TSA Tutorial 4

17.5.2021, Manuel Brenner

Organizational Question

Is someone still missing their correction of exercise sheet 2? Moodle shows me there is one still missing, but I can’t find it.

Log Likelihood

$$(7.34)\ \log \mathrm{L}(\{\alpha_i\},\sigma) = -\frac{T-p}{2}\log(2\pi) - \frac{T-p}{2}\log(\sigma^2) - \frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\sigma^{-2},$$

ML estimator $\hat{\sigma}^2 = \sum_{t=p+1}^{T} \hat{\varepsilon}_t^2 / (T-p)$;

Latent Variable Models

“A **latent variable model** is a statistical model that relates a set of observable variables (so-called *manifest variables*) to a set of latent variables. “

allows us to look beyond the “data surface”

Is what we observe generated by some more complex/simple underlying process that we cannot observe directly?

Examples

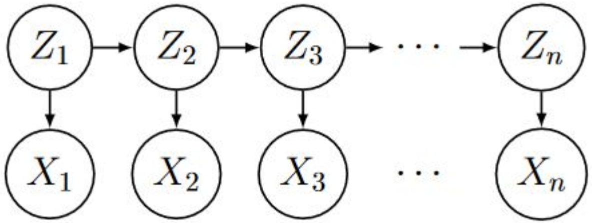
Example 1: Covid Infections numbers are observable, but underlying dynamics are generally unknown/depend on many factors

Example 2: Rise in global temperature vs. underlying climate models

Example 3: Neuronal firing patterns

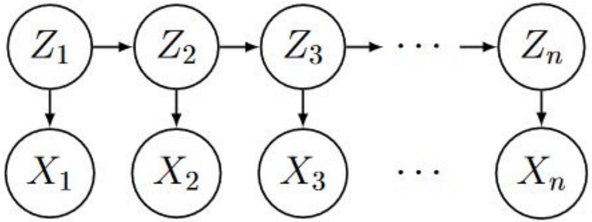
Factorisation of dependencies

State Space Models with first order Markov property



$$p(x_{1:n}, z_{1:n}) = p(z_1)p(x_1|z_1)\prod_{j=2}^{n} p(z_j|z_{j-1})p(x_j|z_j).$$

Observation Models



How do we obtain an observable that we can actually measure from an underlying “true” state?

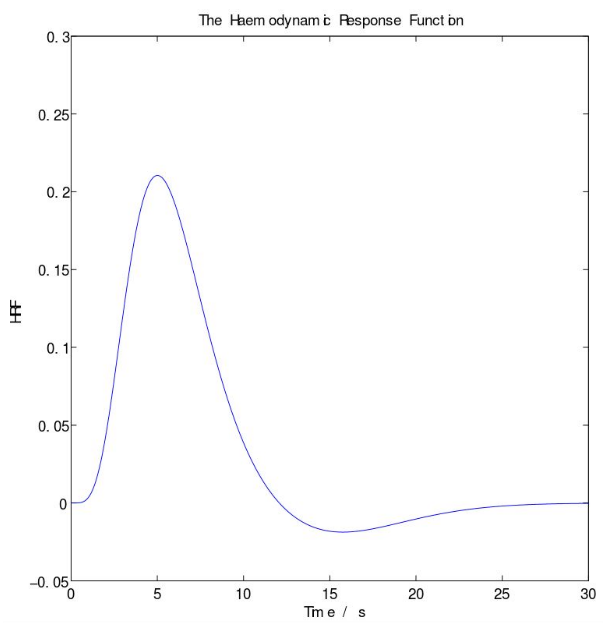
- positive covid tests relate to underlying infections through
- we measure spike trains that relate in some way to synaptic dynamics/fMRI measurements

Example: Hemodynamic Response Function

In fMRI measurements, we only measure blood oxygenation level, not activity, but assume activity is the underlying cause of changes in oxygen

→ Consider a time window tau-t

$$\mathbf{x}_t = \mathbf{B}(hrf * \mathbf{z}_{\tau:t}) + \mathbf{J}\mathbf{r}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(0, \boldsymbol{\Gamma})$$



Latent Factor Analysis

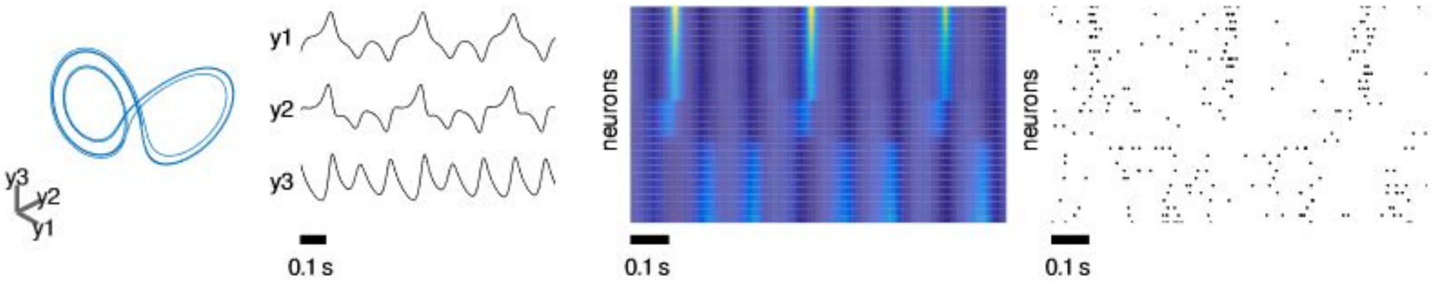
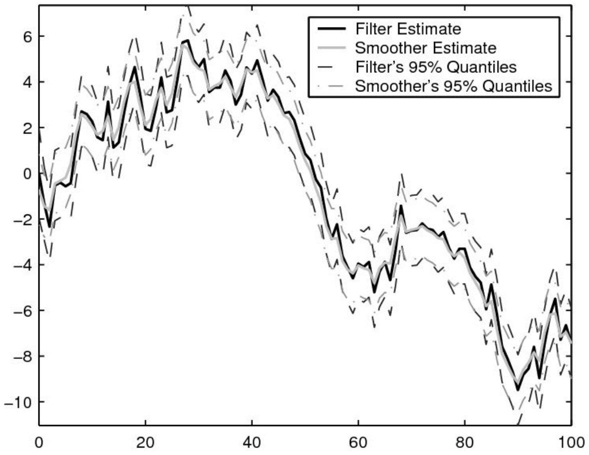


Figure 3: **Overview of Lorenz task.** An example trial illustrating the evolution of the Lorenz system in its 3-dimensional state space (far left) and its dynamic variables as a function of time (left middle). Firing rates for the 30 simulated neurons are computed by a linear readout of the latent variables followed by an exponential nonlinearity (right middle; with neurons sorted according to their weighting for the first Lorenz dimension). Spike times for the neurons are shown at the right.

Kalman Filter



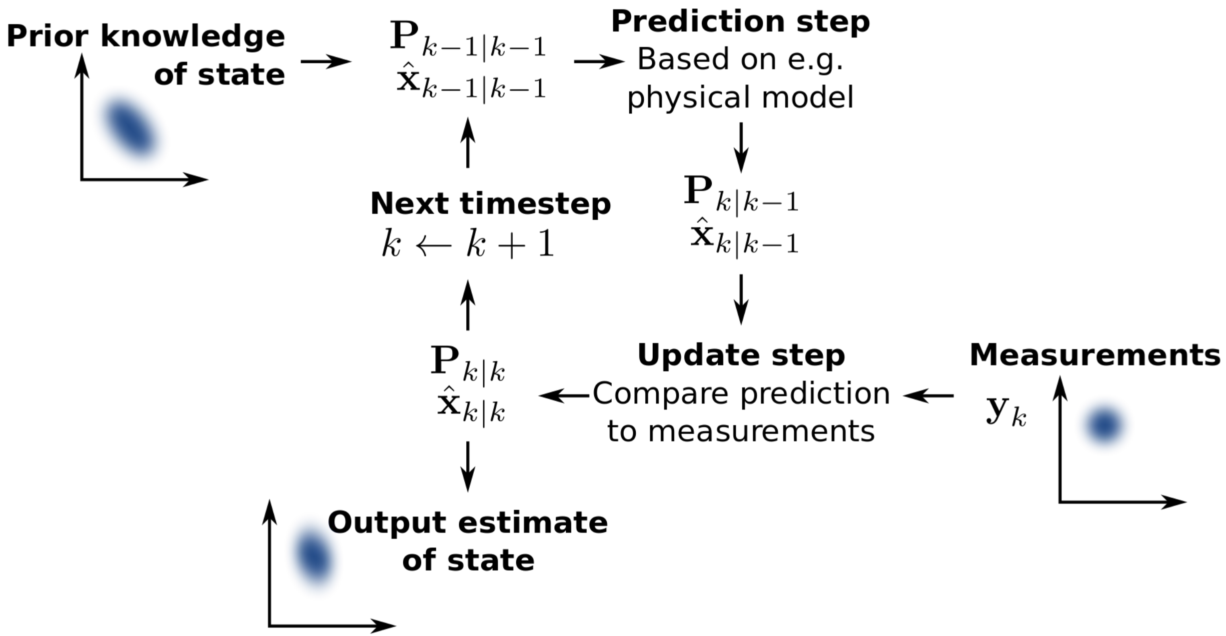
Assume errors are Gaussian, dynamics are linear

→ try to obtain optimal state estimates

Kalman filter is best possible linear estimator in MSE sense

Originally in a control theory context (helped land us on the moon)

Kalman Filter: estimating latent states



Kalman Filter+Smoother

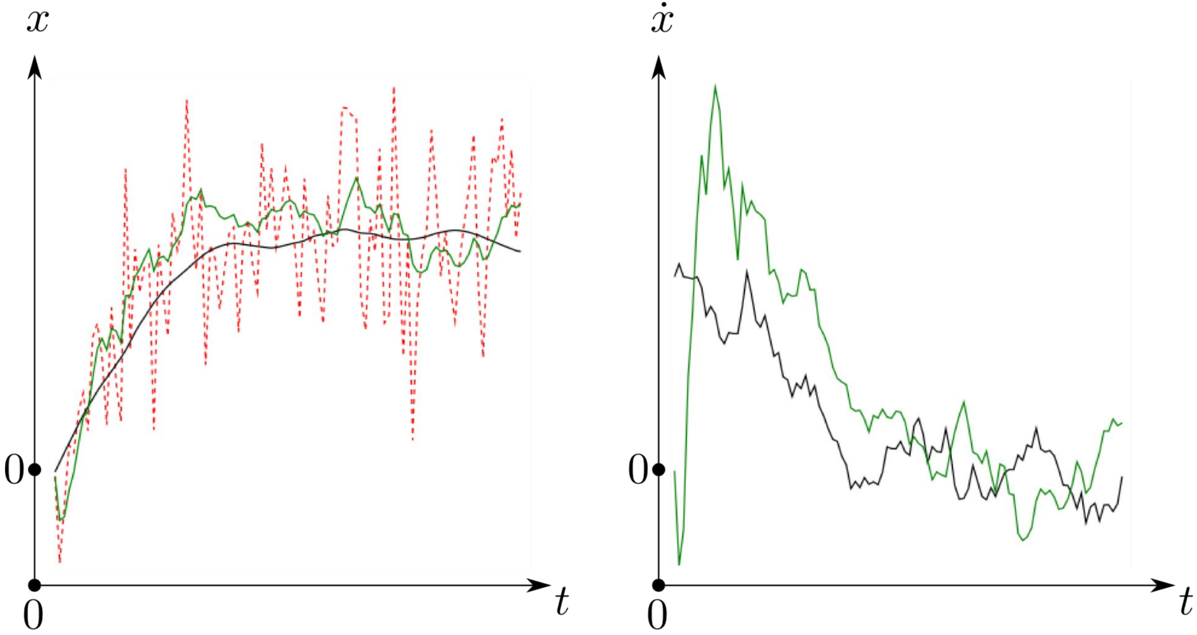
Filtering: looking into the past

How do I use my past observations to obtain the best estimate of the current state? $p(z_j|x_{1:j})$

Smoothing: looking back from the future

How do I use future observations to find the best estimate of the current state?

Kalman Filter+Smoother



Latent Variables+Intractability

Conditional Dependencies follow Bayes' law

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \frac{p(\boldsymbol{X} \mid \boldsymbol{Y})p(\boldsymbol{Y})}{p(\boldsymbol{X})} = \frac{p(\boldsymbol{X}, \boldsymbol{Y})}{\int p(\boldsymbol{X}, \boldsymbol{Y}) d\boldsymbol{Y}}.$$

Integrating over all latent states is generally intractable:

$$p(\boldsymbol{X}) = \int p(\boldsymbol{X}, \boldsymbol{Y}) d\boldsymbol{Y},$$

Intractability in the EM

In the E-step, we try to create an expectation over the intractable density over the latent states

$\longrightarrow$ we need some way to approximate the posterior distribution

$\longrightarrow$ approximate posterior $q(z)$

In the M-step, we use that estimate to obtain the maximum likelihood parameters of the model

Creating an expectation by sampling

We usually try to build unbiased estimators of quantities of interest

$$\mu = \mathbb{E}[f(\mathrm{X})] = \int p(x) f(x) dx. \qquad \hat{\mu} = \frac{1}{N} \sum_{\ell=1}^{N} f(x^{(\ell)}), \quad x^{(\ell)} \sim p(x),$$

We can also f.e. build an estimate of the ELBO by sampling z's

$$\widehat{\mathrm{ELBO}}_{\boldsymbol{X}}(\phi, \boldsymbol{\theta}) = \frac{1}{L} \sum_{\ell=1}^{L} \left(\log p_{\boldsymbol{\theta}}(\boldsymbol{X}, \boldsymbol{Z}^{(\ell)}) - \log q_{\phi}(\boldsymbol{Z}^{(\ell)} \mid \boldsymbol{X}) \right),$$

Variational Inference and Autoencoders