Fast Variational Bayesian Linear State-Space Model

Aalto University
School of Science

Jaakko Luttinen, `jaakko.luttinen@aalto.fi`
Department of Information and Computer Science, Aalto University

Summary

- *Model:* Linear state-space model
  - used for multivariate dynamical systems
- *Problem:* Variational Bayesian (VB) estimation is slow because
  - the variables are strongly coupled
  - the variables are updated one at a time
  - the iteration zigzags and proceeds slowly
- *Solution:* Jointly optimize several variables based on how they are coupled
  - that is, optimize the rotation of the latent subspace
  - in general, known as parameter expansion
- *Effect:* 100–10000 times faster convergence

Model

- A sequence of high-dimensional observations $(\mathbf{y}_1,\ldots,\mathbf{y}_N)$ is assumed to be generated from latent low-dimensional states $(\mathbf{x}_1,\ldots,\mathbf{x}_N)$:

$$\mathbf{y}_n = \mathbf{C}\mathbf{x}_n + \text{noise}\,.$$

  where $\mathbf{C}$ is the loading matrix
- The latent states follow a first-order Markov process:

$$\mathbf{x}_n = \mathbf{A}\mathbf{x}_{n-1} + \text{noise}\,.$$

  where $\mathbf{A}$ is the state dynamics matrix.
- Variables $\mathbf{x}_n$, $\mathbf{A}$ and $\mathbf{C}$ are unknown and estimated from the data.

Speed-up rotation / Parameter expansion

- Motivate the parameter expansion by the rotational ambiguity of the latent sub-space.
- The states can be rotated by compensating it in the loadings:

$$\mathbf{y}_n = \mathbf{C}\mathbf{x}_n = \mathbf{C}\mathbf{R}^{-1}\mathbf{R}\mathbf{x}_n\,,$$

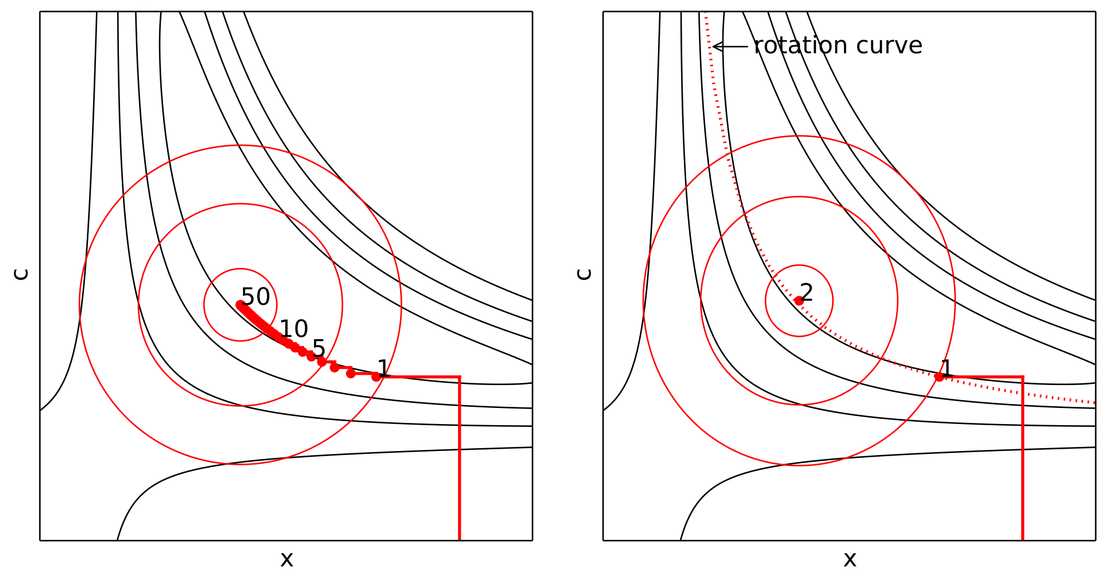
  thus rotate as $\mathbf{C} \to \mathbf{C}\mathbf{R}^{-1}$ and $\mathbf{x}_n \to \mathbf{R}\mathbf{x}_n$.
- Keep the dynamics of the latent states unaffected:

$$\mathbf{R}\mathbf{x}_n = \mathbf{R}\mathbf{A}\mathbf{R}^{-1}\mathbf{R}\mathbf{x}_{n-1}\,,$$

  thus rotate as $\mathbf{A} \to \mathbf{R}\mathbf{A}\mathbf{R}^{-1}$.
- Parameterize the VB posterior by $\mathbf{R}$ and maximize the VB lower bound with respect to $\mathbf{R}$.

Simple illustration

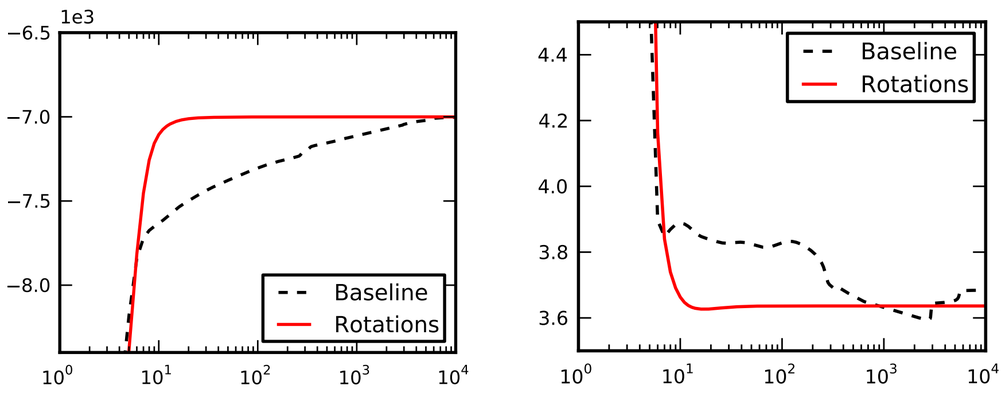
- VB for a simple 1-dimensional model $y = cx + \text{noise}$
- Rotate as $x \to Rx$ and $c \to c/R$
- Compare the VB iterations with and without rotation:



(a) Standard VB (b) VB using rotations

Artificial experiment

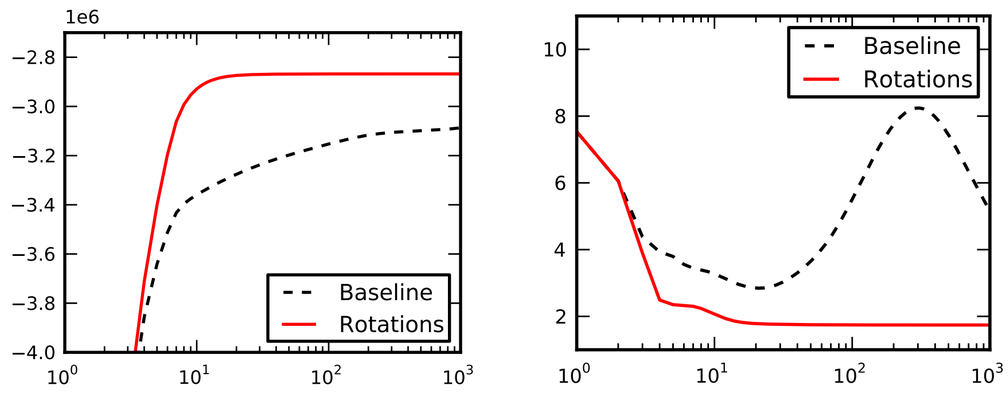
- 400 observations with 30 dimensions
- 8-dimensional latent space
- Performance as a function of VB iterations (log-scale):



(a) VB lower bound (b) Test RMSE

Weather data experiment

- 89202 observations with 66 dimensions
- 10-dimensional latent space
- Performance as a function of VB iterations (log-scale):



(a) VB lower bound (b) Test RMSE

The method was implemented as a part of a variational Bayesian Python package, BayesPy, available at GitHub and PyPI.
The scripts and data for reproducing the results are available at `http://users.ics.aalto.fi/jluttine/ecml2013`.