -\tau_1^3 \\ 1 & \tau_2^1 & 0 & 0 & -1 & -\tau_2^3 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \\ a_3 \\ b_3 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}.$$

Note that $\hat{x} = A^{-1} \hat{\ell}$ would in this case have produced the same solution, as $A$ is square and invertible.
Of course also this system will prove to be singular. Removing the singularity can be done by fixing all three parameters $b$ and one parameter $a^3$.

The phenomenon that no solution can be found unless something is fixed, is called a *datum defect*. Fixing something suitable will define a certain *datum*. Between different datums exists a *transformation formula*, e.g., in the case of one orbit error parameter per track, this transformation is a simple *translation* of all tracks up or down.

The situation is somewhat similar as when defining a height system, vertical reference system, for a country: one has to fix one point, e.g., Helsinki harbour. If one fixes another point, e.g., Turku harbour, the result is another *datum*, in which all height values differ from the corresponding ones in the first datum by a certain fixed amount.

The argument continues to hold if there is a large number of tracks: say, ten Northgoing and ten Southgoing tracks, crossing in $10 \times 10$ crossover points. Here, for two parameters per track, we would have 40 unknowns and no less than 100 observations. Still, we must constrain the absolute level and the various trends and possible other deformations of the whole network of tracks. It gets complicated, but a simple approach is to attach *a priori* uncertainties to the unknowns $a_i, b_i$ to be estimated, e.g., the known uncertainties of the orbit prediction available. The least-squares adjustment equation then becomes

$$\hat{x} = \left( A^T A + \Sigma^{-1} \right)^{-1} A^T \ell,$$

where $\Sigma$ is the diagonal matrix containing the *a priori* variances $\sigma^2_{a,i}, \sigma^2_{b,i}$ of the parameters of each track $i$. This is referred to as Tikhonov\(^4\) regularization.

### 12.2.1 Example

In diagram 12.4 describing a satellite altimetry geometry, there are 16 crossover points. We attempt a crossover adjustment.

\(^3\)In order to understand this, build, e.g., a three track “wire-frame model” from pieces of iron wire, tied together by pieces of string at the crossover points. Crossover conditions don’t in any way fix the values of the trends $b$, and the whole absolute level of the frame continues to be unconstrained.

\(^4\)Andrey Nikolayevich Tikhonov (1906–1993) was a Russian Soviet mathematician and geophysicist.
Questions:

1. If the orbital error $\Delta h$ of each satellite track is described by a model with a single bias term, how many unknowns are there?

2. If we have available 16 "observations", i.e., crossover differences, how many of them are redundant?

3. Is it geometrically possible to calculate this network?

4. If we fix one track in advance (so-called *a priori* information), how many redundant observations are there? Can this network be calculated?

5. If every track has two unknowns, a bias as well as an error growing linearly with time, i.e., a "trend" or "tilt", what then needs to be fixed in order to make the network calculable? How many redundancies are there then?

6. If, in case (3), we fix one track, which one would you choose? Propose a solution where you do not have to make a choice.

Answers:

1. As many as there are tracks: 8.
2. $16 - 8 = 8$.

3. No, because the absolute level of the whole network is indeterminate.

4. $16 - (8 - 1) = 9$. Now the network can be calculated.

5. If we assume that the tracks are “straight” in $(x, y)$ co-ordinates, then the set of allowable transformations on the whole network is

$$\Delta h = a_{00} + a_{10}x + a_{01}y + a_{11}xy$$

with four degrees of freedom. So, fix one bias and three trends, not all North- or all Southgoing.

6. Any such choice would be arbitrary (however, the two parallel-track trends should be chosen far apart). Rather use the method described above instead, Tikhonov regularization.

### 12.2.2 Global crossover adjustment

In a global crossover adjustment, often a still more sophisticated model is used,

$$\Delta h = a + b \sin \tau + c \cos \tau,$$  \hspace{1cm} (12.1)

where now $\tau$ is an angular measure, e.g., the place along the track measured from the last South-North equator crossing. See Schrama (1989), where this problem is treated more extensively. In this model, $a$ represents the size of the orbit, while $(b, c)$ describe the offset of the centre of the orbit from the geocentre. This model is three-dimensional: the orbital arcs with their crossovers form a spherical network surrounding the Earth. The degrees of freedom left by the crossover conditions are now the size of this sphere and the offset of its centre from the geocentre: with $(X, Y, Z)$ geocentric co-ordinates, we have

$$\Delta h = a_0 + a_1X + a_2Y + a_3Z$$  \hspace{1cm} (12.2)

with four degrees of freedom\(^5\).

\(^5\)One could argue that, in eq. (12.1), the parameter $a$ should be zero, as Kepler’s third law allows a very precise determination of the orbital size, see section 12.3. Then, also $a_0 = 0$ in eq. (12.2).
12.3 Choice of satellite orbit

In choosing a satellite orbit, Kepler’s orbital laws are central. Kepler’s third law says:

\[ GM P^2 = 4\pi^2 a^3, \]  

(12.3)

where \( a = a_E + h \) is the satellite orbit’s semi-major axis (i.e., the mean distance from the geocentre), while \( h \) is called the satellite’s mean height. \( P \) is the orbital period, \( a_E \) the equatorial radius of the Earth.

From equation (12.3) one can already infer that using satellite observations one can precisely determine the quantity \( GM \). The period \( P \) is precisely measurable from long observation series; also the size of the orbit \( a \) can be obtained very precisely, e.g., from satellite laser ranging (SLR) observations. For this purpose have been used, e.g., the well known Lageos (Laser Geodynamic Satellite) satellites, which orbit the Earth at a height of 6000 km.

The orbits of altimetry satellites are chosen much lower, as is seen from table 12.1 at the start of the chapter. The height is fine tuned using thrusters, so that the satellite passes over the same place, e.g., once a day, after 14 orbital periods. Alternatively one chooses an orbit that flies over the same place every third, seventeenth, 168th day... this is called the repeat period.

The choice of the repeat period depends on the mission objective:
• if one wishes to study the precise shape of the mean sea surface, one chooses a long repeat period, in order to get the tracks as close together as possible on the Earth’s surface;

• if one wishes to study the variability of the sea surface, one chooses an orbit that returns to the same location after a short time interval. Then, the grid of tracks on the Earth’s surface will be sparser.

Also parameters describing the figure of the Earth affect satellite motion, e.g., the quantity $J_2$, the dynamic flattening, having a value of $J_2 = 1082.6267 \cdot 10^{-6}$. It is just one of many so-called spherical harmonic coefficients that describe the figure of the Earth and affect satellite orbits. In the case of $J_2$ the effect is, that the plane of the satellite orbit rotates at a certain rate (orbital precession), which make the satellite, if she flies over the same location the next day, doing so several minutes earlier. The equation is, for a circular orbit of radius \( a \):

$$\frac{d\Omega}{dt} = -\frac{3}{2} \sqrt{\frac{GM a_E^2}{a^3}} J_2 \cos i,$$

where again \( a_E \) is the equatorial radius of the Earth, and \( i \) the inclination of the orbital plane relative to the equator. If we substitute numerical values into this, we obtain

$$\frac{d\Omega}{dt} = -1.31895 \cdot 10^{18} \frac{\cos i}{(a_E + h)^{3.5}} \left[ \text{m}^{3.5} \text{s}^{-1} \right],$$

where \( h \) is the height of the satellite orbit, conventionally above the equatorial radius \( a_E \). If we substitute into this, e.g., the satellite height \( h = 800 \text{ km} \) (and use \( a_E = 6378137 \text{ m} \)) we obtain

$$\frac{d\Omega}{dt} = -1.33102 \cdot 10^{-6} \cos i \left[ \text{rad s}^{-1} \right] = -6^\circ.589 \text{ day}^{-1} \cos i. \quad (12.4)$$

For practical reasons (solar panels!) we often choose the satellite orbit such, that the orbital plane turns along with the annual apparent motion of the Sun, \( \frac{360^\circ}{365.25 \text{ days}} = \frac{0^\circ.9856}{\text{day}} \). See figure 12.6.

If the inclination \( i \) is chosen in the range \( 96^\circ - 102^\circ \), depending on the orbital height, then the Earth’s dynamic flattening \( J_2 \) will cause just the suitable rotational motion of the orbital plane (“no-shadow/Sun-synchronous/Sun-stationary orbit”), see figure 12.7.

An orbit with an inclination \( i > 90^\circ \) is called a retrograde orbit: the satellite is moving Westward in longitude, opposite to the direction of the Earth’s rotation, which is Eastward. The orbital inclination \( i \), or for a retrograde orbit,
its supplement $180^\circ - i$, is also the greatest Northern or Southern latitude a satellite can fly over. This means that, unless the inclination is precisely $90^\circ$, there will be areas around both poles that the satellite will never overfly.

A drawback of a Sun-stationary orbit is, that the altimetric observations are made at always the same local time of day. E.g., the diurnal and semidiurnal tides caused by the Sun will always be in the same phase, and thus they cannot be observed with this kind of satellite (“resonance”). Therefore the oceanographic satellite TOPEX/Poseidon, and its follow-up Jason satellites, were placed in non-sun-stationary orbits.

Figure 12.6. The mechanism of a Sun-stationary orbit.

Figure 12.7. Geometry of a “no-shadow” orbit. Season names are boreal.
Chapter 12. Satellite altimetry and satellite gravity missions

Figure 12.8. A satellite in a retrograde orbit, crossing the equator South to North three successive times. The angle between the orbit and the equator, the inclination \(i\), or for a retrograde orbit, its supplement \(180^\circ - i\), is also the highest Northern or Southern latitude that the satellite can fly over. The unreachable polar “caps” are indicated by white dashed lines.

12.4 Example

A satellite moves in an Sun-stationary orbit, i.e., she always, day after day, flies over the same latitude at the same local (mean) Solar time.

Questions:

1. What is the period of the satellite if she always after 14 revolutions flies again over the same spot?

2. Same question if she always flies over the same spot after 43 revolutions (3 days)?

3. And after 502 revolutions (35 days)?

4. What is the height of the satellite in a “three-day orbit”? Use Kepler’s third law, equation (12.3). \(GM = 3986005 \cdot 10^8 \text{m}^3\text{s}^{-2}\), and the height of the satellite is \(h = a_S - a\), where \(a = 6378137 \text{m}\).

5. What is the satellite height in a “35-day orbit”? And the height difference with the previous question?

6. What is, for the three-day orbit, the mean separation between North-going orbital tracks (i.e., at what level of detail is the altimeter able to image the sea surface)?
7. Same question for a 35-day orbit.

8. Questions for reflection:

   (a) for what purpose would you use a 35-day orbit, for what purpose a three-day orbit?

   (b) Would it be possible (easy) to fly both orbits with the same satellite (see question 5)?

Answers:

1. The satellite completes 14 orbits per day, i.e., per 1440 minutes: \( P = \frac{1440}{14} \text{ min} = 102.857 \text{ min}. \)

2. The satellite completes 43 orbits in three days, i.e., per \( 3 \times 1440 \) minutes: \( P = \frac{3 \times 1440}{43} \text{ min} = 100.465 \text{ min}. \)

3. The satellite completes 502 orbits in 35 days, i.e., per \( 35 \times 1440 \) minutes:
\[
P = \frac{35 \times 1440}{502} \text{ min} = 100.398 \text{ min}.
\]

4. Execute the following octave code:
\[
\begin{align*}
\%format long
GM &= 3986005e8; \\
 ae &= 6378137; \\
P &= 100.465*60; \ % \text{ seconds} \\
fac &= 4*\text{pi*pi}; \ % \text{ four } \pi \text{ square} \\
a &= (GM*P*P/fac)^{0.33333333}; \\
h &= a - ae; \\
printf('\n\nOrbital height: \ %8.3f \text{ km}.\n', h/1000);
\end{align*}
\]
The result is 780.604 km.

5. The same code, with \( P = 100.398*60 \), yields 777.421 km. The difference with the previous is 3.183 km.

6. There are 43 orbits with different ground tracks. That means a separation of \( \frac{360}{43} = 8.372 \) degrees, or at the equator, \( \frac{40000}{43} = 930 \) km; less at higher latitudes.

7. \( \frac{360}{502} = 0.717 \) degrees, or \( \frac{40000}{502} = 80 \) km.

8. (a) The 35-day orbit would be excellent for detailed mapping. The three-day orbit would be able to see, e.g., tides or weather-related phenomena, but at poor resolution.

   (b) The difference in height being only 3 km, and in period, 4 s, the change in orbit between the two repeat periods should be easily within reach of even small on-board thrusters. So, yes.
12.5 Retracking

The results of a satellite altimetry mission are published already during flight in the form of a so-called geophysical data record (GDR) file, containing everything related to the measurement and, e.g., atmospheric correction terms, tidal corrections, wave parameters, etc.

It is common practice today to process already collected altimetry measurements again, in order to extract more useful information. In this, the complete return pulse is analyzed again. The method is called retracking\(^6\).

The standard method of analysis is based on that point on the return pulse, which is at half height from the maximum value of the pulse. This is according to experience a good way to get the travel time associated with the point in the centre of the footprint, directly underneath the satellite. In the back part of the pulse are reflections from the further-away peripheral areas of the footprint.

There are however two situations where this method doesn’t work well during flight, and a more careful \textit{a posteriori} analysis of the pulse is worthwhile:

1. Archipelagos like Indonesia, Åland, … Here it may happen, e.g., that the centre point of the footprint is on land. Then, the first strong bounces will come under an angle from the nearest coast. A precise coastline file is then essential for processing.

2. Sea ice areas in the Arctic and Antarctic seas. Bounces may come from the surface of the sea ice, in which case one should consider freeboard in the processing, i.e., how much the ice sticks out of the water.

In both cases the traditional on-board processing produces erroneous measurements, as the travel time of the return pulse varies too rapidly as the

satellite flies on. With retracking, such measurements have been saved, and the area covered by altimetric measurements has been extended to the Arctic and Antarctic seas.

*Freeboard* is an important quantity in determining the thickness of the ice. As the density of ice is about $920 \text{ kg/m}^3$ and the density of sea water about $1027 \text{ kg/m}^3$, the ice thickness is about $8 \times$ freeboard. If additionally there is remote-sensing data on the area of ice cover, one can calculate the total volume and mass of sea ice.

The Arctic ice cover has diminished radically over the last decades. However, the most radical reduction has been that of ice volume, see figure 12.10: in addition to surface area, also thickness goes down, and especially of the multi-year, thicker ice, a large part has already vanished.

### 12.6 Oceanographic research using satellite altimetry

The interest of geodesy into satellite altimetry has traditionally been its use for determining the geoid. This works only if we assume that the sea surface

1. is constant

2. coincides with an equipotential surface, i.e., is the same as the geoid.

In practice however the ocean surface is variable and not an equipotential surface. For this reason, other perspectives have appeared.
1. The variability of the sea surface can be studied by satellite altimetry using three methods:

(a) Repeat tracks from the same satellite. The tracks can be stacked using a simple orbit error model, and the remaining per-track residuals tell something (but, not everything!) about the variability of the sea surface.

(b) Also the crossovers may provide information on sea-surface variability. When the sea surface varies, the results from the crossover adjustment will get poorer: the average \textit{a posteriori} (after calculation) crossover discrepancy will become larger. Using this method to actually \textit{study} sea-surface variability is more difficult: it is mostly just able to establish that it exists, and estimate its magnitude.

(c) Nowadays altimetry satellites always carry a GNSS positioning instrument, providing the absolute, geocentric location of the microwave radar device at the moment of measurement. With it, the variations of sea level can be monitored by direct measurement, assuming that both temporal and spatial measurement densities are sufficient.

2. The deviations of sea level from an equipotential surface – the geoid – can be studied only, if we have access to independent information on the \textit{true} geoid surface. If dense, high-quality gravity measurements are available for an area, this is the case, and we may estimate the \textit{sea-surface topography}.

Collecting sufficiently precise and dense gravimetric data is possible with a ship gravimeter or with \textit{airborne gravimetry}. Also, measurement with a special satellite (gravity gradiometry, GOCE satellite) has long been planned and was finally realized, see section 12.7.

12.7 Satellite gravity missions

During the early years of the 21st century three satellites were launched for investigating the fine structure of the Earth’s gravity field or geopotential, i.e., for determining a global, high resolution model of the geoid.

\textbf{CHAMP} (Challenging Minisatellite Payload for Geophysical Research and
Applications, 2000-099A) was a German satellite project under the auspices of the German Research Centre for Geosciences GFZ. She was launched into orbit from Plesetsk, Russia, in 2000. The orbit height of CHAMP was initially 454 km, diminishing over the mission time to $\sim 300$ km due to atmospheric drag. The orbital inclination was $87^\circ$. On September 19, 2010 the satellite returned into the atmosphere. A project description is found at the address http://op.gfz-potsdam.de/champ.

CHAMP contained a GPS receiver used for determining the precise orbit of the satellite, which again allows the calculation of her location in space $x(t)$ for any point in time $t$. From this, one may calculate the geometric acceleration $a(t)$ by differentiation:

$$a(t) = \frac{d^2}{dt^2} x(t).$$

The differentiation is done numerically in the way that was described in the part on airborne gravimetry, equation (10.8).

The satellite also contained an accelerometer, which eliminated the satellite accelerations caused by the atmosphere’s aerodynamic forces (i.e., the deviations from free-fall motion). Then, only the accelerations caused by the Earth’s gravitational field remain, from which a precise geopotential or geoid model may be calculated by the techniques described earlier.

A number of global geopotential models based on CHAMP data have been calculated and published.

**GRACE** (Gravity Recovery And Climate Experiment Mission, 2002-012A...
and B) measures *temporal changes* in the gravity field of the Earth at intervals of about a month extremely precisely, but at a rather crude geographic resolution. These temporal changes are caused by motions in the Earth’s “blue film”, i.e., atmosphere and hydrosphere. The quantity measured is also called the “sea-floor pressure”, a somewhat surprising expression, until one sees that it really represents the total mass of a column of air and water. The project is described at the University of Texas web site http://www.csr.utexas.edu/grace/. It is a collaborative American–German undertaking under the leadership of the Center for Space Research, University of Texas at Austin.

GRACE is a *satellite pair* (“Tom & Jerry”): the satellites fly in the same orbit in a tandem configuration at initially about 500 km height, at an inter-satellite separation of 220 km. The orbital inclination is 89°, i.e., the orbit is almost polar, providing complete global coverage. The changes in distance between the satellites are measured by a microwave link at a precision of 1 µm s⁻¹. Both satellites also carry sensitive accelerometers for measuring and eliminating the effect of atmospheric drag.

The measurement system is so sensitive, that even the movement of a water layer of one millimetre thickness can be noticed, as long as it extends to an area the size of a continent (some 500 km).

Published results show impressively, e.g., the wet and dry monsoons,
seasonal variations in opposite phase in the Northern and Southern hemispheres, in the great tropical river basins: Amazonas, Congo, the Mekong, Indonesia... see http://grace.jpl.nasa.gov/. Animation: GIF.

By 2016 the orbital height had come down to 360 km and the end of the mission is near. A GRACE follow-up mission is being planned.

**GOCE** (Geopotential and Steady-state Ocean Circulation Explorer, 2009-013A) was the most ambitious of all the satellites. Built by the European Space Agency ESA she was launched successfully from Plesetsk in March 2009. The orbital height was only 270 – 235 km during the mission and the satellite contained a rocket engine (an ionic engine) and a stock of propellant in order to maintain the orbit against atmospheric drag. The orbital inclination was 96.7°, i.e., the orbit was Sun-stationary. GOCE carried a very sensitive gravity gradiometer, a device that measured precisely components of the gradient of the Earth’s attraction, i.e., the dependence of components of the attraction vector on the various co-ordinates of place. The gradiometer consisted of six extremely sensitive, three-axes accelerometers mounted pairwise on a frame. The mission ended in 2013 and the satellite burned up in the atmosphere November 11 over the Falkland islands.

Theoretical analysis has shown that a gradiometer is the best way to measure the very local features of the Earth’s gravity field, better than orbital tracking by GNSS. The smallest details in the geoid map seen by GOCE are only 100 km in diameter, and their precision is as good as ±2 cm.

With a global geoid model this precise, we may calculate the deviations of the sea surface from the geoid or equipotential surface, at the same precision. We saw that the true location in space of the sea surface is obtained from satellite radar altimetry, also at a few centimetres precision. This separation between sea surface and equipotential sur-

---

7http://commons.wikimedia.org/wiki/File:Global_Gravity_Anomaly_Animation_over_LAND.gif

8Because of this inclination angle, there was a cap of radius 3.3° at each pole within which no measurements were obtained.

9http://blogs.esa.int/rocketscience/2013/11/11/goce-burning-last-orbital-view/
Figure 12.13. Determining the Earth’s gravity field with the gravity gradiometer on the GOCE satellite.

face can again be inverted to ocean currents, see section 11.5. This is the background for the name of the GOCE satellite.
13. Tides, atmosphere and Earth crustal movements

13.1 Theoretical tide

We may write the tidal potential $W$ as follows:

$$W = \frac{GMr^2}{d^3} P_2(\cos z) + \ldots = \frac{GMr^2}{2d^3} (3\cos^2 z - 1) + \ldots,$$

where $d$ is the distance to either the Moon or the Sun; $R$ the radius of the Earth, and $z$ the local zenith angle of the Sun or Moon. $P_2(\cos z)$ is the Legendre polynomial of degree two. $GM$ is the mass of the Sun or Moon multiplied by Newton’s constant. In the case of Sun and Moon the extra terms ($\ldots$) can be neglected, because these are such remote bodies: $d \gg R$.

According to spherical trigonometry

$$\cos z = \sin \phi \sin \delta + \cos \phi \cos \delta \cos t,$$

where $\phi$ is latitude, $\delta$ is the declination$^1$ of the Moon, and $t$ is the hour angle$^2$ of the Moon. Substitution gives

$$P_2(\cos z) = \frac{1}{2} (\cos^2 z - 1) =$$

$$= \frac{1}{2} (3\sin^2 \phi - 1) \frac{1}{2} (3\sin^2 \delta - 1) + \frac{3}{4} \cos^2 \phi \cos^2 \delta \cos 2t +$$

$$+ \frac{3}{4} \sin 2\phi \sin 2\delta \cos t,$$

$^1$The declination is the geocentric latitude of the Moon.

$^2$The hour angle is the difference in longitude between the Moon and the local meridian. It is zero when the Moon is due South in upper culmination.
Low tide

High tide

Low tide

High tide

Figure 13.1. Theoretical tide. $z$ is the local zenith angle of the Moon (or Sun).

from which

$$W = \frac{GM R^2}{4d^3} \left[ (3 \sin^2 \phi - 1)(3 \sin^2 \delta - 1) + 3 \sin 2\phi \sin 2\delta \cos t + 3 \cos^2 \phi \cos^2 \delta \cos 2t \right].$$

This is the so-called Laplace tidal equation.

It has three parts:

1. A slowly varying part,

$$W_1 = \frac{GM R^2}{4d^3} \left[ (3 \sin^2 \phi - 1)(3 \sin^2 \delta - 1) \right],$$

that still depends on $\delta$ and therefore is periodic with a 14-day (half-month) period. Again by using spherical trigonometry:

$$\sin \delta = \sin \epsilon \sin \ell \Rightarrow \sin^2 \delta = \sin^2 \epsilon \sin^2 \ell = \sin^2 \epsilon \left( \frac{1}{2} - \frac{1}{2} \cos 2\ell \right),$$

where $\ell$ is the longitude of the Moon in its orbit, reckoned from the ascending node (equator crossing), and $\epsilon$ is the inclination of the Moon’s orbit with respect to the equator, on average $23^\circ$ but rather variable, between $18^\circ.3$ and $28^\circ.6$. Thus we obtain

$$W_1 = \frac{GM R^2}{4d^3} \left[ (3 \sin^2 \phi - 1) \left( 3 \sin^2 \epsilon \left( \frac{1}{2} - \frac{1}{2} \cos 2\ell \right) - 1 \right) \right],$$
where we have used result \((13.1)\). We split \(W_l = W_{l_1} + W_{l_2}\), a constant \(^3\) and a periodic or semi-monthly (“fortnightly”) part:

\[
W_{l_1} = \frac{GMR^2}{4d^3} \left[ (3 \sin^2 \phi - 1) \left( \frac{3}{2} \sin^2 \epsilon - 1 \right) \right]; \quad (13.2)
\]

\[
W_{l_2} = -\frac{GMR^2}{4d^3} \left[ (3 \sin^2 \phi - 1) \left( \frac{3}{2} \sin^2 \epsilon \cos 2\ell \right) \right].
\]

2. Additionally we have a couple of terms in which the hour angle \(t\) appears (periods roughly a day and roughly half a day):

\[
W_2 = \frac{GMR^2}{4d^3} \left[ 3 \sin 2\phi \sin 2\delta \cos t \right],
\]

\[
W_3 = \frac{GMR^2}{4d^3} \left[ 3 \cos^2 \phi \cos^2 \delta \cos 2t \right].
\]

In both, we have in addition to \(t\), still \(\delta\) as a “slow” variable. These equations could be written out as sums of various functions of the Lunar longitude \(\ell\).

Use again basic trigonometry, equation \((13.1)\):

\[
\cos^2 \delta = 1 - \sin^2 \delta = 1 - \sin^2 \epsilon \sin^2 \ell = 1 - \sin^2 \epsilon \left[ \frac{1}{2} - \frac{1}{2} \cos 2\ell \right];
\]

\[
\cos 2\ell \cos 2t = \frac{1}{2} \left[ \cos(2\ell + 2t) + \cos(2\ell - 2t) \right];
\]

\[
\sin 2\delta = 2 \sin \delta \cos \delta = 2 \sin \delta \sqrt{\cos^2 \delta} =
\]

\[
= 2 \sin \epsilon \sin \ell \sqrt{1 - \sin^2 \epsilon \left[ \frac{1}{2} - \frac{1}{2} \cos 2\ell \right]},
\]

leading to a trigonometric expansion in \(\ell\), and so on. See, e.g., Melchior’s\(^4\) famous book Melchior (1978).

From the above equations, often the coefficient

\[
D \equiv \frac{3GMR^2}{4d^3},
\]

“Doodson’s\(^5\) constant” is removed. The value for the Moon is \(D = 26.75 \text{ cm} \times g\) and for the Sun 12.3 cm \(\times g\). See figure 13.2.

---

\(^3\)Not precisely, because \(\epsilon\) is (slowly) time dependent.

\(^4\)Paul Melchior (1925–2004) was an eminent Belgian geophysicist and Earth tides researcher.

\(^5\)Arthur Thomas Doodson (1890–1968) was a British oceanographer, a pioneer of tidal theory, also involved in designing machines for computing the tides. He was stone deaf.
Table 13.1. The various periods in the theoretical tide. The widely used symbols were standardized by George Darwin.

<table>
<thead>
<tr>
<th>Changing function</th>
<th>Period</th>
<th>Darwin symbol</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W_{1a} )</td>
<td>-</td>
<td>-</td>
<td>( M_0 ) Sun, ( S_0 ) Moon</td>
</tr>
<tr>
<td>( W_{1b} ) cos 2( \ell )</td>
<td>14(^d) 182(^d)</td>
<td>( Mf^a ) Sun, ( Ssa^b ) Moon</td>
<td>Declination tide</td>
</tr>
<tr>
<td>( W_2 ) cos ( t )</td>
<td>24(^h)50(^m) 24(^h)</td>
<td>( K_1, O_1 ) Moon, ( S_1, P_1 ) Sun</td>
<td>Diurnal</td>
</tr>
<tr>
<td>( W_3 ) cos 2( t )</td>
<td>12(^h)25(^m) 12(^h)</td>
<td>( M_2 ) Sun, ( S_2 ) Moon</td>
<td>Semi-diurnal</td>
</tr>
</tbody>
</table>

\(^a\)Lunar fortnightly
\(^b\)Solar semi-annual

The periods are tabulated in table 13.1 with their Darwin\(^6\) symbols.

In practice, the diurnal and semi-diurnal tides can be divided further into many “spectral lines” close to each other, also because the Lunar orbit (like the Earth orbit) is an ellipse, not a circle.

13.2 Deformation caused by the tidal force

The tidal force, or theoretical tide, of which we spoke above, is not the same as the deformation it causes in the solid Earth. This deformation will depend upon the elastic properties inside the Earth. These elastic properties are often described by so-called (elastic) Love\(^7\) numbers (Melchior, 1978).

Let us first write the external (tidal, or generally, disturbing) potential in the following way:

\[
W = \sum_{n=2}^{\infty} \left( \frac{r}{R} \right)^n W_n,
\]

– where now the index \( n \) denotes the degree number of spherical harmonics!
– and we call the displacement of an element of matter of the solid Earth in the radial direction, \( u_r \), in the North direction, \( u_\phi \), and in the East direction,

\(^6\)Sir George Howard Darwin (1845 – 1912) was an English astronomer and mathematician, son of Charles Darwin of Origin of Species fame.

\(^7\)Augustus Edward Hough Love (1863 – 1940) was a British mathematician and student of Earth elasticity.
Deformation caused by the tidal force

\[ W_{1a}, \text{permanent} \]

\[ W_{1b}, \text{fortnightly, } \epsilon = 23^\circ \]

\[ W_2, \text{diurnal, } \delta = 23^\circ \]

\[ W_3, \text{semi-diurnal, } \delta = 23^\circ \]

**Figure 13.2.** The main components of the theoretical tide. These values must still be multiplied by Doodson’s constant \( D \).

\( u_\lambda \). The following equations apply:

\[
\begin{align*}
  u_r &= \sum_{n=2}^{\infty} H_n (r) \frac{W_n}{8}, \\
  u_\phi &= \frac{1}{8} \sum_{n=2}^{\infty} L_n (r) \frac{\partial W_n}{\partial \phi}, \\
  u_\lambda &= \frac{1}{8} \sum_{n=2}^{\infty} L_n (r) \frac{\partial W_n}{\cos \phi \partial \lambda}.
\end{align*}
\]

Here \( r \) is the distance from the geocentre. It is assumed here that the Love numbers \( H_n, L_n \) depend only on \( r \), i.e., the elastic properties of the Earth are spherically symmetric.

The deformation of the Earth causes also a change (the “indirect effect”, in addition to the original potential \( W \)) in the gravity potential. We write

\[
\delta W = \sum_{n=2}^{\infty} K_n (r) W_n,
\]

where we use already a third type of Love numbers.
On the Earth surface $r = R$ we make the following specialization:

\[ H_n (R) = h_n, \]
\[ L_n (R) = \ell_n, \]
\[ K_n (R) = k_n. \]

In practice, because of the large distances to Sun and Moon, the only important part of the tidal potential $W$ is the part for the degree number $n = 2$, i.e., $W_2$.

The Love numbers will depend still on the frequency, i.e., the tidal period $P$:

\[ h_n = h_n (P), \]
\[ \ell_n = \ell_n (P), \]
\[ k_n = k_n (P). \]

The tides offer an excellent means of determining all these Love numbers $h_2 (P), \ell_2 (P), k_2 (P)$ empirically, because, being periodic variations, they cause in the Earth deformations at the same periods, but different amplitudes and phases. In this way we may determine at least those Love numbers that correspond to periods occurring in the theoretical tide.

The $h$ and $\ell$ numbers are nowadays obtained, e.g., by GNSS positioning. The GNSS processing software contains a built-in reduction for this phenomenon. From gravity measurements one obtains information on a certain linear combination of $h$ and $k$ (vertical displacement changes gravity though its gradient, and deformation of the Earth, the shifting of masses, also changes gravity). A useful research instrument is also the long water-tube clinometer, like the tube of the Finnish Geodetic Institute that has long been in use in the Tytyri limestone mine in Lohja (Kääriäinen and Ruotsalainen, 1989). The same applies for sensitive clinometers in general, like the Verbaandert-Melchior pendulum etc. A clinometer measures the change in orientation between the Earth’s crust and the local plumbline.

Measuring the absolute direction of the plumbline, e.g., with a zenith tube, can again give information on a certain linear combination of $\ell$ and $k$, but only after various reductions (Earth polar motion).
13.3 The permanent part of the tide

As shown above, the theoretical tide equation contains a constant part that doesn’t even vary in a long-period way. Of course the Earth responds also to this part of the tidal force; however, because the deformation isn’t periodic, it is not possible to measure it. And the mechanical theory of the solid Earth, and our knowledge of the state of matter inside of the Earth, just aren’t good enough for prediction of the response.

For this reason the understanding is generally accepted that the effect of the permanent part of the tide on the Earth’s state of deformation should not be included in any tidal reduction (Ekman, 1992). However, often, e.g., in the processing of GNSS observations or in defining spherical harmonic expansions of the Earth’s gravity field, the tidal reduction does include this term which it is theoretically and practically impossible to know. See Poutanen et al. (1996).

More generally we can say, that a geodetic quantity, e.g., the height of the geoid, can be reduced for the permanent part of the tide in three different ways:

- No reduction whatever is made for the permanent part; the quantity thus obtained is called the “mean geoid”. The surface obtained is in the hydrodynamic sense an equilibrium surface, and is therefore the best surface to use in oceanography.

- The effect of the gravitational field emanating from celestial objects is removed in its entirety from the quantity, but the Earth’s deformation it causes is left completely uncorrected; the quantity thus obtained is called the “zero geoid”.

- Both the gravitational effect of a celestial body, and the indirect effect of the deformation it causes, can be calculated according to a certain deformation model (Love number), and corrected for. The result obtained is called the “tide-free geoid”. Its problem is precisely the empirical indeterminacy of the elasticity model used.

It is good to be critical and precisely analyze in which way the data reduction has been done!
13.4 Loading of the Earth’s crust by sea and atmosphere

In addition to the deformation caused by the tidal force, the Earth’s crust also deforms due to the loading by sea and atmosphere. Especially close to the coast, the tidal motion of the sea causes a multi-period deformation that moves the Earth’s crust up and down by as much as centimetres.

This phenomenon can be computationally modelled if the elastic properties of the solid Earth, the tidal motion of the sea, and the precise shape of the coastline are known. One known programme for this purpose is the package Eterna written by the German H.-G. Wenzel\(^8\), which has been used also in Finland.

On the other hand, when such a tool exists, then tidal loading offers also an excellent opportunity for studying precisely the very local elastic properties.

For measuring the deformation, generally a registering gravimeter is used. The Earth’s crust moves up and down, which changes gravity in proportion to the free-air gradient value \(-0.3\ \text{mGal/m}\) (Torge, 1992).

The use of GNSS for measuring the ocean tidal loading has not yet become common.

Like the ocean, also the atmosphere causes, through changes in air pressure, varying deformations of the Earth’s crust. The phenomenon is very small, at most a couple of cm. Gravity measurement is not a very good way to study this phenomenon, because many more local, often poorly known, factors affect local gravity. Measurement by GNSS is promising but also challenging.

---

\(^8\)Hans-Georg Wenzel (1945 – 1999) was a German physical geodesist and geophysicist.
14. Earth gravity field research

14.1 Internationally

In the framework of the IAG, the International Association of Geodesy, research into the Earth’s gravity field is currently the responsibility of the International Gravity Field Service. The IGFS was created in 2003 at the IUGG General Assembly in Sapporo, Japan, and it operates under the IAG’s new Commission 2 “Gravity Field”. The United States National Geospatial-Intelligence Agency (NGA) serves as its technical centre.

An important IAG service of great reputation is the International Gravity Bureau, BGI, Bureau Gravimétrique International, located in Toulouse, France (http://bgi.omp.obs-mip.fr/). The bureau works as a sort of clearing-house, to which countries can submit their gravimetric materials. If some researcher needs gravimetric material from another country, e.g., for geoid computation, he can request it from the BGI, who will provide it with the permission of the country of origin, provided the country of the researcher has in its turn submitted its own gravimetric materials for BGI use.

The French state has invested significant funds into this vital international activity.

Another important IAG service in this field is the ISG, International Service for the Geoid. It has in fact already operated since 1992 under the name International Geoid Service (IGeS), the executive arm of the International Geoid Commission (IGeC). The ISG office is located in Milano (http://www.isgeoid.polimi.it/) also with substantial support by the Italian state. The task of this service is to support geoid calculation in different countries, for which purpose existing geoid solutions are collected in a common data base,
and international research schools are organized to develop knowledgability
and skills in the art of geoid computation, especially in developing countries.
Both services, BGI and ISG, are under the auspices of the International Grav-
ity Field Service IGFS, as two of the many official services of the IAG. Other
IGFS services are the International Center for Earth Tides (ICET), the In-
ternational Center for Global Earth Models (ICGEM), and the International
Digital Elevation Model Service (IDEMS).

14.2 Europe

In Europe operates the EGU, the European Geosciences Union, in the frame
of which much publication and meeting activities relating to gravity field
and geoid are being co-ordinated. The EGU organizes annually symposia,
where always also sessions are included on subjects related to gravity field
and geoid. Also American scientists participate. Conversely the American
Geophysical Union’s (AGU) fall and spring meetings\(^1\) are also favoured by
European researchers.

To be mentioned is the Geodetic Institute (“Institut für Erdmessung”) of
Leibniz University in Hannover, Germany, which since 1990 has acted as
the European computing centre of the International Geoid Commission,
and produced high quality geoid models for use in Europe (Denker, 1998);
(http://www.ife.uni-hannover.de/gravity.html?&L=1).

14.3 The Nordic countries

In the Nordic countries, important work is being co-ordinated by the NKG,
Nordiska Kommissionen för Geodesi and its Working Group for Geoid and
Height Systems. To its activities belongs the calculation of new geoid mod-
els, studying the preconditions for still more precise geoid models, new
levelling technologies, and the study of post-glacial land uplift.

The group has for a long time computed, at its computing centre in Copen-
hagen, high-quality geoid models, the last such being NKG2004 (Forsberg

\(^1\)Fall (autumn) meetings in San Francisco, spring meetings somewhere in the world.
The AGU, though American, is a very cosmopolitan player.
The newest model, working name NKG2015, which is still a work in progress, will be the result of calculations by the computing centres of several countries including Sweden and Estonia.

14.4 Finland

In Finland the study of the Earth’s gravity field has mainly been in the hands of the Finnish Geodetic Institute, founded in 1918, one year after Finnish independence. The institute has been responsible for the national fundamental levelling and gravimetric networks and their international connections. In 2001 the Finnish Geodetic Institute’s gravity and geodesy departments were joined into a new department of geodesy and geodynamics, to which also gravity research belongs. Among matters studied are, e.g., solid-Earth tides, the free oscillations of the solid Earth, post-glacial land uplift, and vertical reference or height systems.

Geoid models have been computed all the time, starting with Hirvonen’s global model (Hirvonen, 1934) and ending, for now, with the Finnish model FIN2005N00 (Bilker-Koivula, 2010). These geoid models are actually based on the Nordic NKG2004 gravimetric geoid, and are only fitted to a Finnish set of GNSS levelling control points through a transformation surface.

In 2015, the Finnish Geodetic Institute was merged into the National Land Survey as its geospatial data centre and research facility. The English-language acronym continues as FGI, the Finnish Geospatial Research Institute (http://www.fgi.fi/fgi).

Also Helsinki University of Technology (today Aalto University) has been active in research on the Earth’s gravity field. W.A. Heiskanen, who was a professor at HUT in 1928–1949, acted 1936–1949 as the director of the International Isostatic Institute. After moving to Ohio State University he worked with many other, including Finnish and Finnish-born, geodesists on calculating the first major global geoid model, the “Columbus geoid” (Kakkuri, 2008).
14.5 Textbooks

There are many good textbooks on the study of the Earth’s gravity field. In addition to the already mentioned classic Heiskanen and Moritz (1967), which is in large part obsolete, we may mention Wolfgang Torge’s book Torge (1989). Difficult but good is Moritz (1980). Similarly difficult is Molodensky et al. (1962). Worth reading also from the perspective of physical geodesy is Vaníček and Krakiwsky (1986). A new book in this field is Hofmann-Wellenhof and Moritz (2006).


---

265


Pratt, J. H. (1855). The attraction of the Himalaya mountains upon the plumbline in India. *Philosophical Transactions of the Royal Society of London*, CXLV:53–100. 96


A. Field theory and vector calculus – core knowledge

A.1 Vector calculus

In physics, we describe many quantities as vector quantities. E.g., force, velocity, electromagnetic field, … A vector behaves in the same way as the location difference between two neighbouring points. Velocity $\mathbf{v}$, force $\mathbf{F}$, location difference $\Delta \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the location vectors of points 1 and 2. In a co-ordinate transformation, the vector considered as an object does not change, but the numeric values of its components are co-ordinate system dependent, see subsection A.2.2.

About notation: in printed text, vectors are most often written in bold. In handwritten text one may use a small upper arrow: $\vec{v}$.

A.1.1 Scalar product

Between two vectors, a scalar product or dot product can be defined, which is itself a scalar value. A scalar is in physics a single numeric value, e.g., pressure or temperature. In the case of a scalar product of two vector fields, the value is tied to a location, but, even if a co-ordinate transformation changes the co-ordinate values of the location, the scalar itself remains unchanged.

An example of a scalar product: work $\Delta E$ is

$$\Delta E = \langle \mathbf{F} \cdot \Delta \mathbf{r} \rangle,$$

the scalar product of force $\mathbf{F}$ and path $\Delta \mathbf{r}$. Often, also in the sequel, we leave the angle brackets $\langle \cdot \rangle$ off.
Later we shall see that if the points 1 and 2, \( \Delta r = r_2 - r_1 \), are very close to each other, we may write

\[
dE = \langle F \cdot dr \rangle,
\]

where \( dr \) and \( dE \) are infinitesimal elements of path and energy. If now there is between the points A and B a curved path, we may get from this an integral equation:

\[
\Delta E_{AB} = \int_A^B F \cdot dr.
\]

This is the work integral.

### A.1.2 The scalar product, formally

Let

\[
s \equiv \langle a \cdot b \rangle
\]

be the scalar product of the vectors \( a \) and \( b \). Then

\[
\langle \mu a \cdot b \rangle = \langle a \cdot \mu b \rangle = \mu \langle a \cdot b \rangle,
\]

\[
\langle a \cdot b \rangle = \langle b \cdot a \rangle,
\]

and often we call

\[
\|a\| \equiv \sqrt{\langle a \cdot a \rangle}
\]

the norm or length of vector \( a \).

The following also applies:

\[
\langle a \cdot b \rangle = \|a\| \|b\| \cos \alpha,
\]

where \( \alpha \) is the angle between vectors \( a \) and \( b \).

### A.1.3 Exterior or vectorial product

The exterior product of two vectors is itself a vector (at least in three-dimensional space \( \mathbb{R}^3 \)). E.g., the angular momentum \( q \):

\[
q = \langle r \times p \rangle,
\]

where \( p = mv \) is linear momentum, \( m \) the mass of the body, and \( v = \frac{dr}{dt} \) is the time derivative of the location, or velocity. We write

\[
q = m \left\langle r \times \frac{dr}{dt} \right\rangle.
\]
A.1.4 The vectorial product, formally

Let
\[ \mathbf{c} \equiv (\mathbf{a} \times \mathbf{b}) \]
be the vectorial product of the two vectors \( \mathbf{a} \) and \( \mathbf{b} \). Then (\( \mu \in \mathbb{R} \)):
\[ (\mu \mathbf{a} \times \mathbf{b}) = (\mathbf{a} \times \mu \mathbf{b}) = \mu (\mathbf{a} \times \mathbf{b}) , \]
\[ (\mathbf{a} \times \mathbf{b}) = -(\mathbf{b} \times \mathbf{a}) , \]
and thus \( (\mathbf{a} \times \mathbf{a}) = 0 \).

The resulting vector \( \mathbf{c} \) is always orthogonal to the vectors \( \mathbf{a} \) and \( \mathbf{b} \); the length of vector \( \mathbf{c} \) corresponds to the surface area of the parallelogram spanned by the vectors \( \mathbf{a} \) and \( \mathbf{b} \). In a formula:
\[ \| \mathbf{c} \| = \| \mathbf{a} \| \| \mathbf{b} \| \sin \alpha , \]
where again \( \alpha \) is the angle between vectors \( \mathbf{a} \) and \( \mathbf{b} \). If the angle is zero, then also the vectorial product is zero (because then, \( \mathbf{a} = \mu \mathbf{b} \) for some suitable value of \( \mu \)).

A.1.5 Kepler’s second law

If \( \mathbf{r} \) is the distance of the body (planet) from the centre of motion (the Sun), and \( \frac{d\mathbf{r}}{dt} \) is the distance travelled in a unit of time, then the product
\[ (\mathbf{r} \times \frac{d\mathbf{r}}{dt}) \]
is precisely twice the surface area of the triangle or “area” swept over.

Let us take the time derivative of this product, i.e., the expression (A.1):
\[ \frac{d}{dt} \left( \mathbf{r} \times \frac{d\mathbf{r}}{dt} \right) = \frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} + \mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} . \]
Appendix A. Field theory and vector calculus – core knowledge

Here the first term vanishes, because $\langle a \times a \rangle = 0$. In the second term we can exploit our knowledge, that the attractive force $F$ emanating from the Sun that causes planetary orbital motion – and also the acceleration it causes, $a \equiv \frac{d^2 r}{dt^2}$ – is central:

$$F = ma = -\frac{GMm}{\|r\|^3} r.$$ 

Substitute into the above:

$$\frac{d}{dt} \left( r \times \frac{dr}{dt} \right) = 0 - \frac{GM}{\|r\|^3} \langle r \times r \rangle = 0.$$ 

So: the quantity $\langle r \times \frac{dr}{dt} \rangle$ – angular momentum per unit of mass $q/m$ – is conserved. Like, e.g., the total amount of energy, electric charge and many other quantities, the amount of angular momentum in a closed system is constant.

$G$ is the universal gravitational constant, $M$ is the mass of the Sun, $m$ the mass of the planet.

A.2 Scalar and vector fields

A.2.1 Definitions

In the space $\mathbb{R}^3$ we may define functions, or fields.

A scalar field is a scalar-valued function, which is defined throughout the space (or a part of it), e.g., temperature $T$:

$$T(r).$$

I.e., for every value of the location vector $r$ there is a temperature value $T$. 
A vector field is a vector-valued function that again is defined throughout space, e.g., the electrostatic field $E$:

$$E(r).$$

### A.2.2 A base in space

In the space $\mathbb{R}^3$ we may choose a base made up of three vectors which span the space in question. Generally we choose three base vectors $i, j, k$, that are orthogonal to each other, and the norms, or lengths, of which are $1$:

$$i \perp j, i \perp k, j \perp k; \|i\| = \|j\| = \|k\| = 1.$$ 

Now we may write vectors out into their components:

$$a = a_1i + a_2j + a_3k$$

and also scalar and vectorial products can now be calculated with the aid of their components:

$$s = \langle a \cdot b \rangle = \langle (a_1i + a_2j + a_3k) \cdot (b_1i + b_2j + b_3k) \rangle = a_1b_1 + a_2b_2 + a_3b_3 = \sum_{i=1}^{3} a_ib_i,$$

using the above identities for the base vectors.

For the vectorial product, the calculation is more involved; we get as the final outcome the determinant

$$c = \langle a \times b \rangle =
\begin{vmatrix}
i & j & k \\
a_1 & a_2 & a_3 \\
b_1 & b_2 & b_3 \\
\end{vmatrix} = (a_2b_3 - a_3b_2)i + (a_3b_1 - a_1b_3)j + (a_1b_2 - a_2b_1)k.$$ 

I.e.,

$$c_1 = a_2b_3 - a_3b_2,$$

$$c_2 = a_3b_1 - a_1b_3,$$

$$c_3 = a_1b_2 - a_2b_1.$$ 

Also these expressions are determinants:

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} a_2 & a_3 & a_1 \\ a_3 & a_1 & a_2 \\ b_2 & b_3 & b_1 \end{bmatrix}^T.$$
Appendix A. Field theory and vector calculus – core knowledge

A.2.3 The nabla operator

The location vector \( \mathbf{r} \) can be written on the \( \{i, j, k\} \) base as follows:

\[
\mathbf{r} = xi + yj + zk,
\]

which defines \( (x, y, z) \) co-ordinates in space.

Let us now define a vector operator called nabla (\( \nabla \)) as follows:

\[
\nabla \equiv i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z}.
\]

The operator, or function, is on its own without meaning. It acquires meaning only when it “operates” on something, in which case the three partial derivatives on the right-hand side can be calculated.

A.2.4 The gradient

Let \( F (\mathbf{r}) = F (x, y, z) \) be a scalar field in space. The nabla operator with give its gradient \( \mathbf{g} \), a vector field in the same space:

\[
\mathbf{g} = \text{grad} F = \nabla F = i \frac{\partial F}{\partial x} + j \frac{\partial F}{\partial y} + k \frac{\partial F}{\partial z}.
\]

So, the field \( \mathbf{g} (\mathbf{r}) = \mathbf{g} (x, y, z) \) is a vector field in the same space, the gradient field of \( F \).

**Interpretation:** the gradient describes the slope of the scalar field. The direction of the vector is the direction in which the value of the scalar field changes fastest, and its length describes the rate of change with location. Imagining a hilly landscape: the height if the ground above sea level is the scalar field, and its gradient is pointing everywhere uphill, away from the valleys toward the hilltops. The \( \mathbf{g} \) arrows are the longer, the steeper is the slope of the ground surface.

The gradient operator (like also the divergence and the curl, see later) is linear:

\[
\text{grad} \ (F + G) = \text{grad} F + \text{grad} G.
\]

A.2.5 The divergence

Given a vector field \( \mathbf{a} (x, y, z) \). We form the **scalar product** \( s \) of this and the nabla operator:

\[
s = \text{div} \mathbf{a} = \langle \nabla \cdot \mathbf{a} \rangle = \frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z}.
\]
**A.2. Scalar and vector fields**

---

**Figure A.3.** The gradient. The level curves of the scalar field in blue.

**Interpretation:** The divergence describes the “sources” of a vector field, both the positive and the negative ones. Imagine the velocity of flow of water as a vector field. At the locations of the “sources” the divergence is positive, at the locations of the “sewer holes” or sinks, negative, everywhere else zero (because liquid cannot appear out of nothing or disappear into nothing).

---

**Figure A.4.** The divergence. Positive divergences (“sources”) and negative ones (“sinks”).
A.2.6 The curl

Given a vector field \( \mathbf{a} (x, y, z) \). We form the vectorial product \( \mathbf{c} \) of this and the nabla operator, producing again a vector field:

\[
\mathbf{c} = \text{rot} \mathbf{a} = \nabla \times \mathbf{a} = \begin{vmatrix}
\mathbf{i} & \mathbf{j} & \mathbf{k} \\
\frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\
a_1 & a_2 & a_3
\end{vmatrix} =
\]

\[
= \left( \frac{\partial a_3}{\partial y} - \frac{\partial a_2}{\partial z} \right) \mathbf{i} + \left( \frac{\partial a_1}{\partial z} - \frac{\partial a_3}{\partial x} \right) \mathbf{j} + \left( \frac{\partial a_2}{\partial x} - \frac{\partial a_1}{\partial y} \right) \mathbf{k},
\]

using the evaluation rules for determinants.

**Interpretation:** the curl describes the eddiness or turbulence present in a vector field.

Imagine a weather map, where low- and high-pressure zones are drawn. Our vector field is the wind field. The wind circulates (on the Northern hemisphere) clockwise around the high-pressure zones, and counterclockwise around the low-pressure zones. We may say that the curl of the wind field is positive at the high pressures and negative at the low pressures.

(This is a poor metaphor as it is two-dimensional. In \( \mathbb{R}^2 \), the curl is a scalar, not a vector. Just like we need only one angle to describe a rotational motion, when in \( \mathbb{R}^3 \) we need the three Euler angles.)

A.2.7 Conservative fields

What happens if a vector field \( \mathbf{a} \) is the gradient of a scalar field \( F \), and we try to calculate its curl? As follows:

\[
\text{rot} \mathbf{a} = \text{rot grad } F = \begin{vmatrix}
\mathbf{i} & \mathbf{j} & \mathbf{k} \\
\frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\
\frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} & \frac{\partial F}{\partial z}
\end{vmatrix} =
\]

\[
= \left( \frac{\partial}{\partial y} \frac{\partial F}{\partial z} - \frac{\partial}{\partial z} \frac{\partial F}{\partial y} \right) \mathbf{i} + \left( \frac{\partial}{\partial z} \frac{\partial F}{\partial x} - \frac{\partial}{\partial x} \frac{\partial F}{\partial z} \right) \mathbf{j} + \left( \frac{\partial}{\partial x} \frac{\partial F}{\partial y} - \frac{\partial}{\partial y} \frac{\partial F}{\partial x} \right) \mathbf{k} = 0!
\]
In other words, if the vector field \( \mathbf{a}(x,y,z) \) is the gradient of the scalar field \( F(x,y,z) \), its curl will vanish.

**Definition:** this kind of vector field \( \mathbf{a} \) is called *conservative*, and the corresponding scalar field \( F \), \( \mathbf{a} = \text{grad} \ F \), is called the *potential* of field \( \mathbf{a} \).

Note that if
\[
\mathbf{a}(x,y,z) = \text{grad} \ F(x,y,z),
\]
then also
\[
\mathbf{a}(x,y,z) = \text{grad} \ (F(x,y,z) + F_0),
\]
where \( F_0 \) is a constant, because
\[
\text{grad} \ F_0 = \mathbf{i} \frac{\partial F_0}{\partial x} + \mathbf{j} \frac{\partial F_0}{\partial y} + \mathbf{k} \frac{\partial F_0}{\partial z} = 0.
\]
So the potential is *not uniquely* defined.

### A.2.8 The Laplace operator

Assume a conservative field \( \mathbf{a} \), i.e., \( \text{rot} \ \mathbf{a} = 0 \). Then we may write
\[
\mathbf{a} = \text{grad} \ F = \nabla F,
\]
where \( F \) is the potential.
Let us now express the divergence of field $\mathbf{a}$ into the potential:

$$\text{div} \, \mathbf{a} = \nabla \mathbf{a} = \nabla \nabla F = \frac{\partial}{\partial x} \frac{\partial}{\partial x} F + \frac{\partial}{\partial y} \frac{\partial}{\partial y} F + \frac{\partial}{\partial z} \frac{\partial}{\partial z} F = \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) F = \Delta F,$$

where we have introduced a new differential operator: the Delta operator invented by the French Pierre Simon de Laplace,

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

For the potential of a “source free” field – e.g., for the gravitational potential in vacuum, the electrostatic potential in an area of space free of electric charges – this Delta operator vanishes.

### A.3 Integrals

#### A.3.1 The curve integral

We saw earlier, that work $\Delta E$ can be written as the scalar product of force $\mathbf{F}$ and path $\Delta \mathbf{r}$:

$$\Delta E = (\mathbf{F} \cdot \Delta \mathbf{r}).$$

The differential form of this is

$$dE = \mathbf{F} \cdot d\mathbf{r},$$

from which one obtains the integral form

$$\Delta E_{AB} = \int_{A}^{B} \mathbf{F} \cdot d\mathbf{r}.$$  

Here is computed the amount of work needed to move a body from point $A$ to point $B$ by integrating $\mathbf{F} \cdot d\mathbf{r}$ along the path $AB$.

If we parametrize the path according to arc length $s$, and the tangent vector to the path is called

$$\mathbf{t} = \frac{dx}{ds} \mathbf{i} + \frac{dy}{ds} \mathbf{j} + \frac{dz}{ds} \mathbf{k},$$

we may also write

$$\Delta E_{AB} = \int_{A}^{B} (\mathbf{F} \cdot \mathbf{t}) \, ds,$$

the parametrized version of the integral.
A.3 Integrals

A.3.2 The surface integral

Assume given again some vector field \( \mathbf{a} \) and a surface in space \( S \). Often, one needs to integrate over the surface \( S \) the normal component of a vector field, the projection of \( \mathbf{a} \) onto the normal vector of the surface.

Let the normal vector of the surface be \( \mathbf{n} \). Then we must integrate

\[
\iint_S (\mathbf{a} \cdot \mathbf{n}) \, dS,
\]

symbolically written

\[
\iint_S \mathbf{a} \cdot d\mathbf{S},
\]

where the notation \( d\mathbf{S} \) is called an oriented surface element. It is a vector pointing in the same direction as the normal vector \( \mathbf{n} \).

Like a curve, also a surface can be parametrized. E.g., the Earth's surface (assumed a sphere) can be parametrized by latitude \( \varphi \) and longitude \( \lambda \): \( \mathbf{r} = \mathbf{r}(\varphi, \lambda) \). In this case we write as the surface element

\[
dS = R^2 \cos \varphi \, d\varphi d\lambda,
\]

where \( R^2 \cos \varphi \) is the Jacobian, or determinant of Jacobi, of the parameter pair \((\varphi, \lambda)\). In this parametrization, the integral is calculated as follows:

\[
\iint_S \mathbf{a} \cdot d\mathbf{S} = \iint_S (\mathbf{a} \cdot \mathbf{n}) \, R^2 \cos \varphi \, d\varphi d\lambda.
\]

Other surfaces and parametrizations have other Jacobians. The Jacobian always describes the true area of a "parameter surface element" \( d\varphi d\lambda \) "in nature". E.g., on the Earth surface, a degree times degree patch is the largest.
near the equator. In polar co-ordinates \((\rho, \theta)\) in the plane \((x = \rho \cos \theta, y = \rho \sin \theta)\) the Jacobian is \(\rho\). In the ordinary \((x, y)\) parametrization in the plane, it is 1 and thus can be left out altogether.

### A.3.3 The Stokes edge integral theorem

Let \(S\) be a surface in space (not necessarily flat) and \(\partial S\) its edge curve. Assume that the surface and its edge are well-behaved enough for all necessary integrations and differentiations to be possible. Then (Stokes):

\[
\iint_S \text{rot} \mathbf{a} \cdot dS = \oint_{\partial S} \mathbf{a} \cdot d\mathbf{r}.
\]

**In words:** The surface integral of the curl of a vector field over a surface is the same as the closed path integral of the field around the edge of the surface.

**Special case:** If \(\text{rot} \mathbf{a} = 0\) everywhere (a conservative force field) then

\[
\oint_{\partial S} \mathbf{a} \cdot d\mathbf{r} = 0,
\]

i.e., also

\[
\int_{A, \text{path 1}}^{B} \mathbf{a} \cdot d\mathbf{r} = \int_{A, \text{path 2}}^{B} \mathbf{a} \cdot d\mathbf{r}.
\]

In other words, *the work integral from point A to point B does not depend on the path chosen. And the work done by a body transported around a closed path is zero.*

This explains perhaps better the essence of a conservative force field. A conservative field can be represented as the *gradient of a potential*: \(\mathbf{a} = \text{grad} F\), where \(F\) is the potential of the field. The Earth’s gravity field \(\mathbf{g} (x, y, z)\) is the gradient of the Earth’s gravity potential \(W (x, y, z)\). At mean sea level – more precisely, at the geoid – the gravity potential is constant; the gravity vector \(\mathbf{g}\) stands everywhere orthogonally on the geoid.

### A.3.4 The Gauss integral theorem

Let \(V\) be a certain volume of space, and \(\partial V\) its closed boundary, a union of surfaces. Assume again that both are mathematically well behaved. Then the following theorem applies (Gauss):

\[
\iiint_V \text{div} \mathbf{a} \, dV = \iint_{\partial V} \mathbf{a} \cdot dS = \int_{\partial V} (\mathbf{a} \cdot \mathbf{n}) \, dS.
\]
A.4. The continuity of matter

An often used equation in, e.g., hydro- or aerodynamics is the continuity equation. This describes that matter cannot just disappear or increase in amount. In the general case, this equation looks like this:

$$\text{div} (\rho \mathbf{v}) + \frac{d}{dt} \rho = 0.$$  

Here, the expression $\rho \mathbf{v}$ describes mass currents; $\rho$ is the matter density, $\mathbf{v}$ is the current flow velocity. The term $\text{div} (\rho \mathbf{v})$ describes how much more matter, in a unit of time, exits the volume element than enters it, per unit of volume. The second term again, $\frac{d}{dt} \rho$, describes the change in the amount of matter inside the volume element over time. Those two terms must balance for the “matter accounting” to close.
If the moving fluid is non-compressible, then \( \rho \) is constant and
\[
\frac{d}{dt} \rho = 0 \quad \text{and} \quad \text{div} (\rho \mathbf{v}) = \rho \text{ div } \mathbf{v},
\]
and we obtain
\[
\rho \text{ div } \mathbf{v} = 0 \implies \text{div } \mathbf{v} = 0.
\]

Remember, however, that not necessarily \( \text{rot } \mathbf{v} = 0 \) – i.e., the flow isn’t necessarily eddy free –, i.e., a potential \( F \) for which \( \mathbf{v} = \text{grad } F \) does not necessarily exist.
B. Function spaces

B.1 An abstract vector space

In an abstract vector space we may create a base, with the help of which each vector can be written as a linear combination of the base vectors: e.g., if the base, in a three-dimensional space, is \( \{ e_1, e_2, e_3 \} \), we may write an arbitrary vector \( r \) in the form:

\[
r = r_1 e_1 + r_2 e_2 + r_3 e_3 = \sum_{i=1}^{3} r_i e_i.
\]

Precisely because three base vectors is always enough, we call the ordinary (Euclidean) space three-dimensional.

In a vector space one can define a scalar product, which is a linear mapping from two vectors to one number (“bilinear form”):

\[
\langle r \cdot s \rangle.
\]

Linearity means that

\[
\langle \alpha r_1 + \beta r_2 \cdot s \rangle = \alpha \langle r_1 \cdot s \rangle + \beta \langle r_2 \cdot s \rangle,
\]

and commutativity, that

\[
\langle r \cdot s \rangle = \langle s \cdot r \rangle
\]

If the base vectors are orthogonal to each other, in other words, \( \langle e_i \cdot e_j \rangle = 0 \) if \( i \neq j \), we may calculate the coefficients \( r_i \) simply:

\[
r = \sum_{i=1}^{3} \frac{\langle r \cdot e_i \rangle}{\langle e_i \cdot e_i \rangle} e_i \quad \text{(B.1)}
\]

If, additionally, also \( \langle e_i \cdot e_i \rangle = \|e_i\|^2 = 1 \forall i \in \{1, 2, 3\} \), i.e., the base vectors are orthonormal, equation (B.1) becomes simpler still:

\[
r = \sum_{n=1}^{3} \langle r \cdot e_i \rangle e_i. \quad \text{(B.2)}
\]
Here, the coefficients are \( r_i = \langle r \cdot e_i \rangle \). The quantity
\[
\|e_i\| = \sqrt{\langle e_i \cdot e_i \rangle}
\]
is called the norm of the vector \( e_i \).

### B.2 Fourier function space

Also functions can be considered elements in a vector space. If we define the scalar product of two functions \( f, g \) as the following integral:
\[
\langle \vec{f} \cdot \vec{g} \rangle \equiv \frac{1}{\pi} \int_{0}^{2\pi} f(x) g(x) \, dx,
\]
it is easily verified that the above requirements for a scalar product are met.

One base in this vector space (a function space) is formed by the so-called Fourier functions,
\[
\begin{align*}
\vec{e}_0 &= \frac{1}{2} \sqrt{2} \quad (k = 0) \\
\vec{e}_k &= \cos kx, \quad k = 1, 2, 3, \ldots \\
\vec{e}_{-k} &= \sin kx, \quad k = 1, 2, 3, \ldots
\end{align*}
\]
This base is orthonormal (proof: exercise). It is also a complete base, which we shall not prove. As the number of base vectors is countably infinite, we say that this function space is infinitely dimensional.

Now every function \( f(x) \) meeting certain conditions can be expanded in the way of equation ((B.2)), i.e.,
\[
f(x) = a_0 \frac{1}{2} \sqrt{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx),
\]
– the familiar Fourier expansion – where the coefficients are
\[
\begin{align*}
a_0 &= \langle \vec{f} \cdot \vec{e}_0 \rangle = \frac{1}{\pi} \int_{0}^{2\pi} f(x) \frac{1}{2} \sqrt{2} dx = \sqrt{2} \cdot \overline{f(x)}, \\
a_k &= \langle \vec{f} \cdot \vec{e}_k \rangle = \frac{1}{\pi} \int_{0}^{2\pi} f(x) \cos kx \, dx, \\
b_k &= \langle \vec{f} \cdot \vec{e}_{-k} \rangle = \frac{1}{\pi} \int_{0}^{2\pi} f(x) \sin kx \, dx, \quad k = 1, 2, \ldots
\end{align*}
\]
This is the familiar way in which the coefficients of a Fourier series are calculated.
B.3 Sturm-Liouville differential equations

B.3.1 The eigenvalue problem

In an abstract vector space we may formulate an eigenvalue problem: if there exists a linear operator (mapping) \( L \), we may write

\[
Lx = \lambda x,
\]

where the problem consists of determining the values \( \lambda_i \) for which a solution \( x_i \) exists.

In a concrete \( n \)-dimensional vector space we may write the vector

\[
x = \sum_{i=1}^{n} x_i e_i,
\]

and, thanks to linearity,

\[
Lx = \left\{ \sum_{i=1}^{n} x_i e_i \right\} = \sum_{i=1}^{n} x_i \cdot L \{ e_i \};
\]

on the other hand we may write \( n \) different vectors \( L \{ e_i \} \) on the base \( \{ e_j \} \) in the following way:

\[
L \{ e_i \} = \sum_{j=1}^{n} a_{ij} e_j.
\]

This defines the coefficients \( a_{ij} \), which may be collected into an \( n \times n \) matrix \( A \).

Now

\[
Lx = \sum_{j=1}^{n} \left[ \sum_{i=1}^{n} a_{ij} x_i \right] e_j, \tag{B.3}
\]

when also

\[
\lambda x = \lambda \sum_{i=1}^{n} x_i e_i = \sum_{j=1}^{n} \left[ \lambda x_j \right] e_j. \tag{B.4}
\]

By combining the vector equations (B.3), (B.4), which must be identical, we obtain

\[
\sum_{i=1}^{n} a_{ij} x_i - \lambda x_j = 0
\]

or, as a matric equation,

\[
Ax - \lambda x = 0, \tag{B.5}
\]

where \( A \) is a matrix consisting of the coefficients \( a_{ij} \), and \( x \) a column vector consisting of the coefficients \( x_i \): \( x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T \).
Of course also equation ((B.5)) represents an eigenvalue problem, but now in the linear vector space consisting of all coefficient vectors $x$. Every $x$ is the numerical representation of a vector $x$ on the chosen base $\{e_i\}$. The matrix $A$ again is the numerical representation of operator $L$ on the same base$^1$.

### B.3.2 A self-adjoint operator

Let $L$ be a linear operator in a vector space where there exists a scalar product, i.e., a bilinear form $\langle x \cdot y \rangle$ which is symmetric or commutative.

Then $L$ is self-adjoint, if for each pair of vectors $x, y$ it holds that

$$\langle x \cdot Ly \rangle = \langle Lx \cdot y \rangle.$$

If the matrix $A$ is self-adjoint, that means that

$$\langle x \cdot Ay \rangle = \langle Ax \cdot y \rangle$$

i.e.,

$$\sum_{i=1}^{n} x_i \left( \sum_{j=1}^{n} a_{ij} y_j \right) = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij} x_j \right) y_i,$$

which is trivially true, if, for all $i, j$ values $1, \ldots, n$,

$$a_{ij} = a_{ji}, \text{ i.e., } A = A^T.$$

In other words:

\begin{quote}
A symmetric matrix is a self-adjoint operator.
\end{quote}

From linear algebra it is undoubtedly familiar, that the eigenvectors $x_p, x_q$ belonging to different eigenvalues $\lambda_p \neq \lambda_q$ of a symmetric $n \times n$ matrix are mutually orthogonal: $x_p \perp x_q$. If all eigenvalues $\lambda_p, p = 1, \ldots, n$ are different, then the eigenvectors $x_p, p = 1, \ldots, n$ will constitute a complete base in the vector space $\mathbb{R}^n$.

The proof is not hard. We start from the equation for the eigenvalue problem for eigenvectors and -values $x_p, \lambda_p$:

$$Lx_p = \lambda_p x_p,$$

and multiply from the left by vector $x_q$:

$$\langle x_q \cdot Lx_p \rangle = \lambda_p \langle x_q \cdot x_p \rangle.$$

$^1$An advantage of the numerical representations is of course that one can really calculate with them.
Similarly for eigenvectors and -values $x_q, \lambda_q$:

$$\langle x_p \cdot L x_q \rangle = \lambda_q \langle x_p \cdot x_q \rangle .$$

If now $L$ is self-adjoint, then

$$\langle x_q \cdot L x_p \rangle = \langle L x_q \cdot x_p \rangle = \langle x_p \cdot L x_q \rangle$$

(remember that the scalar product is symmetric) and thus, that

$$\lambda_p \langle x_q \cdot x_p \rangle = \lambda_q \langle x_p \cdot x_q \rangle$$

i.e.,

$$\left( \lambda_p - \lambda_q \right) \langle x_p \cdot x_q \rangle = 0.$$  

If $\lambda_p \neq \lambda_q$, we thus must have $\langle x_p \cdot x_q \rangle = 0$, or $x_p \perp x_q$. What was to be proven.

**Example:** the variance matrix in the plane. The variance matrix of the co-ordinates of point $P$ in the plane is

$$\text{Var} \left( x_P \right) = \Sigma = \begin{bmatrix} \sigma^2_x & \sigma_{xy} \\ \sigma_{xy} & \sigma^2_y \end{bmatrix},$$

a symmetric matrix. Here, $\sigma^2_x$ and $\sigma^2_y$ are the variances, or squares of the mean errors, of the $x$ and $y$ co-ordinates, whereas $\sigma_{xy}$ is the covariance between the co-ordinates.

The eigenvalues of this matrix $\Sigma$ are the solutions of the characteristic equation

$$\det \begin{bmatrix} \sigma^2_x - \lambda & \sigma_{xy} \\ \sigma_{xy} & \sigma^2_y - \lambda \end{bmatrix} = 0$$

i.e.,

$$\left( \sigma^2_x - \lambda \right) \left( \sigma^2_y - \lambda \right) - \sigma^2_{xy} = 0.$$  

This yields

$$\lambda_{1,2} = \frac{1}{2} \left( \sigma^2_x + \sigma^2_y \right) \pm \sqrt{\left[ \frac{1}{2} \left( \sigma^2_y - \sigma^2_x \right) \right]^2 + \sigma^2_{xy}}.$$  

The variance matrix has a *variance or error ellipse*. The semi-lengths of its principal axes are $\sqrt{\lambda_1}, \sqrt{\lambda_2}$ and the directions of the principal axes are the eigenvectors of $\Sigma$, $x_1, x_2$, mutually orthogonal. If the coordinate axes are turned in such a way, that the axes of the ellipse are in the directions of $x_{1,2}$, then the matrix $\Sigma$ will assume the form

$$\Sigma' = \begin{bmatrix} \sigma^2_x & 0 \\ 0 & \sigma^2_y \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$
The sum of the eigenvalues, $\lambda_1 + \lambda_2 = \sigma_x^2 + \sigma_y^2$, is an invariant called the point variance.

B.3.3 Self-adjoint differential equations

In function space there are also self-adjoint or “symmetric” differential equations. In fact, the most famous equations of physics are of this type.

Take a good look at, e.g., the oscillation equation
\[
\frac{d^2}{dt^2} x - \lambda x = 0. \tag{B.6}
\]

The solution has the general form
\[
x(t) = \alpha \sin(t\sqrt{\lambda}) + \beta \cos(t\sqrt{\lambda}),
\]
on the interval $[0, 2\pi]$ the solution (we require $x(0) = x(2\pi) = 0$) has the general form
\[
x(t) = \alpha \sin(t\sqrt{\lambda}),
\]
which is possible only for certain values of $\lambda$.

Note that equation (B.6) is an eigenvalue problem, form-wise:
\[
Lx - \lambda x = 0,
\]
where the operator is
\[
L = \frac{d^2}{dt^2}.
\]

We first show that this operator is on the interval $[0, 2\pi]$ self-adjoint. If the scalar product is defined as follows:
\[
\langle x \cdot y \rangle \equiv \int_0^{2\pi} x(t) y(t) \, dt,
\]
it holds that (integration by parts):
\[
\langle x \cdot Ly \rangle = \int_0^{2\pi} x(t) \frac{d^2}{dt^2} y(t) \, dt = \left[ x(t) \frac{d}{dt} y(t) \right]_0^{2\pi} - \int_0^{2\pi} \frac{d}{dt} x(t) \frac{d}{dt} y(t) \, dt,
\]
\[
\langle Lx \cdot y \rangle = \int_0^{2\pi} \left( \frac{d}{dt} x(t) \right) y(t) \, dt = \left[ \left( \frac{d}{dt} x(t) \right) y(t) \right]_0^{2\pi} - \int_0^{2\pi} \frac{d}{dt} x(t) \frac{d}{dt} y(t) \, dt.
\]

As on the right-hand side the first terms vanish and the second terms are identical, we have
\[
\langle x \cdot Ly \rangle = \langle Lx \cdot y \rangle,
\]
which was to be proven.
Self-adjoint operators have eigenvalues and eigenvectors, in this case functions, that are mutually orthogonal. They are just the solution functions
\[ \sin \left( t \sqrt{\lambda_p} \right) = \sin (2\pi pt), \quad p \in \mathbb{N} \]
where
\[ \lambda_p = (2\pi p)^2 \]
represents the energy level of a certain oscillation of quantum number \( p \).

In physics there is a broad class of partial differential equations that are self-adjoint in some function space. The class is known as “Sturm-Liouville type problems”. To it belongs, e.g., the oscillation equation, Legendre’s equation, Bessel’s equation, and many more. Every one of them generates, in a natural way, its own set of orthogonal functions that serve as the base functions for the equation’s general solution.

### B.4 Legendre polynomials

Also the ordinary Legendre polynomials \( P_n(t) \) constitute a base in function space, with the scalar product definition
\[ \left\langle \overrightarrow{f} \cdot \overrightarrow{g} \right\rangle \equiv \int_{-1}^{+1} f(t) g(t) \, dt. \]
They don’t however constitute an orthonormal base, but only an orthogonal one:
\[ \| \overrightarrow{P_n} \|^2 = \left\langle \overrightarrow{P_n} \cdot \overrightarrow{P_n} \right\rangle = \int_{-1}^{+1} P_n^2(t) \, dt = \frac{1}{2n + 1}. \]

Unlike ordinary space, which is three-dimensional, a function space is an \( \infty \)-dimensional, abstract vector space, that nevertheless helps us make more concrete certain abstract, but very useful fundamentals of function theory!

### B.5 Spherical harmonics

On the surface of the sphere also all functions can be considered elements of a function space. Every function meeting certain “well-behavedness proper-

---

2) Jacques Charles François Sturm (1803 – 1855) was an eminent French mathematician, one of the 72 names engraved on the Eiffel Tower.

3) Joseph Liouville (1809 – 1882) was an eminent French mathematician.
ties” – like integrability – is an element. The functions

\[ R_{nm}(\phi, \lambda) = P_{nm}(\sin \phi) \cos m\lambda, \]
\[ S_{nm}(\phi, \lambda) = P_{nm}(\sin \phi) \sin m\lambda, \]

together, for the values \( n = 0 \ldots \infty, m = 0 \ldots n \), form a \textit{complete base} for this vector space in such a way, that every function can be written as a (if necessary infinite) linear combination of these base functions.

An alternative, more compact way of writing this is

\[
Y_{nm}(\phi, \lambda) = \begin{cases} 
P_{nm}(\sin \phi) \cos m\lambda & m \geq 0, \\
P_{n||m||}(\sin \phi) \sin ||m|| \lambda & m < 0, 
\end{cases}
\]

for values \( n = 0 \ldots \infty, m = -n \ldots n \).

(The situation is analogous to three-dimensional space, where a complete base consists of three vectors not in the same plane.)

In the function space, a \textit{scalar product} is defined:

\[
\langle \vec{v} \cdot \vec{w} \rangle = \frac{1}{4\pi} \int_\sigma V(\phi, \lambda) W(\phi, \lambda) \, d\sigma,
\]

where \( \sigma \) is the surface of the unit sphere ("directional sphere", or even "celestial sphere"). According to this definition we can show, that two different functions, namely \( Y_{nm}, Y_{sr} \), are orthogonal with respect to each other: e.g.,

\[
\langle \vec{y}_{nm} \cdot \vec{y}_{sr} \rangle = \frac{1}{4\pi} \int_\sigma Y_{nm}(\phi, \lambda) Y_{sr}(\phi, \lambda) \, d\sigma = 0
\]

if \( n \neq s \) or \( m \neq r \), etc.

The base \( \{ \vec{y}_{nm} \} \) is \textit{orthogonal} but not \textit{orthonormal}: the “length” of every vector differs from 1.

\[
||\vec{y}_{nm}||^2 = \langle \vec{y}_{nm} \cdot \vec{y}_{nm} \rangle = \frac{1}{4\pi} \int_\sigma Y_{nm}(\phi, \lambda)^2 \, d\sigma = \begin{cases} 
\frac{1}{2n+1} & m = 0, \\
\frac{1}{2(2n+1)} \frac{(n+||m||)!}{(n-||m||)!} & m \neq 0,
\end{cases}
\]

see also Heiskanen and Moritz (1967). Proving this with the help of equation (2.11) is a long process.

If we now divide the functions \( Y_{nm} \) (or, equivalently, \( R_{nm}, S_{nm} \)) by the square roots of the above factors, we obtain the \textit{fully normalized} surface spherical harmonics \( \overline{Y}_{nm} \).

With those it is again easy to calculate the coefficients \( \overline{f}_{nm} \) of a given general function \( f(\phi, \lambda) \) (the overbar means that these are fully normalized coefficients):

\[
\overline{f}_{nm} = \langle f \cdot \overline{y} \rangle \Rightarrow \overline{f}_{nm} = \frac{1}{4\pi} \int_\sigma f(\phi, \lambda) \overline{Y}_{nm}(\phi, \lambda) \, d\sigma. \quad (B.7)
\]
This is a straightforward projection on the unit vectors of the base (geometric analogue).

In the above integral, \( f(\phi, \lambda) \) is the function \( f \) on the Earth’s surface, i.e., if the mean radius of the Earth is \( R \), \( f(\phi, \lambda) = f(\phi, \lambda, R) \).

The equation corresponding to expansion (2.8) is

\[
V(\phi, \lambda, r) = \sum_{n=0}^{\infty} \frac{1}{n^{n+1}} \sum_{m=0}^{n} \bar{P}_{nm}(\sin \phi) \left( \bar{\alpha}_{nm} \cos m\lambda + \bar{\beta}_{nm} \sin m\lambda \right).
\]

We may write also

\[
\bar{Y}_{nm}(\phi, \lambda) = \begin{cases} 
\bar{P}_{nm}(\sin \phi) \cos m\lambda & m \geq 0, \\
\bar{P}_{n||m||}(\sin \phi) \sin ||m|| \lambda & m < 0,
\end{cases}
\]

which corresponds to the definition of the fully normalized Legendre functions:

\[
\bar{P}_{n0}(\sin \phi) = \sqrt{2n+1} P_{n0}(\sin \phi),
\]

\[
\bar{P}_{nm}(\sin \phi) = \sqrt{2(2n+1) \frac{(n-m)!}{(n+m)!}} P_{nm}(\sin \phi), \quad m > 0.
\]

Then the above equation for the potential becomes

\[
V(\phi, \lambda, r) = \sum_{n=0}^{\infty} \frac{1}{n^{n+1}} \sum_{m=-m}^{n} (\bar{Y}_{nm} \bar{Y}_{nm}(\phi, \lambda)),
\]

in which

\[
\bar{Y}_{nm} = \begin{cases} 
\bar{\alpha}_{nm} & m \geq 0, \\
\bar{\beta}_{n||m||} & m < 0,
\end{cases}
\]
C. Why does FFT work?

There are alternatives in choosing the precise FFT method. The fastest FFT requires a grid the size of which is a power of 2, i.e., a grid of size $2^m \times 2^n$. Alternative, “mixed-radius” methods, may also be considered and perform well if the grid size is something like $360 \times 480$, e.g., $N = 360 = 2 \times 2 \times 2 \times 3 \times 3 \times 5$. If the grid size is a prime number, FFT does not give any advantage over ordinary discrete FFT.

If the function $f(x)$ is given on the interval $[0, 1)$ on an equi-spaced grid as values $f(x_k), k = 0, \ldots, N - 1$, then the discrete Fourier transform in one dimension is

$$\mathcal{F}\{f(x)\} = F(\omega),$$

where

$$F(\omega_j) = \frac{1}{N} \sum_{k=0}^{N-1} f(x_k) e^{2\pi i \omega_j k / N}, \quad j = 0, \ldots, N - 1. \quad (C.1)$$

Here, also the frequency argument $\omega_j, j = 0, \ldots, N - 1$ is defined on the interval $[0, 1]$. (Remember that $i$ is the imaginary unit $i^2 = -1$.)

Correspondingly, the inverse discrete Fourier transform

$$\mathcal{F}^{-1}\{F(\omega)\}$$

is

$$f(x_k) = \sum_{j=0}^{N-1} F(\omega_j) e^{-2\pi i \omega_j k / N}, \quad k = 0, \ldots, N - 1. \quad (C.2)$$

FFT is just a very efficient method for computing both these formulas (C.1), (C.2). A brute-force calculation of these formulas requires of order $N^2$ “standard operations”, each of them a single multiplication plus a single addition.

---

1 Or some other interval.
If $N$ is even, we may write

$$F(\omega_j) = \frac{1}{N} \left[ \sum_{k=0}^{N/2-1} f(x_k) e^{-2\pi i k \frac{N}{2N}} + \sum_{k=N/2}^{N-1} f(x_k) e^{-2\pi i k \frac{N}{2N}} \right] =$$

$$= \frac{1}{N} \left[ \sum_{k=0}^{N/2-1} f(x_k) e^{-2\pi i k \frac{N}{2N}} + e^{-2\pi i j \frac{N}{2N}} \sum_{k'=0}^{N/2-1} f(x_{k'+N/2}) e^{-2\pi i k' \frac{N}{2N}} \right] =$$

$$= \frac{1}{N} \left[ \sum_{k=0}^{N/2-1} f(x_k) e^{-2\pi i k \frac{N}{2N}} + e^{-\pi ij} \sum_{k=0}^{N/2-1} f(x_{k+N/2}) e^{-2\pi i k \frac{N}{2N}} \right] =$$

$$= \frac{1}{N} \sum_{k=0}^{N/2-1} \left[ f(x_k) \pm f\left(x_{k+N/2}\right) \right] e^{-2\pi i k \frac{N}{2N}}, \quad \begin{cases} + & \text{if } j \text{ even} \\ - & \text{if } j \text{ odd} \end{cases}$$

(C.3)

the computation of which sum requires only $\frac{N^2}{2}$ multiplications and $\frac{N^2}{2} + \frac{N}{2} \approx \frac{N^2}{2}$ additions / subtractions. Here we used Euler’s identity $e^{-\pi i} = -1$, i.e., $e^{-\pi ij} = (e^{-\pi i})^j = (-1)^j$, either $+1$ or $-1$.

Note that the expression in square brackets $[\cdot]$ is the same separately for all either even or odd values of $j$, and will be evaluated only $N/2$ times. Altogether some $\frac{N^2}{2}$ standard operations, half the original number.

Equation (C.3) is itself recognised as a Fourier series, but the number of support points is instead of $N$ only $\frac{N}{2}$. If also $\frac{N}{2}$ is even, we may repeat the above trick, resulting in an expression requiring only of order $\frac{N^2}{4}$ operations. Lather, rinse, repeat, and the number of operations becomes $\frac{N^2}{8}$, $\frac{N^2}{16}$, $\frac{N^2}{32}$, etc... A more precise analysis shows that, if $N$ is a power of 2, then the whole discrete Fourier transform may be computed in order $N \times 2 \log N$ operations!

In the literature, smart algorithms are found implementing the above method, e.g., **fftw** (“Fastest Fourier Transform in the West”, [http://www.fftw.org](http://www.fftw.org)).
D. Helmert condensation

D.1 The interior potential of the topography

In order to derive the equation for Helmert condensation, we first derive the equation for the interior potential of the topography, i.e., the masses between sea level and the terrain surface:

\[ T_t(r, \phi, \lambda) = G \int \int \int_{\text{top}} \frac{\rho(r', \phi', \lambda')}{\ell(r, r', \psi)} dV, \]

where \( \psi \) is the angular distance between the evaluation point \( (r, \phi, \lambda) \) and the data point \( (r', \phi', \lambda') \). The spatial distance \( \ell \) between those points again is written using the interior expansion (equation (7.4)):

\[ \frac{1}{\ell} = \frac{1}{r} \sum_{n=0}^{\infty} \left( \frac{r}{r'} \right)^{n+1} P_n(\cos \psi), \]

This expansion converges uniformly\(^{1}\) with respect to \( \psi \) if \( r < r' \). Substitute:

\[ T_t^{\text{int}}(r, \phi, \lambda) = G\rho \int \int \int_{\text{top}} \frac{1}{r} \sum_{n=0}^{\infty} \left( \frac{r}{r'} \right)^{n+1} P_n(\cos \psi) dV = \]

\[ = G\rho \int_\sigma \int_R^{R+H(\phi', \lambda')} \frac{1}{r} \sum_{n=0}^{\infty} \left( \frac{r}{r'} \right)^{n+1} (r')^2 \ln r' P_n(\cos \psi) d\sigma = \]

\[ = G\rho \int_\sigma \left[ \sum_{n=0}^{\infty} r^n \left( -\frac{1}{n-2} (r')^{-(n-2)} + r^2 \ln r' \right) \right]_R^{R+H} P_n(\cos \psi) d\sigma = \]

\[ = G\rho \int_\sigma \left[ \sum_{n=0}^{\infty} \frac{r^n}{n-2} \left( R^{-(n-2)} - (R + H)^{-(n-2)} + r^2 \ln \frac{R+H}{r} \right) \right] P_n(\cos \psi) d\sigma. \]

\(^{1}\)Uniform convergence means that, given \( r, r' \), for every \( \epsilon > 0 \) there is an \( N_{\min} \), for which

\[ \left| \frac{1}{\ell} - \frac{1}{r} \sum_{n=0}^{N} \left( \frac{r}{r'} \right)^{n+1} P_n(\cos \psi) \right| < \epsilon \]

for all \( N > N_{\min} \), and for all values of \( \psi \). This is a stronger property than mere convergence.
Here, we shall expand the following expression into a Taylor series:

\[(R + H)^{-(n-2)} = R^{-n} \left[ 1 - (n-2) \frac{H}{R} + \frac{(n-2)(n-1)}{2} \frac{H^2}{R^2} - \frac{(n-2)(n-1)n}{23} \frac{H^3}{R^3} + \ldots \right].\]

Also the special case \(n = 2\),

\[r^2 \ln \frac{R + H}{R} = r^2 \left[ \frac{H}{R} - \frac{1}{2} \frac{H^2}{R^2} + \frac{1}{3} \frac{H^3}{R^3} - \frac{1}{4} \frac{H^4}{R^4} + \ldots \right] = \frac{r^n}{R^{n-2}} \left[ \frac{H}{R} - \frac{n-1}{2} \frac{H^2}{R^2} + \frac{(n-1)n}{23} \frac{H^3}{R^3} - \frac{(n-1)n(n+1)}{234} \frac{H^4}{R^4} + \ldots \right],\]
is cleanly included into the following expression obtained by substitution:

\[T_{int}^\text{ext}(r, \phi, \lambda) = G \rho \int \sum_{n=0}^{\infty} \frac{r^n}{R^{n-2}} \left[ \frac{H}{R} - \frac{n-1}{2} \frac{H^2}{R^2} + \frac{(n-1)n}{6} \frac{H^3}{R^3} - \ldots \right] P_n(\cos \psi) d\sigma.\]

(D.1)

**D.2 The exterior potential of the topography**

In the same way we may derive the exterior potential of the topography. Now we use for the expansion of the inverse distance (equation (7.4)):

\[\frac{1}{r} = \sum_{n=0}^{\infty} \frac{1}{r} \left( \frac{r'}{r} \right)^n P_n(\cos \psi) = \sum_{n=0}^{\infty} \frac{1}{r} \left( \frac{r'}{r} \right)^n P_n(\cos \psi),\]

which converges uniformly if \(r > r'\). Substitution yields

\[T_{ext}^\text{ext} = G \rho \int \int_{\sigma \in \text{top}} \sum_{n=0}^{\infty} \frac{1}{r} \left( \frac{r'}{r} \right)^n P_n(\cos \psi) d\gamma =\]

\[= G \rho \int_{\sigma} \int_{R}^{R+H} \left( \frac{r'}{r} \right)^n d\gamma P_n(\cos \psi) d\sigma =\]

\[= G \rho \int_{\sigma} \left[ \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \frac{1}{n+3} \left( \frac{r'}{r} \right)^{n+3} \right]^{R+H}_R P_n(\cos \psi) d\sigma =\]

\[= G \rho \int_{\sigma} \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \frac{1}{n+3} \left( (R + H)^{n+3} - R^{n+3} \right) P_n(\cos \psi) d\sigma.\]

Here again we use the Taylor expansion:

\[(R + H)^{n+3} = R^{n+3} \left[ 1 + (n+3) \frac{H}{R} + \frac{(n+3)(n+2)}{2} \frac{H^2}{R^2} + \frac{(n+3)(n+2)(n+1)}{23} \frac{H^3}{R^3} + \ldots \right].\]

Substitution yields

\[T_{ext}^\text{ext} = G \rho R^2 \sum_{\sigma} \frac{r^{n+1}}{r} \left[ \frac{H}{R} + \frac{n+2}{2} \frac{H^2}{R^2} + \frac{(n+2)(n+1)}{6} \frac{H^3}{R^3} + \ldots \right] P_n(\cos \psi) d\sigma.\]

(D.2)

This is thus the exterior potential of the topography, or, inside the topographic masses, the harmonic downward continuation of the exterior potential, assuming that this is mathematically possible (in the case of mountainous topography, generally not) and doesn’t diverge.
D.3 The exterior potential of the condensation layer

This is derived by specializing equation (D.2) to the case $H \to 0$, but nevertheless $\rho \to \infty$, so that $\kappa = \rho H$ is finite. in this limit, all terms containing $H^2, H^3$ etc. are going to zero. The result is then

$$T_{ext}^c = G\rho R^2 \int \sum_{\nu=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} \frac{H}{R^2} P_n (\cos \psi) d\sigma =$$

$$= GR \int \sum_{\nu=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} \kappa P_n (\cos \psi) d\sigma. \quad (D.3)$$

Earlier on we had a more precise formula (5.4) for $\kappa$ on the surface of a spherical Earth:

$$\kappa = \rho H \left( 1 + \frac{H}{R} \right);$$

by substituting this in the previous, we obtain

$$T_{ext}^c = G\rho R^2 \int \sum_{\nu=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} \left[ \frac{H}{R} + \frac{H^2}{R^2} + \frac{(n+2)(n+1) H^3}{6 R^3} + \ldots \right] P_n (\cos \psi) d\sigma. \quad (D.3)$$

D.4 The total potential of Helmert condensation

This is obtained by subtracting equations (D.2) and (D.3) from each other. The result (which thus applies in the exterior space) is

$$T_{Helmert}^{ext} = T_{t}^{ext} - T_{c}^{ext} =$$

$$= G\rho R^2 \int \sum_{\nu=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} \left[ \frac{n+2}{2} - 1 \right] \frac{H^2}{R^2} + \frac{1}{6} \frac{(n+2)(n+1) H^3}{R^3} + \ldots \right] P_n (\cos \psi) d\sigma =$$

$$= G\rho \int \sum_{\nu=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} \left[ \frac{n+2}{2} H^2 + \frac{(n+2)(n+1) H^3}{6} \frac{H^3}{R} + \ldots \right] P_n (\cos \psi) d\sigma. \quad (D.4)$$

Often we define the degree constituents of powers of height $H$ (compare the degree constituent equation (2.19)), as follows:

$$H_n^\nu = \frac{2n+1}{4\pi} \int H^\nu P_n (\cos \psi) d\sigma,$$

after which we may expand

$$H^\nu = \sum_{n=0}^{\infty} H_n^\nu.$$
Then

\[
T_{\text{Helmert}}^\text{ext} = 4\pi G \rho \sum_{n=0}^{\infty} \left( \frac{R}{r} \right)^{n+1} \left[ \frac{n}{2(2n+1)} H_n^2 + \frac{(n+2)(n+1)}{6(2n+1)} \frac{H_2^3}{R} + \ldots \right].
\]  

(D.6)

If the topography is constant, then all terms for which \( n \neq 0 \) vanish; in the above expression, also the first term vanishes, and the second and further terms are very small. So, in practice\(^2\)

\[
T_{\text{Helmert}}^\text{ext} = \frac{4}{3} \pi H^3 \cdot \frac{G \rho}{r} + \ldots \approx 0
\]

as was to be expected. The exterior field remains in this special case almost unchanged as a result of Helmert condensation.

**D.4.1 The gravity effect of Helmert condensation**

Let us calculate gravity anomalies from the Helmert potential, but only using the first term of equation (D.6):

\[
\Delta g_{\text{Helmert}} = \frac{\partial T}{\partial r} + \frac{2}{r} T = \frac{2}{r} T = 2\pi G \rho \sum_{n=0}^{\infty} \frac{n}{2n+1} \left[ \frac{-(n+1)}{r} + \frac{2}{r} \right] \left( \frac{R}{r} \right)^{n+1} H_n^2 =
\]

\[
= -2\pi G \rho \sum_{n=0}^{\infty} \frac{n}{2n+1} \left( \frac{R}{r} \right)^{n+1} H_n^2.
\]

Now, also \( n = 1 \) gives a zero result, expected as gravity anomalies do not contain any components of degree number 1.

*Note* in this equation also a dependence upon \( n \): The gravity effect of Helmert condensation is dominated by short wavelengths or the very local features of the topography. The appearance of the square of height in this equation is again related to the terrain correction, in which also the square of the terrain height figures. When we are summing squares, leaving out terms will always cause a *systematic error*: also very short wavelengths, i.e., high values of \( n \), must be taken along in the summation.

**D.4.2 The interior potential of Helmert condensation**

This quantity is evaluated on the level of the geoid. It represents the indirect effect of Helmert condensation, i.e., the shift of the geoid surface caused by

\(^2\)As a curiosity, this result can be interpreted as the potential of a sphere of crustal matter with radius \( H \) (the average height of the topography) located at the geocentre. Even for a topographic mean height of 10 km the effect on the geoid would only be 12 mm (exercise!).
the mass shift in space.

\[ T_{\text{Helmert}}^{\text{int}} = T_{\text{Helmert}}^{\text{int}} - T_{\text{c}}^{\text{ext}} = \]

\[ = G\rho R^2 \int \int \sum_{n=0}^{\infty} \left( \frac{H}{R} - \frac{n - 1}{2} \left( \frac{H^2}{R^2} + \frac{(n - 1) n H^3}{6 R^3} \right) \right) P_n(\cos \psi) d\sigma - \]

\[ - G\rho R^2 \int \int \sum_{n=0}^{\infty} \left( \frac{H}{R} + \frac{H^2}{R^2} \right) P_n(\cos \psi) d\sigma = \]

\[ G\rho \int \int \sum_{n=0}^{\infty} \left[ - \frac{n + 1}{2} H^2 + \frac{(n - 1) n H^3}{6 R} - \ldots \right] P_n(\cos \psi) d\sigma. \]

Using again the definition of the degree constituents of the powers of height \( H \), equation (D.5):

\[ H_n^\nu = \frac{2n + 1}{4\pi} \int \int H^\nu P_n(\cos \psi) d\sigma, \]

we obtain

\[ T_{\text{Helmert}}^{\text{int}} = 4\pi G\rho \sum_{n=0}^{\infty} \left[ - \frac{n + 1}{2 (2n + 1)} H_n^2 + \frac{(n - 1) n H_n^3}{6 (2n + 1)} \right], \]

from which one obtains the indirect effect of Helmert condensation:

\[ \delta N_{\text{HC}} = - \frac{T_{\text{Helmert}}^{\text{int}}}{\gamma} = - \frac{4\pi G\rho}{\gamma} \sum_{n=0}^{\infty} \left[ - \frac{n + 1}{2 (2n + 1)} H_n^2 + \frac{(n - 1) n H_n^3}{6 (2n + 1)} \right] = \]

\[ = \frac{4\pi G\rho}{\gamma} \cdot \left[ \frac{1}{2} \sum_{n=0}^{\infty} \frac{n + 1}{2n + 1} H_n^2 - \frac{1}{6R} \sum_{n=0}^{\infty} \frac{(n - 1) n H_n^3}{2n + 1} \right]. \]  

(D.7)

The term \( n = 0 \) yields the indirect effect of a constant terrain \( H = \overline{H} = H_0 \): using only the first term inside the brackets yields

\[ \delta N_{\text{HC,Const}} = \frac{2\pi G\rho}{\gamma} \overline{H}^2, \]

which can not be neglected.

### D.5 The dipole method

As a sanity test, we may describe the effect of Helmert condensation in first approximation as a dipole layer field \( \mu \). The topographic mass, density \( \kappa = H\rho \), moves downward by on average \( \frac{1}{2} H \). The effect would be the same if mean sea level\(^3\) was covered by a double mass density layer

\[ \mu = \frac{1}{2} H^2 \rho. \]  

(D.8)

\(^3\)In fact, a better place for this layer would be level \( \frac{1}{4} H \). This is one of the approximations made here.
The potential of this layer is, in spherical approximation,

\[ T = G \int_S \mu \frac{\partial}{\partial n} \left( \frac{1}{r} \right) dS \approx GR^2 \int_\sigma \mu \frac{\partial}{\partial n} \left( \frac{1}{r} \right) d\sigma. \]

Written more explicitly:

\[ T_P = GR^2 \int_\sigma \mu_Q \frac{\partial}{\partial \sigma_Q} \left( \frac{1}{r_{PQ}} \right) d\sigma_Q. \]

We use the expansion into Legendre polynomials, equation (7.4):

\[ \frac{1}{r_{PQ}} = \frac{1}{r_Q} \sum_{n=0}^{\infty} \left( \frac{r_Q}{r_P} \right)^n P_n (\cos \psi_{PQ}), \]

differentiate with respect to \( r_Q \), and substitute:

\[ T_P = GR^2 \int_\sigma \frac{1}{r_Q} \mu_Q \sum_{n=0}^{\infty} n \left( \frac{r_Q}{r_P} \right)^n P_n (\cos \psi_{PQ}) d\sigma_Q. \]

By substituting into this the equation (D.8) for the \( \mu_Q \) double mass density layer, we obtain, by taking the limit \( r_P, r_Q \downarrow R \):

\[ T = \frac{1}{4\pi} \sum_{n=0}^{\infty} n \int_\sigma (2\pi G \rho) H P_n (\cos \psi) d\sigma = \frac{1}{4\pi} \sum_{n=0}^{\infty} n \int A_B H P_n (\cos \psi) d\sigma. \]

Here, we have left the designations \( P, Q \) again off as no longer needed for clarity.

The symbol \( A_B \) denotes the attraction of a Bouguer plate of thickness \( H \) and density \( \rho \).

Let us develop the quantity \([A_B H]\) into a spherical harmonic expansion (Heiskanen and Moritz, 1967 equation 1-71). Then, each degree constituent is (equation (2.19)):

\[ [A_B H]_n = \frac{2n + 1}{4\pi} \int_\sigma [A_B H] P_n (\cos \psi) d\sigma, \]

in which case (note that the term \( n = 0 \) vanishes):

\[ T = \sum_{n=1}^{\infty} n \frac{n}{2n + 1} [A_B H]_n \approx \frac{1}{2} [A_B H], \]

at least for the higher \( n \) values, i.e., regionally if not globally.

Thus is obtained again the indirect effect of Helmert condensation, in geoid computation by means of this method the shift in geoid surface caused by the condensation, which must be accounted for with opposite algebraic sign.

In other words, looked upon as a \textit{Remove-Restore} method, its “Restore” step:

\[ \delta N_{HC} = \frac{T}{\gamma} \approx \frac{1}{2} \frac{A_B H}{\gamma} = \frac{\pi G \rho H^2}{\gamma}. \]
For comparison, the more precise expansion (D.7) yields in approximation for larger $n$ values

$$\delta N_{HC} \approx \frac{4\pi G \rho}{\gamma} \cdot \frac{1}{2} \sum_{n=0}^{\infty} \frac{n+1}{2n+1} H_n^2 \approx \frac{\pi G \rho H^2}{\gamma},$$

essentially the same result.
Åland, 246
Académie française, 86
acceleration
   aircraft, 212
accelerometer, 214, 249–251
action at a distance, 1, 29
adjustment
   crossover, 248
Agulhas Stream, 217
air drag compensation, 214
Airy, George Biddell, 97
Airy-Heiskanen hypothesis, 97, 103
algorithm
   recursive, 38
altimetry, satellite, 169
Amazonas (Brazil), 164, 251
American Geophysical Union (AGU), 262
amplifier, 127
amplitude, 223
tidal, 258
Amsterdam, The Netherlands, 110
analysis
   a posteriori, 246
angular distance, 129, 132, 178, 228
angular momentum, 272, 274
   conservation of, 274
anisotropy, 132
Antarctic Sea, 246
Antarctica, 164, 213, 226, 229
Apollo project, 3
Arabelos, Dimitris, 162
Archimedes’ law, 99
Arctic, 164, 213
Arctic Sea, 246, 247
arrest (gravimeter), 203
asthenosphere, 222, 227
Astronomer Royal, 97
Atlantic Ocean
   North, 219
atmosphere
   attraction, 210
   mass, 211
atmospheric drag, 249–251
atmospheric tidal loading, 260
autocovariance, 171
average
   a posteriori, 248
axis
   semi-major, 36
   semi-minor, 36
azimuth, 173, 174

Baltic Sea, 213, 219, 223, 232
barometer, 77
  inverted, 217
Barzaghi, Riccardo, 161
base, 285, 288
  complete, 286, 288, 292
  orthogonal, 291
  orthonormal, 4, 8, 49, 286, 291
base vector, 285, 286
bathymetry, 93
bench mark, 236
Bergensbanen, 100
Bessel function, 34
Bessel’s equation, 291
Bessel, Friedrich Wilhelm, 34
bias, 239
bilinear form, 285, 288
Bjerhammar sphere, 191
Bjerhammar, Arne, 191
block, gravity anomaly averaging, 147
blue film, Earth’s, 250
Blue Road Geotraverse project, 222
bordering, 161
Bose-Einstein condensate, 126, 207
Bothnian Bay, 223
Bouguer hypothesis, 221, 222
Bouguer plate, 1, 87, 88, 92, 106, 118, 119, 210, 302
Bouguer plate correction, 92
Bouguer reduction, 102, 143
Bouguer shell, 94
Bouguer, Pierre, 86, 95
Boulder, University of Colorado at, 204, 233
boundary condition, 32, 80
boundary surface, 85

boundary-value problem, 26, 27, 30, 85
free, 77
  of Dirichlet, 27, 79
  of Neumann, 46, 79
  of physical geodesy, 79
  spectral solution, 80
  third, 78, 79
box
  rectangular, 19
brick
  Legotm, 20
Brovelli, Maria, 161
Bruns equation, 61, 75, 129
Bruns, Ernst Heinrich, 61, 75
calculus
  vector, 271
calibration
  radar altimeter, 234
Canada, 227
cantilever beam, 200
cantilever motion, 227
carbon dioxide, 224
cat
  Schrödinger’s, 30
Cavendish, Henry, 3
centre of mass
  of Earth, 129
  of the Earth, 49, 50, 220
CHAMP (satellite), 51, 164, 248, 249
Chasles theorem, 1, 26, 28
Chasles, Michel, 26
checkerboard, 41
China, 232
clock
  climate, of Earth, 225
  clinometer, 258
atomic, 126
optical lattice, 126
closing error, 203
co-geoid
  of isostatic reduction, 103
co-ordinate transformation, 271
co-ordinates
  ellipsoidal, 36, 51
  geodetic, 35, 77
  geographical, 34, 65
  rectangular, 30
  spherical, 34, 37
co-rotating reference system, 56
coastline file, 246
Collocation, Fast, 194
collocation, least-squares (LSC), 103, 175, 178, 180, 183, 191
Columbus geoid (geoid model), 263
commutative diagram, 143, 155
commutativity, 285
compensation depth, 105
condensation layer, 299
Congo (Africa), 251
conservative force field, 4, 110
continental shelf, 100
convection
  in the Earth’s mantle, 104
convection current, 104
convolution, 12, 151, 166, 167, 193, 228
convolution theorem, 153, 158, 159, 161
Copenhagen, Denmark, 262
Coriolis effect, 223
Coriolis force, 218, 224
Coriolis, Gaspard-Gustave, 57
correlation, 177
correlation length, 178, 179
covariance function, 173, 174, 185, 188
  empirical, 190
  Gauss-Markov, 179
  Hirvonen’s, 178
  in space, 189
  on the surface, 189
cross covariance, 171
crossover adjustment, 235
crossover point, 236, 238
Cryosat-2 (satellite), 232
curvature
  of equipotential surface, 215
damping, of gravimeter, 204, 213
Danish straits, 223
Darwin, Sir George, 256
data point, 129, 132, 151
datum, 238
datum defect, 238
datum point, 110
datum transformation, 238, 240
De Maupertuis, Pierre, 86
Defense Mapping Agency (DMA), 50
definition
  for viscous loading, 230
deglaciation, 102, 230
degree constituent, 190
  of power of height, 299, 301
degree constituent equation, 47, 105, 129, 136, 137, 187, 188, 299
degree variance, 186, 189, 192
  of gravity anomalies, 190
degree variance formula, 191
degree, harmonic, 37
degrees of freedom, 240
delta function, Dirac’s, 21, 136, 193
density
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>crustal</td>
<td>122</td>
</tr>
<tr>
<td>rock</td>
<td>121</td>
</tr>
<tr>
<td>standard crustal</td>
<td>121</td>
</tr>
<tr>
<td>dice</td>
<td>172</td>
</tr>
<tr>
<td>differential equation</td>
<td></td>
</tr>
<tr>
<td>self-adjoint</td>
<td>290</td>
</tr>
<tr>
<td>dipole</td>
<td>15, 49, 50</td>
</tr>
<tr>
<td>potential field of dipole</td>
<td>49</td>
</tr>
<tr>
<td>dipole layer</td>
<td>301</td>
</tr>
<tr>
<td>dipole moment</td>
<td>49, 63</td>
</tr>
<tr>
<td>Dirichlet’s principle</td>
<td>27</td>
</tr>
<tr>
<td>Dirichlet’s problem</td>
<td>1, 27</td>
</tr>
<tr>
<td>Dirichlet, Peter Gustav Lejeune</td>
<td>27</td>
</tr>
<tr>
<td>disk</td>
<td></td>
</tr>
<tr>
<td>circular</td>
<td>87</td>
</tr>
<tr>
<td>dislocation (crystal)</td>
<td>203</td>
</tr>
<tr>
<td>distance</td>
<td></td>
</tr>
<tr>
<td>angular</td>
<td>173, 174</td>
</tr>
<tr>
<td>divergence</td>
<td>12, 17</td>
</tr>
<tr>
<td>domain</td>
<td>171</td>
</tr>
<tr>
<td>Doodson, Arthur Thomas</td>
<td>255</td>
</tr>
<tr>
<td>DORIS (instrument)</td>
<td>232, 233</td>
</tr>
<tr>
<td>downward continuation</td>
<td>164</td>
</tr>
<tr>
<td>downward continuation, harmonic</td>
<td>118</td>
</tr>
<tr>
<td>134, 135, 139, 298</td>
<td></td>
</tr>
<tr>
<td>drift (gravimeter)</td>
<td>203, 210</td>
</tr>
<tr>
<td>Eötvös (unit)</td>
<td>59, 215</td>
</tr>
<tr>
<td>Eötvös, Loránd</td>
<td>58, 215</td>
</tr>
<tr>
<td>earthquake</td>
<td>210</td>
</tr>
<tr>
<td>eddy</td>
<td>233</td>
</tr>
<tr>
<td>EGM2008</td>
<td>42, 51</td>
</tr>
<tr>
<td>EGM96</td>
<td>50–52</td>
</tr>
<tr>
<td>Eiffel Tower</td>
<td>12, 13, 26, 32, 38, 57, 61, 291</td>
</tr>
<tr>
<td>Eiffel Tower’s 72 names</td>
<td>32</td>
</tr>
<tr>
<td>eigenvalue</td>
<td>288, 289, 291</td>
</tr>
<tr>
<td>eigenvalue problem</td>
<td>287, 288, 290</td>
</tr>
<tr>
<td>eigenvector</td>
<td>288, 289, 291</td>
</tr>
<tr>
<td>Einstein summation convention</td>
<td>177</td>
</tr>
<tr>
<td>El Niño Southern Oscillation (ENSO)</td>
<td>217, 233</td>
</tr>
<tr>
<td>elasticity</td>
<td>203</td>
</tr>
<tr>
<td>electric charge</td>
<td>274</td>
</tr>
<tr>
<td>electric currents in the Earth’s core</td>
<td>104</td>
</tr>
<tr>
<td>ellipsoidal harmonic expansion</td>
<td>51, 62</td>
</tr>
<tr>
<td>ellipsoid Hayford</td>
<td>83</td>
</tr>
<tr>
<td>Envisat (satellite)</td>
<td>232</td>
</tr>
<tr>
<td>equations of motion</td>
<td></td>
</tr>
<tr>
<td>of satellites, 49</td>
<td></td>
</tr>
<tr>
<td>equatorial radius (Earth ellipsoid)</td>
<td>66</td>
</tr>
<tr>
<td>equatorial radius (Earth)</td>
<td>36, 50, 77, 241, 242</td>
</tr>
<tr>
<td>equipotential surface</td>
<td>55, 58, 59, 65, 74, 120, 218, 247</td>
</tr>
<tr>
<td>equivalence principle</td>
<td>3, 58</td>
</tr>
<tr>
<td>ergodicity</td>
<td>173</td>
</tr>
<tr>
<td>error ellipse</td>
<td>289</td>
</tr>
<tr>
<td>ERS-1 (satellite)</td>
<td>232</td>
</tr>
<tr>
<td>ERS-2 (satellite)</td>
<td>232</td>
</tr>
<tr>
<td>estimate</td>
<td></td>
</tr>
<tr>
<td>a priori</td>
<td>183</td>
</tr>
<tr>
<td>estimation</td>
<td>175</td>
</tr>
<tr>
<td>estimator</td>
<td>176</td>
</tr>
<tr>
<td>Eterna (software)</td>
<td>260</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>164</td>
</tr>
<tr>
<td>Euclidean space</td>
<td>285</td>
</tr>
<tr>
<td>Euler’s identity</td>
<td>296</td>
</tr>
<tr>
<td>Eurajoki, Finland</td>
<td>208</td>
</tr>
</tbody>
</table>
Europe, 225
  Southern, 110
  Western, 110
European Geosciences Union (EGU), 262
European Space Agency (ESA), 232, 251
eustatic rise, of mean sea level, 220, 221
evaluation point, 129, 132, 151
Everest, Mount, 100
expectancy, 171
  of a stochastic process, 173
exterior normal, of a surface, 17
factorial, 38
Falkland islands, 251
Faller, James E., 204
fast collocation, 103
Fast Fourier Transform (FFT), 12, 103,
  143, 157, 162, 164, 165, 167,
  191, 193, 219, 295
mixed-radius, 295
Fastest Fourier Transform in the West
  (fftw, software), 296
Fennoscandia, 101, 219, 229, 230
Fennoscandian Shield, 103
field, 1, 29
  harmonic, 13, 27
  vector, 271
field equations, 1
  of electromagnetism, 13
  of gravitation, 13
field theory, 1
  electromagnetism, 30
figure of the Earth, 242
FIN2000 (geoid model), 219
FIN2005No0 (geoid model), 263
finite elements, 103
Finland, 225
Finnish Geodetic Institute (FGI), 204,
  209, 258, 263
Finnish Geospatial Research Institute
  (FGI), 263
first eccentricity, 36
flattening, 36
dynamic
  of Earth, 242
  of a planet, 10
  of Earth, 34, 35, 51–53, 93
  of Earth ellipsoid, 66
footprint
  radar altimeter, 233, 234, 246
footscrew, 201
force
  centrifugal, 56, 63
  Coriolis, 57
Forsberg, René, 161, 162
Fourier coefficient, 33, 37
Fourier expansion, 286
Fourier function, 286
Fourier series, 286, 296
Fourier sine expansion, 32
Fourier transform, 161, 166
  discrete, 161, 295, 296
  inverse, 295
  inverse, 167
Fourier, Joseph, 32
free oscillations, of the solid Earth,
  210, 263
free-air hypothesis, 221
free-air reduction, 92
freeboard, 232, 246, 247
function space, 286, 291
function theory, 291
functional, 170
  linear, 27, 171, 188
fundamental equation of physical geodesy,
  80, 141
Galilei, Galileo, 3
Gauss integral theorem, 1, 17, 18, 20, 21, 23, 26
Gauss-Markov process, 179
Gelder, Martin van, 161
General Relativity, 13
generating function, 105, 133
  of the Legendre polynomials, 133
geodesy, 24, 27, 51, 56, 247
generating function, 105, 133
  of the Legendre polynomials, 133
geometric, 69
  physical, 8, 13, 27, 34, 69, 79, 80, 170, 264
  space, 51
goedetic forward problem
  on the sphere, 173
Geodetic Reference System 1967 (GRS67), 67
Geodetic Reference System 1980 (GRS80),
  4, 66, 69, 211, 219
geodynamics, 263
geographic mean, 172–174
geographic variance, 172
geoid, 47, 58, 79, 111, 120, 217, 218
  free-air, 118
goid determination, 127, 139
  gravimetric, 130, 139, 223
ged height, 73, 118, 123, 129, 233
ged model, 75, 162, 219, 249
  global, 248
  local, 218
ged rise, 221
gedological map, 122
gophysical data record (GDR), 235,
covariance function, 195
  free-air, 121
  vertical, 83
gradiometer, gravity, 251, 252
gravimeter, 59
  absolute, 204, 207, 222
  airborne, 204, 212
  astatized, 200, 201
  atomic, 207, 208
  ballistic, 198, 204, 208
  IMGC-02, 207
  JILA, 204
  LaCoste-Romberg, 200
  pendulum, 197
  quantum, 207
  registering, 260
  relative, 205, 222
  sea, 204
  ship, 248
  spring, 198, 203, 207
  superconducting, 208
gravimetry, 28
  airborne, 164, 212, 214, 248, 249
  satellite, 214
gravitation, 46, 56
gravitational field, 56
gravitational lense, 2
gravity, 56
  anomalous, 56
  equatorial, 66
  normal, 56, 64, 66, 82, 122
gravity acceleration, 47
gravity anomaly, 37, 77, 78, 82, 132, 136, 140, 160, 170, 172, 185, 211, 221
  Bouguer, 88, 91, 92, 102, 119, 123
  computation, 91
  simple, 88
  spherical, 93
  terrain corrected, 89
  free-air, 82–84, 86, 88, 92, 102, 103, 123
  isostatic, 102, 103
  gravity disturbance, 76, 132
  spectral representation, 77
  gravity field, 56
  ellipsoidal, 69
  normal, 74
  of Earth, 169
  gravity formula, 55, 65, 77, 83
  gravity reduction, 139
  gravity-gradient kernel, 138
  gravity-gradient tensor, 214
  GRAVSOFT (software), 162
  Green’s function, 228
  of sea level, 229
  of the geopotential, 229
  of vertical displacement, 229
  Green’s theorems, 1
  first, 22
  second, 22
  third, 22, 85
Greenland, 164, 226, 229
Greenland Aerogeophysics Project, 213
Greenwich meridian, 35
grid integration, 103
Gulf of Finland, 213, 223
Gulf Stream, 217, 224–226
Guyana, French, 197
GWR iGrav (gravimeter), 210
Haagmans, Roger, 161
half-height point, 246
Hannover, Germany, 262
Harangervidden plateau (Norway), 100
harmonic continuation, 136
harmonic oscillator, 31
Hayford ellipsoid, 67
Hayford, John Fillmore, 96
heat transport, 224, 225
Heaviside function, 230
height
above the reference ellipsoid, 120
Helmert, 118, 121, 122
mean, of satellite, 241
normal, 112, 116, 118, 120, 122, 123, 125
orthometric, 62, 73, 111, 118–120, 122–124
height anomaly, 81, 82, 113, 116, 120, 123
height system, 263
national, 219
height, orbital, 242
Heiskanen, Weikko A., 97, 263
helicopter, 213
helium
liquid, 209, 210
Helmert condensation, 1, 94, 139, 297, 299, 300
Helmert heights, 219
Helmert’s formula, 67, 120
Helmert, Friedrich Robert, 94
Helsinki, 219
Helsinki astronomical observatory, 110, 111
Helsinki harbour, 110, 219, 238
Helsinki University of Technology (HUT, TKK), 263
Helsinki, Finland, 110, 219, 238
hierarchy, network, 207
Himalayas, 95
Hirvonen, Reino Antero, 178
Hofmann-Wellenhof, Bernhard, 264
homogeneity, 172–174, 194
Hubble Space Telescope, 2
Huygens, Christiaan, 197
HY-2A (satellite), 232
ice age, 230
ice cap, 220
ice sheet
continental, 101, 220, 226
Fennoscandian, 227
Laurentide, 227
ice-load history, 229
inclination, orbital, 242, 244
India, 232
indirect effect, 86, 104, 139, 140, 257
of a constant terrain, 301
of Helmert condensation, 300, 301
of isostatic reduction, 106
Indonesia, 246, 251
inertial reference system, 56
information
a priori, 239
Institut für Erdmessung, 262
integrability, 292
integral theorem, 1
integration by parts, 290
intercomparison, gravimeter, 207–209
International Association of Geodesy (IAG), 261
International Center for Earth Tides (ICET), 262
International Center for Global Earth Models (ICGEM), 262
International Digital Elevation Model Service (IDEMS), 262
International ellipsoid (1924), 67
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Geoid Commission (IGeC) Kater, Henry</td>
<td>198</td>
</tr>
<tr>
<td>International Gravity Bureau (Bureau Gravimétrique International, BGI)</td>
<td>261</td>
</tr>
<tr>
<td>International Gravity Field Service (IGFS)</td>
<td>261</td>
</tr>
<tr>
<td>International Isostatic Institute</td>
<td>263</td>
</tr>
<tr>
<td>International Service for the Geoid (ISG)</td>
<td>261</td>
</tr>
<tr>
<td>International Union of Geodesy and Geophysics (IUGG)</td>
<td>261</td>
</tr>
<tr>
<td>Internet</td>
<td>127</td>
</tr>
<tr>
<td>invariance</td>
<td></td>
</tr>
<tr>
<td>gauge</td>
<td>13</td>
</tr>
<tr>
<td>invariant</td>
<td>290</td>
</tr>
<tr>
<td>inversion</td>
<td>252</td>
</tr>
<tr>
<td>ionospheric propagation delay</td>
<td>234</td>
</tr>
<tr>
<td>isostasy</td>
<td>96, 100</td>
</tr>
<tr>
<td>isostatic co-geoid</td>
<td>107</td>
</tr>
<tr>
<td>isostatic compensation</td>
<td>96, 97, 99, 102, 104</td>
</tr>
<tr>
<td>isostatic equilibrium</td>
<td>101</td>
</tr>
<tr>
<td>isostatic geoid</td>
<td>103, 104</td>
</tr>
<tr>
<td>isostatic hypothesis</td>
<td>103</td>
</tr>
<tr>
<td>isostatic reduction</td>
<td>102–104, 143</td>
</tr>
<tr>
<td>isotropy</td>
<td>7, 132, 174, 178, 186, 229</td>
</tr>
<tr>
<td>iteration</td>
<td>112</td>
</tr>
<tr>
<td>Jacobian</td>
<td>131, 151</td>
</tr>
<tr>
<td>Jason (satellites)</td>
<td>243</td>
</tr>
<tr>
<td>Jason-1 (satellite)</td>
<td>232</td>
</tr>
<tr>
<td>Jason-2 (satellite)</td>
<td>232</td>
</tr>
<tr>
<td>Java Sea, Indonesia</td>
<td>198</td>
</tr>
<tr>
<td>Jerry (mouse)</td>
<td>250</td>
</tr>
<tr>
<td>Joensuu, Finland</td>
<td>208</td>
</tr>
<tr>
<td>Kaivopuisto (Helsinki)</td>
<td>110, 111, 219</td>
</tr>
<tr>
<td>Kaula, William</td>
<td>191</td>
</tr>
<tr>
<td>Kepler’s laws of orbital motion</td>
<td>241</td>
</tr>
<tr>
<td>Kepler’s orbital elements</td>
<td>241</td>
</tr>
<tr>
<td>Kepler’s second law</td>
<td>273, 274</td>
</tr>
<tr>
<td>Kepler’s third law</td>
<td>240, 241, 244</td>
</tr>
<tr>
<td>Kepler, Johannes</td>
<td>148, 241</td>
</tr>
<tr>
<td>Kevo, Finland</td>
<td>208</td>
</tr>
<tr>
<td>Kirkkonummi, Finland</td>
<td>219</td>
</tr>
<tr>
<td>KKJ</td>
<td>83</td>
</tr>
<tr>
<td>Knudsen, Per</td>
<td>162</td>
</tr>
<tr>
<td>Krasowsky’s ellipsoid</td>
<td>67</td>
</tr>
<tr>
<td>Kronecker delta</td>
<td>193</td>
</tr>
<tr>
<td>Kronstadt, Russia</td>
<td>110</td>
</tr>
<tr>
<td>Kuusamo, Finland</td>
<td>208</td>
</tr>
<tr>
<td>LaCoste, Lucien</td>
<td>201</td>
</tr>
<tr>
<td>Lageos (satellite)</td>
<td>241</td>
</tr>
<tr>
<td>land uplift</td>
<td></td>
</tr>
<tr>
<td>absolute</td>
<td>220, 221</td>
</tr>
<tr>
<td>Fennoscandian</td>
<td>101</td>
</tr>
<tr>
<td>post-glacial</td>
<td>219, 220, 262, 263</td>
</tr>
<tr>
<td>mechanism</td>
<td>222</td>
</tr>
<tr>
<td>relative</td>
<td>220, 221</td>
</tr>
<tr>
<td>Laplace equation</td>
<td>12, 29, 30, 33, 37, 51, 216</td>
</tr>
<tr>
<td>Laplace operator</td>
<td>29, 170</td>
</tr>
<tr>
<td>Laplace tidal equation</td>
<td>254</td>
</tr>
<tr>
<td>Laplace, Pierre Simon</td>
<td>12</td>
</tr>
<tr>
<td>Lapland</td>
<td>86</td>
</tr>
<tr>
<td>laser interferometer</td>
<td>204</td>
</tr>
<tr>
<td>last glacial maximum</td>
<td>101</td>
</tr>
<tr>
<td>latitude</td>
<td></td>
</tr>
<tr>
<td>astronomical</td>
<td>61, 73</td>
</tr>
<tr>
<td>geocentric</td>
<td>35, 37, 50, 65</td>
</tr>
</tbody>
</table>
geodetic, 65, 157
geographic, 65, 157
reduced, 36, 65
reference, 158, 159
least-squares adjustment, 238
least-squares method, 220, 236
least-squares principle, 178
Legendre function, 37, 38
associated, 40, 43
fully normalized, 293
of the second kind, 51
Legendre polynomial, 38, 291, 302
Legendre’s equation, 291
Legendre, Adrien-Marie, 38
Leibniz University, 262
level surface, 59
curvature of, 60
level, bull’s eye, 201
levelling, 109, 236
goestrophic, 223
precise, 127, 223
relativistic, 126
steric, 223
levelling staff, 109
levelling, of gravimeter, 201
linear algebra, 288
linear regression, 233
linearity, 31, 285, 287
linearization, 82
Liouville, Joseph, 291
Lohja, Finland, 258
long water-tube clinometer, 258
longitude
astronomical, 61, 73
Lunar, 255
Love number, 256–259
Love, Augustus, 256
magnetic field
of Earth, 104, 209, 213
Maledives, 227
mareograph (tide gauge), 220
Marussi tensor, 214
Marussi, Antonio, 214
mass density layer, 1, 106, 139
double, 15, 25, 302
single, 14, 16, 17, 25, 94, 104, 105
surface, 130
matrix
symmetric, 288
Maxwell, James Clerk, 13, 30
mean geoid, 259
mean sea level, 111, 120, 217–219
global, 218–220
mean sea surface, 232, 242
mechanics
celestial, 10
quantum, 30
Mekong (river), 251
Melchior, Paul, 255
mesoscale eddies, 217, 226
metal fatigue, 203
Metsähovi research station, Finland, 110, 208–210, 219
microgal, 59
microseismicity, 203, 204
microwave link, 250
mid-Holocene highstand, 227
Milano, Italy, 261
milligal, 59
Min, Erik de, 161
mirror symmetry, 44
missile, ballistic, 232
modification coefficients, 146
modification degree, 144
<table>
<thead>
<tr>
<th>Phrase</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mohorovičić, Andrija</td>
<td>107</td>
</tr>
<tr>
<td>Mohorovičić discontinuity</td>
<td>167</td>
</tr>
<tr>
<td>Mohorovičić discontinuity (&quot;Moho&quot;)</td>
<td>107</td>
</tr>
<tr>
<td>Molodensky’s method</td>
<td>138</td>
</tr>
<tr>
<td>Molodensky’s proof</td>
<td>116</td>
</tr>
<tr>
<td>Molodensky, Mikhail Sergeevich</td>
<td>26, 81, 85, 113, 264</td>
</tr>
<tr>
<td>momentum, linear</td>
<td>272</td>
</tr>
<tr>
<td>Mongolia</td>
<td>164</td>
</tr>
<tr>
<td>monopole</td>
<td>49, 50</td>
</tr>
<tr>
<td>monsoon</td>
<td>250</td>
</tr>
<tr>
<td>Moon</td>
<td>218</td>
</tr>
<tr>
<td>Moritz, Helmut</td>
<td>264</td>
</tr>
<tr>
<td>mu-metal</td>
<td>208</td>
</tr>
<tr>
<td>Munk, Walter</td>
<td>232</td>
</tr>
<tr>
<td>N2000 (height system)</td>
<td>110, 164, 219</td>
</tr>
<tr>
<td>N60 (height system)</td>
<td>110, 164, 219</td>
</tr>
<tr>
<td>nabla</td>
<td>8</td>
</tr>
<tr>
<td>National Geospatial-Intelligence Agency</td>
<td>50, 261</td>
</tr>
<tr>
<td>(NGA)</td>
<td>10</td>
</tr>
<tr>
<td>National Imagery and Mapping Agency</td>
<td>50</td>
</tr>
<tr>
<td>(NIMA)</td>
<td>245</td>
</tr>
<tr>
<td>National Land Survey of Finland</td>
<td>210, 263</td>
</tr>
<tr>
<td>natural co-ordinates</td>
<td>62</td>
</tr>
<tr>
<td>Navy, United States</td>
<td>232</td>
</tr>
<tr>
<td>Neumann, Carl Gottfried</td>
<td>46</td>
</tr>
<tr>
<td>Newton’s law of gravitation</td>
<td>2, 166, 215</td>
</tr>
<tr>
<td>Newton’s theory of gravitation</td>
<td>1, 13</td>
</tr>
<tr>
<td>Newton, Isaac</td>
<td>2</td>
</tr>
<tr>
<td>Niethammer’s method</td>
<td>121</td>
</tr>
<tr>
<td>Niethammer, Theodor</td>
<td>121</td>
</tr>
<tr>
<td>NKG WG for Geoid and Height Systems</td>
<td>262</td>
</tr>
<tr>
<td>NKG2004 (geoid model)</td>
<td>262, 263</td>
</tr>
<tr>
<td>NKG2015 (geoid model)</td>
<td>263</td>
</tr>
<tr>
<td>NN (height system)</td>
<td>110</td>
</tr>
<tr>
<td>noise</td>
<td>176</td>
</tr>
<tr>
<td>observational</td>
<td>177</td>
</tr>
<tr>
<td>Nordiska Kommissionen för Geodesin (NKG)</td>
<td>262</td>
</tr>
<tr>
<td>norm of a vector</td>
<td>272, 286</td>
</tr>
<tr>
<td>Normaal Amsterdams Peil (NAP)</td>
<td>110</td>
</tr>
<tr>
<td>normal</td>
<td></td>
</tr>
<tr>
<td>ellipsoidal</td>
<td>56, 113</td>
</tr>
<tr>
<td>normal correction</td>
<td>125</td>
</tr>
<tr>
<td>normal derivative</td>
<td>24</td>
</tr>
<tr>
<td>normal equation</td>
<td>206</td>
</tr>
<tr>
<td>North American Vertical Datum 1988 (NAVD88)</td>
<td>110</td>
</tr>
<tr>
<td>Norwegian Sea</td>
<td>100</td>
</tr>
<tr>
<td>ocean tidal loading</td>
<td>260</td>
</tr>
<tr>
<td>National Imagery and Mapping Agency oceanography</td>
<td>259</td>
</tr>
<tr>
<td>oceanography</td>
<td>259</td>
</tr>
<tr>
<td>National Land Survey of Finland</td>
<td>210, 263</td>
</tr>
<tr>
<td>Ohio State University</td>
<td>50, 263</td>
</tr>
<tr>
<td>Ohio, United States of America</td>
<td>179</td>
</tr>
<tr>
<td>national co-ordinates</td>
<td>62</td>
</tr>
<tr>
<td>operator</td>
<td></td>
</tr>
<tr>
<td>Neumann, Carl Gottfried</td>
<td>46</td>
</tr>
<tr>
<td>Newton’s law of gravitation</td>
<td>2, 166, 215</td>
</tr>
<tr>
<td>Newton’s theory of gravitation</td>
<td>1, 13</td>
</tr>
<tr>
<td>Newton, Isaac</td>
<td>2</td>
</tr>
<tr>
<td>Niethammer’s method</td>
<td>121</td>
</tr>
<tr>
<td>Niethammer, Theodor</td>
<td>121</td>
</tr>
<tr>
<td>NKG WG for Geoid and Height Systems</td>
<td>262</td>
</tr>
<tr>
<td>35-day</td>
<td>244, 245</td>
</tr>
<tr>
<td>no-shadow</td>
<td>242, 243</td>
</tr>
</tbody>
</table>
retrograde, 242, 244
satellite, 241, 242
Sun-stationary, 242–244, 251
Sun-synchronous, 242
three-day, 244, 245
orbit error, 235, 237
satellite, 235
orbit perturbation
satellite, 42, 50, 169
orbit-error correction, 233, 235
order, harmonic, 37
orthogonal function, 291
orthogonality, 144, 187, 189, 285
  of Legendre polynomials, 44, 48
orthometric correction, 124, 125
orthonormality, 187, 285
oscillation equation, 199, 202, 290, 291
Ostrobothnia (Finland), 219
OSU model, 50
Päijänne, Lake, 113
Pacific Ocean, 217
parallelogram, 273
partial potentials, 70
Peltier effect, 227
Peltier, W. Richard, 227
pendulum, 197
period
  free oscillations of the Earth, 210
  orbital, 241
  oscillation, 199, 202, 203
  theoretical tide, 256, 258
  tidal, 254, 258
periodicity, 161
Peru, 86
phase
  tidal, 258
Pizzetti, Paolo, 65
plasma, 127
Plesetsk Cosmodrome (Russia), 249, 251
plumbline, 56, 59, 60, 111, 122, 212, 258
deflection of the, 56, 73, 95, 131, 170, 232
direction of the, 61
absolute, 258
point mass, 48
point variance, 290
Pointe-au-Père (Rimouski, Quebec, Canada), 110
Poisson equation, 13, 20, 58, 122
Poisson kernel, 136, 138
Poisson, Siméon Denis, 13
polar co-ordinates, 33
circular motion, 258
circular radius (Earth ellipsoid), 66
polynomial, 220
potential, 4, 19
  anomalous, 56
centrifugal, 56–58
disturbing, 37, 70, 73, 129, 132, 170, 185, 188
gravitational, 1, 58
gravity, 56, 58, 59, 69, 70, 207, 257, 282
normal, 55, 62, 66
normal gravitational, 56, 69
normal gravity, 56, 64, 73
rotational, 56
tidal, 253, 258
true gravity, 73
potential energy, 8
Potsdam system, 198, 208
PRARE (instrument), 232
Pratt, John Henry, 96
Pratt-Hayford hypothesis, 96
precession
  orbital, 242
prediction, 102, 175, 184
  homogeneous, 183
orbit, 238
Prey reduction, 121
Prey, Adalbert, 121
principal axes
  of error ellipse, 289
prism integration, 90, 104
product
  dot, 271
  exterior, 272
  scalar, 146, 147, 271, 272, 285, 286, 289, 291, 292
  vectorial, 272, 273
projection
  stereographic, 152
propagation of covariances, 190
propagation of variances, 177, 185
propeller aircraft, 213
pseudo-force, 212
quadrature, 147
quadrupole, 50
quantum number, 291
quasi-geoid, 82, 112, 117–119
quasi-geoid height, 114, 123
quasi-geoid model, 162
radar altimeter, 213
radar altimetry
  satellite, 223, 225, 233
radar, microwave, 231
radius of curvature
  transversal, 35
Raman effect, 207
range, measured, 233
Rapp, Richard H., 50
recursive computation, 43
reduction
  Bouguer, 86
  geophysical, 85
reduction to sea level, 138
redundancy, 239
reference ellipsoid, 35, 55, 63–66, 69, 74, 122, 233
refraction
  atmospheric, 109
regularization, 167
  Tikhonov, 238, 240
Relativity, General, 126
relaxation time, 230
“Remove” step, 140, 153
Remove-Restore method, 86, 142, 143, 219, 302
renormalization, 167
repeat period, 241
repeat track, 248
residual gravity field, 142
Residual Terrain Modelling (RTM), 139, 140
resolution
  of measurement, 169
  of spherical harmonic expansion, 42
resonance, 243
“Restore” step, 104
“Restore” step, 139, 140, 153, 302
retracking, 246, 247
reversion pendulum, 198
rheology, 229
Richer, Jean, 197
Index

rising dough model, 221
Robin, Victor Gustave, 79
rocket engine, 251
ionic, 251
Romberg, Arnold, 202
Romberg, Werner, 148
root (of mountain), 95
rotation rate (Earth), 66
rotational symmetry, 44, 63, 191
salinity (sea water), 218, 224
salinity gradient, 223
San Francisco, United States of America, 262
Sandwell, David, 164
Sansò, Fernando, 161
SARAL/AltiKa (satellite), 232
satellite altimetry, 231, 247
satellite geodesy, 191
satellite laser ranging (SLR), 241
satellite orbit, 233
satellite-to-satellite tracking (SST), 250
scalar product, 288
Schrödinger equation, 30
Schrödinger, Erwin, 30
Schwarz, Klaus Peter, 161
Schwarzschild metric, 126
sea-floor pressure, 250
sea-level equation, 227
sea-level rise, 232, 233
global, 224
sea-surface variability, 235, 242, 248
Seasat (satellite), 232
secular effect, 218
seismicity, 230
seismology, 28, 96
seismometer, 210
semi-major axis, 241
sensitivity, instrument, 200, 201
separation of variables, 31
shoebox world, 32
Sideris, Michael, 161
signal, 176
signal variance, 178
signal variance matrix, 175
significant wave height (SWH), 233, 234
Simpson’s rule, 148
Simpson, Thomas, 148
singularity, 238
sink, 18
Skylab (space station), 232
Sodankylä, Finland, 208
solar panel, 242
Somigliana, Carlo, 65
Somigliana-Pizzetti equation, 65
source, 18
Soviet Union, 85
space
   Euclidean, 4, 8
spectral coefficients, 37
spectral expansion
   of geopotential, 45
spectral form, 129, 130, 132, 135
spectral representation, 132, 186, 189
sphere
   celestial, 292
spherical co-ordinates, 35
Spherical FFT, 157, 158
spherical harmonic, 291
   sectorial, 41
solid, 37
surface, 37, 41, 42, 44, 47, 135, 144
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>fully normalized, 292</td>
<td></td>
</tr>
<tr>
<td>tesseral, 41</td>
<td></td>
</tr>
<tr>
<td>zonal, 41</td>
<td></td>
</tr>
<tr>
<td>spherical harmonic coefficient, 48, 49,</td>
<td></td>
</tr>
<tr>
<td>171, 186, 188, 242</td>
<td></td>
</tr>
<tr>
<td>fully normalized, 50</td>
<td></td>
</tr>
<tr>
<td>spherical harmonic expansion, 1, 42–44, 50,</td>
<td></td>
</tr>
<tr>
<td>53, 68, 69, 142, 153, 179, 259, 302</td>
<td></td>
</tr>
<tr>
<td>of ellipsoidal field, 68</td>
<td></td>
</tr>
<tr>
<td>rotationally symmetric, 44</td>
<td></td>
</tr>
<tr>
<td>spherical shell</td>
<td></td>
</tr>
<tr>
<td>attraction of, 6</td>
<td></td>
</tr>
<tr>
<td>potential of, 5</td>
<td></td>
</tr>
<tr>
<td>spherical-cap integration, 103</td>
<td></td>
</tr>
<tr>
<td>spheroid</td>
<td></td>
</tr>
<tr>
<td>Bruns, 69</td>
<td></td>
</tr>
<tr>
<td>Helmert, 69</td>
<td></td>
</tr>
<tr>
<td>spring balance, 198</td>
<td></td>
</tr>
<tr>
<td>stability (gravimeter), 210</td>
<td></td>
</tr>
<tr>
<td>stabilized platform, 204, 212</td>
<td></td>
</tr>
<tr>
<td>stationarity, 175</td>
<td></td>
</tr>
<tr>
<td>Sterneck, Robert von, 198</td>
<td></td>
</tr>
<tr>
<td>stochastic process, 171, 175</td>
<td></td>
</tr>
<tr>
<td>stochastic quantity, 171</td>
<td></td>
</tr>
<tr>
<td>Stokes equation, 81, 129–131, 139, 143,</td>
<td></td>
</tr>
<tr>
<td>151, 183, 219, 221</td>
<td></td>
</tr>
<tr>
<td>Stokes function</td>
<td></td>
</tr>
<tr>
<td>modified, 157</td>
<td></td>
</tr>
<tr>
<td>Stokes kernel, 129, 130, 132, 136, 147,</td>
<td></td>
</tr>
<tr>
<td>158</td>
<td></td>
</tr>
<tr>
<td>modified, 144, 145</td>
<td></td>
</tr>
<tr>
<td>Stokes theorem, 1, 27</td>
<td></td>
</tr>
<tr>
<td>Stokes, George Gabriel, 26, 81</td>
<td></td>
</tr>
<tr>
<td>Strang van Hees, Govert, 156, 157, 161</td>
<td></td>
</tr>
<tr>
<td>Sturm, Jacques, 291</td>
<td></td>
</tr>
<tr>
<td>Sturm-Liouville equation, 287</td>
<td></td>
</tr>
<tr>
<td>Sturm-Liouville problem, 291</td>
<td></td>
</tr>
<tr>
<td>submarine, 232</td>
<td></td>
</tr>
<tr>
<td>superconduction, 208</td>
<td></td>
</tr>
<tr>
<td>superspring, 204</td>
<td></td>
</tr>
<tr>
<td>Sverdrup (unit), 225</td>
<td></td>
</tr>
<tr>
<td>tangent plane, 60</td>
<td></td>
</tr>
<tr>
<td>tapering function, 161</td>
<td></td>
</tr>
<tr>
<td>tea break, 210</td>
<td></td>
</tr>
<tr>
<td>telephony, 127</td>
<td></td>
</tr>
<tr>
<td>telluroid, 117</td>
<td></td>
</tr>
<tr>
<td>telluroid mapping, 81</td>
<td></td>
</tr>
<tr>
<td>temperature, ocean, 218</td>
<td></td>
</tr>
<tr>
<td>terrain correction, 88, 89, 92, 123, 164,</td>
<td></td>
</tr>
<tr>
<td>165, 300</td>
<td></td>
</tr>
<tr>
<td>terrain effect, 12, 121, 167</td>
<td></td>
</tr>
<tr>
<td>thermal expansion, of sea water, 220</td>
<td></td>
</tr>
<tr>
<td>thermostat (gravimeter), 204</td>
<td></td>
</tr>
<tr>
<td>thruster, 241</td>
<td></td>
</tr>
<tr>
<td>on-board, 245</td>
<td></td>
</tr>
<tr>
<td>tidal field, 216</td>
<td></td>
</tr>
<tr>
<td>tide, 10, 217, 233, 235</td>
<td></td>
</tr>
<tr>
<td>diurnal, 243, 256</td>
<td></td>
</tr>
<tr>
<td>ocean, 234</td>
<td></td>
</tr>
<tr>
<td>permanent part, 259</td>
<td></td>
</tr>
<tr>
<td>semi-diurnal, 256</td>
<td></td>
</tr>
<tr>
<td>semidiurnal, 243</td>
<td></td>
</tr>
<tr>
<td>solid-Earth, 234, 263</td>
<td></td>
</tr>
<tr>
<td>theoretical, 253, 254, 256, 257</td>
<td></td>
</tr>
<tr>
<td>equation, 259</td>
<td></td>
</tr>
<tr>
<td>tide gauge, 218, 220, 223</td>
<td></td>
</tr>
<tr>
<td>tide-free geoid, 259</td>
<td></td>
</tr>
<tr>
<td>Tikhonov, Andrey Nikolayevich, 238</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td></td>
</tr>
<tr>
<td>co-ordinate, 126</td>
<td></td>
</tr>
<tr>
<td>proper, 126</td>
<td></td>
</tr>
<tr>
<td>Solar</td>
<td></td>
</tr>
</tbody>
</table>
Index

mean, 244
Toeplitz circulant matrix, 193
Toeplitz, Otto, 193
tomography
  seismic, 103
TOPEX/Poseidon (satellite), 218, 232, 243
topography, 12
Torge, Wolfgang, 264
torsion balance, 3, 215
total mass (Earth), 66
Toulouse, France, 261
transformation surface, 219
translation, 238
trench, ocean, 198
trend, 239
triangle inequality, 10
Trieste, Italy, 110
tropospheric propagation delay
  dry, 234
  wet, 234
Tscherning, Carl Christian, 162, 191
Tscherning-Rapp formula, 191
Turku harbour, 238
Turku, Finland, 238
Tuvalu, 227
Tytyri limestone mine (Lohja, Finland), 258
uncertainty, 169, 175, 181
  a priori, 238
  of observation, 184
unknown, 238, 239
upward continuation, 188
Vaasa, Finland, 208
variance, 171
  a priori, 238
  matrix, 289
  of prediction, 181, 184
vector space, 147, 285, 286, 288
  n-dimensional, 287
  abstract, 285, 287, 291
  linear, 147, 288
velocity, 272
Vening Meinesz, Felix Andries, 131, 198
Vening-Meinesz equations, 131, 148
Verbaandert-Melchior pendulum, 258
very long wire pendulum, 198
viscosity, 229
viscous relaxation mode, 230
water vapour radiometer, 234
wave number, 166
wavelength
  of spherical harmonics, 41
weightlessness, 214
Wenzel, Hans-Georg, 260
wind pile-up, 217
wire-frame model, 238
work, 271
work integral, 272
World Geodetic System 1984 (WGS84), 49, 66
zenith tube, 258
zero geoid, 259
zero-length spring, 201, 202