Machine Learning with Signal Processing

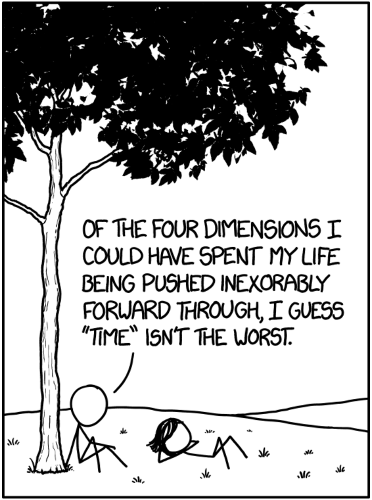
Part I: Signal Processing Tooling

Arno Solin
Assistant Professor in Machine Learning
Department of Computer Science
Aalto University

ICML 2020 Tutorial

@arnosolin

arno.solin.fi


OF THE FOUR DIMENSIONS I COULD HAVE SPENT MY LIFE BEING PUSHED INEXORABLY FORWARD THROUGH, I GUESS "TIME" ISN'T THE WORST.

Scope of this tutorial

This tutorial is about:

- (Mostly) temporal models
- Various tools in signal processing
- Pointers to application areas
- Links to aspects in statistical ML

This tutorial is *not* about:

- Audio processing
- Image processing
- Encyclopedic overview

Goals

- Teach basic principles of direct links between signal processing and machine learning
- Provide an intuitive hands-on understanding of what stochastic differential equations are all about
- Show how these methods have real benefits in speeding up learning, improving inference, and model building

Disclaimer

- I do not have time to discuss many important and relevant works
- If you think I should have included some of those, please send me email and I will try to include it the next time
- The content of the tutorial is based on my own biased opinion (and expertise)
- Many examples are based on my own work

Structure

Part I

Tools and discrete-time models

Part II

SDEs (continuous-time models)

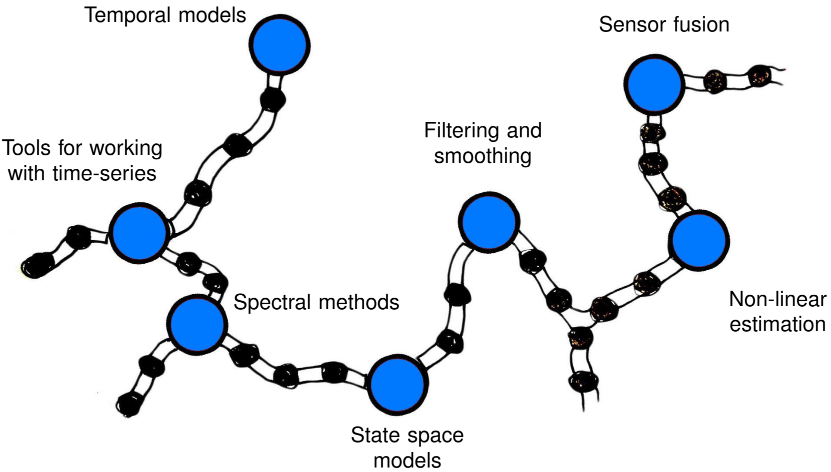
Part III

Gaussian processes

Part IV

Application examples

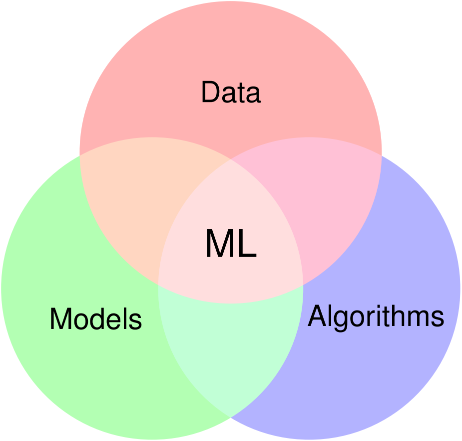
Outline


Temporal models
Sensor fusion
Tools for working
with time-series
Filtering and
smoothing
Spectral methods
Non-linear
estimation
State space
models

Motivation: Temporal models

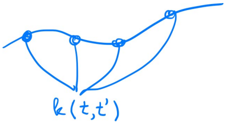
- **One-dimensional problems**
  (the data has a natural ordering)
- **Spatio-temporal models**
  (something developing over time)
- **Long / unbounded data**
  (sensor data streams, daily observations, *etc.*)

Machine learning

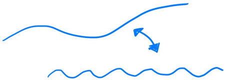


Tools for dealing with time-series

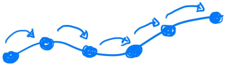
- Moment representation
  Considering the statistical properties of the input data jointly over time



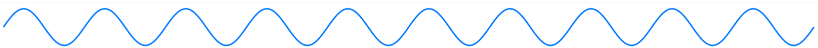
- Spectral (Fourier) representation
  Analyzing the frequency-space representation of the problem/data



- State space (path) representation
  Description of sample behaviour as a dynamic system over time



Spectral (Fourier) representation



- Fourier transform $\mathcal{F}[\cdot]$:

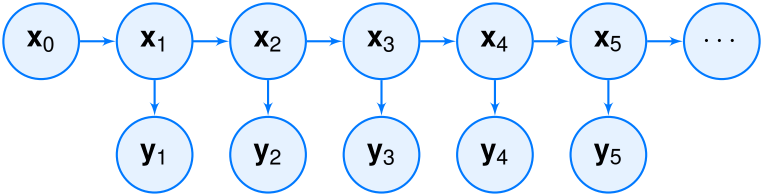
$$\tilde{f}(\omega) = \int f(\mathbf{x}) \, \exp(-\mathrm{i} \, \omega^{\mathsf{T}} \mathbf{x}) \, \mathrm{d}\mathbf{x}$$

- Analyzing properties of 'systems' (input–output mappings) by transfer functions:

$$H(s) = \frac{Y(s)}{X(s)} = \frac{\mathcal{L}[y(t)](s)}{\mathcal{L}[x(t)](s)},$$

where $\mathcal{L}[\cdot]$ is the Laplace transform

Discrete-time state space models

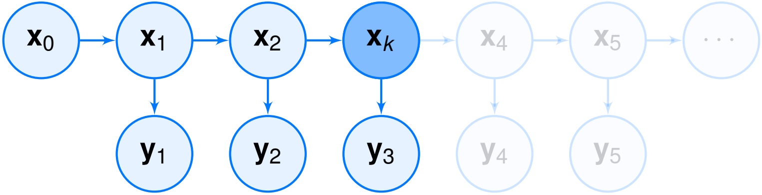


- A canonical state space model:

$$\begin{aligned} \text{Dynamics:} \quad & \mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{q}_k), \quad && \mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_k), \\ \text{Measurement:} \quad & \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k), \quad && \mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_k) \end{aligned}$$

- The key to efficiency is the directed graph:
  The Markov property.

Kalman filtering and smoothing

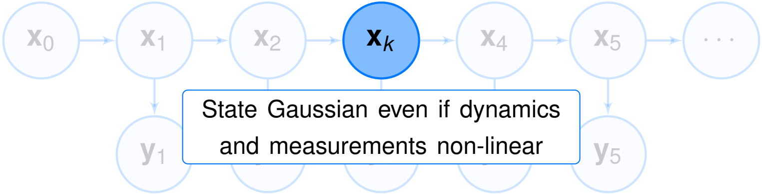


- Closed-form solution to linear-Gaussian filtering problems

$$\mathbf{x}_k = \mathbf{A}\,\mathbf{x}_{k-1} + \mathbf{q}_k, \qquad \mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_k),$$
$$\mathbf{y}_k = \mathbf{H}\,\mathbf{x}_k + \mathbf{r}_k, \qquad \mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_k)$$

- Filtering solution: $p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_{k|k}, \mathbf{P}_{k|k})$
- Smoothing solution: $p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) = \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_{k|T}, \mathbf{P}_{k|T})$

Non-linear filtering



- Typically $\mathbf{x}_k$ is assumed Gaussian:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{q}_k), \qquad \mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_k),$$
$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k), \qquad \mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_k)$$

- Filtering solution: $p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \simeq \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_{k|k}, \mathbf{P}_{k|k})$
- Smoothing solution: $p(\mathbf{x}_k \mid \mathbf{y}_{1:T}) \simeq \mathrm{N}(\mathbf{x}_k \mid \mathbf{m}_{k|T}, \mathbf{P}_{k|T})$

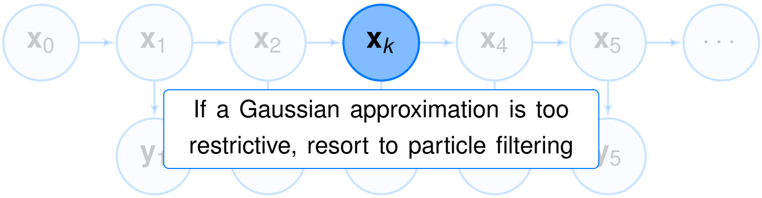
Sensor fusion by non-linear Kalman filtering

- Dynamics $\mathbf{f}(\cdot, \cdot)$ governed by high-frequency accelerometer and gyroscope sensor samples (Newton's laws)
- Observations $\mathbf{y}_k$ are camera frames
- Online learning problem (sensor noises and biases)

Real-time visual-inertial motion tracking

Probabilistic inertial-visual odometry
for occlusion-robust navigation (`https://youtu.be/_ywmtVzxURk`)

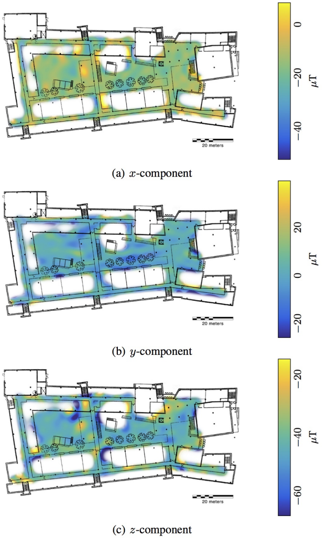
Non-linear filtering by sequential Monte Carlo



- Sequential Monte Carlo characterizes the state distribution with a swarm of 'particles' (particle filtering)
- Can deal with multi-modality, severe non-linearities, *etc.*
- All the usual MC caveats apply

Sensor fusion by sequential Monte Carlo

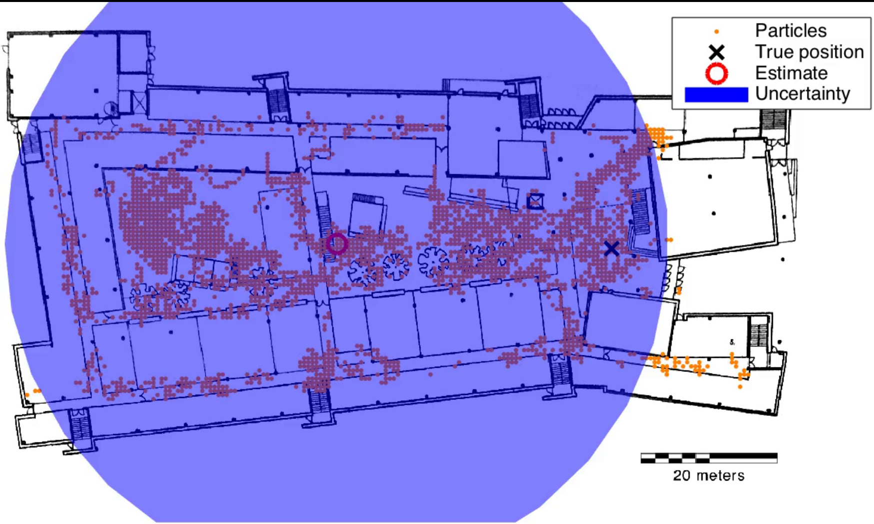
- Absolute position tracking of a smartphone by using the compass sensor (magnetometer)
- Match observations to a map of local anomalies of the magnetic field inside a building
- The ‘anomaly track’ becomes unique when the phone has moved a long-enough distance



(a) x-component

(b) y-component

(c) z-component



Terrain matching in the magnetic landscape
by sequential Monte Carlo (https://youtu.be/UuUo9QOOT1Q)

Up next

2

- Going from discrete-time to continuous-time
- A gentle introduction to stochastic differential equations (SDEs)

Bibliography

These references are sources for finding a more detailed overview on the topics of this part :

- T. Glad and L. Ljung (2000). *Control Theory: Multivariable and Nonlinear Methods*. Taylor & Francis, New York.
- G. Wahba (1990). *Spline Models for Observational Data*. Siam.
- S. Särkkä (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press. Cambridge, UK.
- T. B. Schön *et al.* (2011). The particle filter in practice. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, UK.