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

Due date: March 2, by midnight

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.

Problem 2. LDS prediction, 20p
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 $\mathbf{x}_{t+2}$.

Problem 3. LDS inference coding, 20p
Given the model parameters:
$A = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}$, $C = \begin{bmatrix} 1.1 & 0.2 \\ 0.1 & 0.95 \end{bmatrix}$, $Q = \sigma_Q^2\mathbf{I}$, $R = \sigma_R^2\mathbf{I}$ with default parameter values $\sigma_q^2 = 0.4$, $\sigma_r = 0.001$ and initial condition parameters $\mu_0 = [00]^t$, $\Sigma_0 = \mathbf{I}$. Use the lab4 code as a starting point for the following:

- draw a sequence of length $t = 10$ observations from this model.
- using this artificial data and the true model parameters, run kalman filtering and smoothing; display the data, observations, and corresponding marginal posteriors (filtering, and smoothing) using the conventions in Bishop figure 13.22: true latent trajectory in blue, observations in green, red cross and ellipse for mean and cov of Kalman filter, same in yellow for the kalman smoother.
- Manipulate parameters $\sigma_{q,r}^2$ such that a) the latent dynamics are close to deterministic and b) the observation noise is very small. Re-plot inference figure for each parameter configuration and comment on results.
- What happens to the forward vs. backward estimates when the latent noise term is large or very small?

Problem 4. LDS dimensionality reduction coding, 10p
Given the model parameters:
$A = \begin{bmatrix} 3 & 0.5 \\ 0.8 & 1.7 \end{bmatrix}$, $C = \begin{bmatrix} 1.1 & 0.2 \\ 0.5 & 0.95 \\ 1 & 1 \end{bmatrix}$, $Q = \mathbf{I}_2$, $R = 0.001^2\mathbf{I}_3$ with initial condition parameters $\mu_0 = [00]^t$, $\Sigma_0 = \mathbf{I}$. Visualize $t = 100$ data points sampled from the model then for the first 10 elements in the sequence show the predicted distributions for $\mathbf{x}_{i+1}|\mathbf{x}_{1:i}$.

Problem 5. EM, 20p
For the same LDS model used in Problem 3, draw t samples from model build your own EM implementation (can rely on code from lab4) to estimate the model parameters. Do your parameter estimates depend strongly on the initial value of the parameters? Comment why (not). How do the parameter estimates improve with more data ($t = 100$ vs. $t = 500$)? For a fixed amount of data $t = 100$, how accurate are the parameter reconstructions as you vary the observation noise from $\sigma_r = 0.001$ to $\sigma_r = 0.1$?

Problem 6. Particle filtering, 20p
For the same LDS model used in Problem 3, implement a full step of the particle filtering algorithm: use the closed form solution for $\mathrm{P}z_i|\mathbf{x}_{1:i-1}$ to draw an ensemble of N particles, combine with observation $\mathbf{x}_i$ and build a sampling-based approximation for the predicting distribution $\mathrm{P}z_{i+1}|\mathbf{x}_{1:i}$, compare the analytical posterior mean (Kalman filtering) to the sampling based estimates, for $N = 10, 100, 1000, 10000$.