DS-GA 3001.001 Special Topics in Data Science: Modeling Time Series
Homework 2

Due date: Oct 11, by 6pm

Problem 1. LDS model, 10p
Consider a special case of LDS with $\mathbf{C} = \mathbf{I}$ and $\mathbf{R} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix. Show that in the limit where there is no observation noise the best estimate for latent $\mathbf{z}_i$ is to simply use the observation $\mathbf{x}_i$: formally, in the limit when $\sigma^2 \to 0$ the posterior for $\mathbf{z}_i$ has mean $\mathbf{x}_i$ and vanishing variance.

Solution:
The prediction for the latent state $\mathbf{z}_i$ can be determined using the Kalman filter equations, with the particular setting $\mathbf{C} = \mathbf{I}$:

$$\begin{aligned} \mu_{i|i} &= \mu_{i|i-1} + \mathbf{K}_i \left(x_i - \mathbf{C}\mu_{i|i-1}\right) = \mu_{i|i-1} + \mathbf{K}_i \left(x_i - \mu_{i|i-1}\right) \\ \Sigma_{i|i} &= \Sigma_{i|i-1} - \mathbf{K}_i\mathbf{C}\Sigma_{i|i-1} = \Sigma_{i|i-1} - \mathbf{K}_i\Sigma_{i|i-1} \end{aligned}$$

where the Kalman gain is:

$$\begin{aligned} \mathbf{K}_i &= \Sigma_{i|i-1}\mathbf{C}^{\mathrm{t}} \left(\mathbf{C}\Sigma_{i|i-1}\mathbf{C}^{\mathrm{t}} + R\right)^{-1} \\ &= \Sigma_{i|i-1} \left(\Sigma_{i|i-1} + \sigma^2\mathbf{I}\right)^{-1} \end{aligned}$$

In the limit $\sigma^2 \to 0$, the gain $\mathbf{K}_i \to \mathbf{I}$, which translates into limits for the estimates

$$\begin{aligned} \lim_{\sigma^2\to 0}\mu_{i|i} &= \mu_{i|i-1} + \mathbf{I}\left(\mathbf{x}_i - \mu_{i|i-1}\right) = \mathbf{x}_i && (1) \\ \lim_{\sigma^2\to 0}\Sigma_{i|i} &= \Sigma_{i|i-1} - \mathbf{I}\Sigma_{i|i-1} = \mathbf{0} && (2) \end{aligned}$$

Hence, in the limit of very low observation noise it is best to listen to the observations and completely ignore temporal dependencies.

Problem 2. LDS prediction, 10p
Given the standard parametrization of the LDS model, and the Kalman filtering estimates $\mu_{i|i}$ and $\Sigma_{i|i}$, obtained for a dataset $\mathbf{x}_{1:t}$ write down the expressions for predicting the following 2 observations in the sequence $\mathbf{x}_{t+1}$ and $\mathbf{x}_{t+2}$.

Solution:
Starting from the posterior marginal for $\mathbf{z}_t$, which is multivariate gaussian with mean $\mu_{t|t}$ and covariance $\Sigma_{t|t}$, the marginal predictive distribution for $\mathbf{z}_{t+1}$ and $\mathbf{z}_{t+2}$ is multivariate normal with parameters (using the properties of a linear gaussian model):

$$\begin{aligned} \mu_{t+1|t} &= \mathbf{A}\mu_{t|t} && (3) \\ \Sigma_{t+1|t} &= \mathbf{A}\Sigma_{t|t}\mathbf{A}^{\mathrm{t}} + \mathbf{Q} && (4) \\ \mu_{t+2|t} &= \mathbf{A}\mu_{t+1|t} = \mathbf{A}^2\mu_{t|t} && (5) \\ \Sigma_{t+2|t} &= \mathbf{A}\Sigma_{t+1|t}\mathbf{A}^{\mathrm{t}} + \mathbf{Q} = \mathbf{A}(\mathbf{A}\Sigma_{t|t}\mathbf{A}^{\mathrm{t}} + \mathbf{Q})\mathbf{A}^{\mathrm{t}} + \mathbf{Q} && (6) \\ &= \mathbf{A}^2\Sigma_{t|t}\mathbf{A}^{2\mathrm{t}} + \mathbf{A}\mathbf{Q}\mathbf{A}^{\mathrm{t}} + \mathbf{Q} && (7) \end{aligned}$$

Marginalizing out the unknown $\mathbf{z}_{t+i}$ yields the prediction for $\mathbf{x}_{t+i}$:

$$\begin{aligned} \mu_{x,t+i} &= \mathbf{C}\mu_{t+i|t} = \mathbf{C}\mathbf{A}^i\mu_{t|t} && (8) \\ \Sigma_{x,t+i} &= \mathbf{C}\Sigma_{t+i|t}\mathbf{C}^{\mathrm{t}} + \mathbf{R} && (9) \end{aligned}$$

Problem 3. LDS inference with missing observations, 10p
Consider a variation of the original LDS graphical model with one single missing value $\mathbf{x}_j$. Everything else is as in the original; the only difference is that the graphical model loses the downward observation

arrow and the corresponding $\mathbf{x}_j$.) How do the Kalman filtering/smoothing updates change?

Solution:
The Kalman filter updates change for index j: $\mu_{j|j} = \mu_{j|j-1}$ and $\Sigma_{j|j} = \Sigma_{j|j-1}$. The smoother updates remain unchanged.

Problem 4. LDS filtering, 10p
Given the model parameters: $A = \begin{bmatrix} 0.65 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$, $C = \begin{bmatrix} 1.1 & 0.2 \\ 0.5 & 0.95 \end{bmatrix}$, $Q = 0.1\mathbf{I}_2$, $R = 0.01\mathbf{I}_2$ with initial condition parameters $\mu_0 = [0\ 0]^{\mathrm{t}}$, $\Sigma_0 = 0.5\mathbf{I}$. Generate 25 samples from the model and plot them. Use these observation for inference (filtering and smoothing). How much does the Kalman gain $\mathbf{K}_i$, and $\mathbf{F}_i$ vary across timepoints? Try playing around with the parameters. Does the result change? Discuss.
Note: you can use code from the lab as starting point, if you want. Not need to submit the code, only your figures.
Solution:
In most scenarios the gains do not vary much, that's why in some practical applications you may see people using a constant. Variation may show up more prominently in the first few steps of filtering, if the initial conditions are very noisy.

Problem 5 Particle filtering, 10p
Consider the usual LDS model, but where inference is done using particle filtering instead of the traditional Kalman filter. Write down pseudocode for the particle updates. Given the generated samples, $\{\mathbf{z}_i^{(k)}\}_{k=1:K, i=1:t}$, how would you go about computing the quantities $\mu_{i|i}$, $\Sigma_{i|i}$ and $\mathbb{E}[\mathbf{z}_i\mathbf{z}_{i+1}^t]$?
Hint: Use the general form from the lecture, and plug in the expressions for the different probabilities of the LDS model. The mean and variance can be written as expectations and approximated accordingly.
Solution:

Algorithm 1 Particle filtering pseudocode

---

function PARTICLEFILTERING($\mathbf{x}_{1:t}$, θ, K) initialization of particles:
  $\mathbf{z}_1^{1:K} \leftarrow \mathrm{rndmvn}(\mu_0, \Sigma_0)$
  for $i = 1 : t$ **do**

Step 1: compute importance weights

$$w_i^{(k)} \leftarrow \frac{\mathrm{P}\left(\mathbf{x}_i|\mathbf{z}_i^{(k)}\right)}{\sum_m \mathrm{P}\left(\mathbf{x}_i|\mathbf{z}_i^{(m)}\right)} \text{ where } \mathrm{P}\left(\mathbf{x}_i|\mathbf{z}_i^{(k)}\right) = \exp\left(-0.5\left(\mathbf{x}_i - \mu\right)\mathbf{R}^{-1}\left(\mathbf{x}_i - \mu\right)\right)^T, \text{ and } \mu = \mathbf{C}\mathbf{z}_i^{(k)}$$

Step 2: propagate particles one time step forward, by sampling from mixture

$c_{i,k} \leftarrow \mathrm{multinomial}(\mathbf{w})$
$\mathbf{z}_{i+1}^k \leftarrow \mathbf{A}\mathbf{z}_i^{(c_{i,k})} + \mathrm{rndmvn}(0, \mathbf{Q})$

N.B. There is a bit of an inconsistency in the notation here relative to the original LDS derivation in that μ_0 and Σ_0 are takes as the prior for $\mathbf{z}_1$ rather than $\mathbf{z}_0$. In the original notation the initial particles would come from $\mathrm{rndmvn}(\mathbf{A}\mu_0, \mathbf{A}\Sigma_0\mathbf{A}^T + \mathbf{Q})$.

The marginal $\mathrm{P}(\mathbf{z}_i|\mathbf{x}_{1:i})$ is a mixture with components parametrized by $w_i^{(k)}$. Any expectation of a function $f$under this distribution can be approximated via importance sampling as

$$\mathbb{E}[f(\mathbf{z}_i)] = \sum_k w_i^{(k)} f\left(\mathbf{z}_i^{(k)}\right).$$

The different moments can be computed by different setting of f: $f(\mathbf{z}) = \mathbf{z}$ for first moment, $f(\mathbf{z}) = \mathbf{z}\mathbf{z}^t$

for the second moment, then applying the formula to get the variance. Practically, this means:

$$\mu_{i|i} = \sum_k w_i^{(k)} \mathbf{z}_i^{(k)}$$

$$\Sigma_{i|i} = \sum_k w_i^{(k)} \mathbf{z}_i^{(k)} \mathbf{z}_i^{(k)T} - \mu_{i|i}^2$$

Finally, for the second moment, we need to use parent-child sample pairs, and weight by the importance weight of the children to take into account the evidence from $\mathbf{x}_{i+1}$:

$$\mathbb{E}[\mathbf{z}_i \mathbf{z}_{i+1}^t] = \sum_k w_{i+1}^{(k)} \mathbf{z}_i^{(c_i,k)} \mathbf{z}_{i+1}^{(k)\ T}$$