

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

Yves Greatti - yg390

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 $\sigma^2 \rightarrow 0$ the posterior for \mathbf{z}_i has mean \mathbf{x}_i and vanishing variance.

Given the LDS model specifications for the posterior \mathbf{z}_i :

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

where $\mathbf{K}_i = \Sigma_{i|i-1} \mathbf{C}^T (\mathbf{C} \Sigma_{i|i-1} \mathbf{C}^T + \mathbf{R})^{-1}$. If $\mathbf{C} = \mathbf{I}$ and $\mathbf{R} = \sigma^2 \mathbf{I}$, then the Kalman gain becomes $\mathbf{K}_i = \frac{\Sigma_{i|i-1}}{\Sigma_{i|i-1} + \sigma^2 \mathbf{I}}$, taking the limit $\sigma^2 \rightarrow 0$ then $\mathbf{K}_i \rightarrow \mathbf{I}$, thus in the limit the posterior \mathbf{z}_i has for parameters:

$$\begin{aligned}\mu_{i|i} &= \mu_{i|i-1} + (\mathbf{x}_i - \mu_{i|i-1}) = \mathbf{x}_i \\ \Sigma_{i|i} &= \Sigma_{i|i-1} - \Sigma_{i|i-1} = \mathbf{0}\end{aligned}$$

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} .

Given the prior with parameters $\mu_{i|i}$ and $\Sigma_{i|i}$, we compute the prior for $\mathbf{z}_{t+1} \approx \mathcal{N}(\mu_{i+1|i}, \Sigma_{i+1|i})$:

$$\begin{aligned}\mu_{i+1|i} &= \mathbf{A}\mu_{i|i} \\ \Sigma_{i+1|i} &= \mathbf{A}\Sigma_{i|i}\mathbf{A}^t + \mathbf{Q}\end{aligned}$$

We obtain the prediction $\mathbf{x}_{t+1} \approx \mathcal{N}(\mathbf{C}\mu_{i+1|i}, \mathbf{C}\Sigma_{i+1|i}\mathbf{C}^t + \mathbf{R})$, we then incorporate the evidence to obtain the posterior parameters:

$$\begin{aligned}\mu_{i+1|i+1} &= \mu_{i+1|i} + \mathbf{K}_{i+1}(\mathbf{x}_{t+1} - \mathbf{C}\mu_{i+1|i}) \\ \Sigma_{i+1|i+1} &= \Sigma_{i+1|i} - \mathbf{K}_{i+1} \mathbf{C} \Sigma_{i+1|i}\end{aligned}$$

where the Kalman gain matrix $\mathbf{K}_{i+1} = \Sigma_{i+1|i} \mathbf{C}^T (\mathbf{C} \Sigma_{i+1|i} \mathbf{C}^T + \mathbf{R})^{-1}$ which we use to update the posterior $\mathbf{z}_{t+1} \approx \mathcal{N}(\mu_{i+1|i+1}, \Sigma_{i+1|i+1})$. With the updated \mathbf{z}_{t+1} we can compute the next prior $\mathbf{z}_{t+2} \approx \mathcal{N}(\mu_{i+2|i+1}, \Sigma_{i+2|i+1})$ where

$$\begin{aligned}\mu_{i+2|i+1} &= \mathbf{A}\mu_{i+1|i+1} \\ \Sigma_{i+2|i+1} &= \mathbf{A}\Sigma_{i+1|i+1}\mathbf{A}^t + \mathbf{Q}\end{aligned}$$

which is used to obtain the prediction $\mathbf{x}_{t+2} \approx \mathcal{N}(\mathbf{C}\mu_{i+2|i+1}, \mathbf{C}\Sigma_{i+2|i+1}\mathbf{C}^t + \mathbf{R})$.

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?

When the observation \mathbf{x}_j is missing the update \mathbf{x}_j and the Kalman gain \mathbf{K}_j are zero. Filtering equations become:

$$\begin{aligned}\mu_{j|j} &= \mu_{j|j-1} = \mathbf{A}\mu_{j-1|j-1} \\ \Sigma_{j|j} &= \Sigma_{j|j-1} = \mathbf{A}\Sigma_{j-1|j-1}\mathbf{A}^t + \mathbf{Q}\end{aligned}$$

And the smoothing equations are now:

$$\begin{aligned}\mu_{j|t} &= \mathbf{A}\mu_{j-1|j-1} + \mathbf{F}_j(\mu_{j+1|t} - \mu_{j+1|j}) = \mathbf{A}\mu_{j-1|j-1} + \mathbf{F}_j(\mu_{j+1|t} - \mathbf{A}\mu_{j|j}) \\ \Sigma_{j|t} &= \Sigma_{j|j} + \mathbf{F}_j(\Sigma_{j+1|t} - \Sigma_{j+1|j})\mathbf{F}_j^t \\ \mathbf{F}_j &= \Sigma_{j|j}\mathbf{A}^t\Sigma_{j+1|j}^{-1} = (\mathbf{A}\Sigma_{j-1|j-1}\mathbf{A}^t + \mathbf{Q})\mathbf{A}^t(\mathbf{A}\Sigma_{j|j}\mathbf{A}^t + \mathbf{Q})^{-1}\end{aligned}$$

We can recursively express $\mu_{j+1|j}$, $\Sigma_{j|j}$ in term of $\Sigma_{j-1|j-1}$ in both $\mu_{j|t}$ and $\Sigma_{j|t}$ up to μ_0 and Σ_0 . So $\mu_{j|t}$ and $\Sigma_{j|t}$ are reduced with the exceptions of the terms $\mu_{j+1|t}$ and $\Sigma_{j+1|t}$, to matrix multiplications of \mathbf{A} , \mathbf{Q} , μ_0 and Σ_0 .

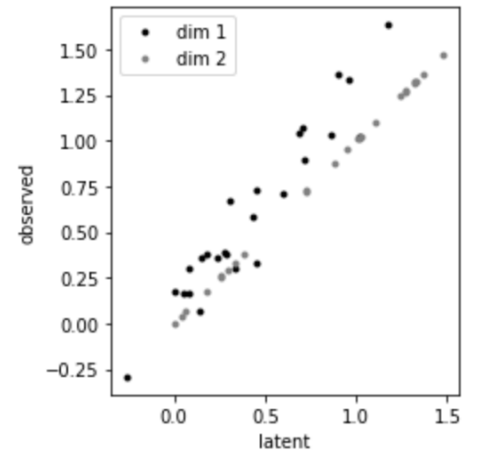
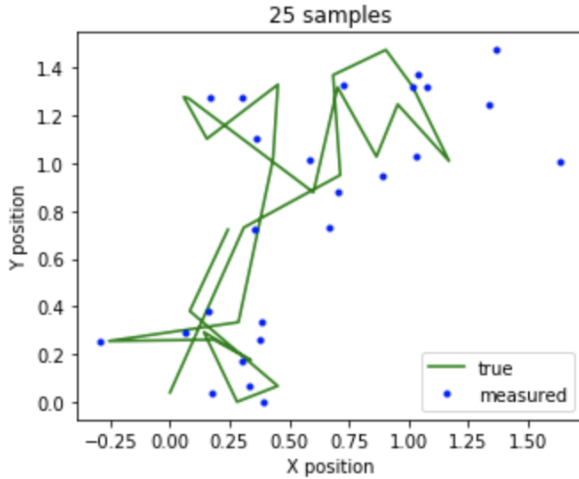
Problem 4. LDS filtering, 10p

Given the model parameters: $\mathbf{A} = \begin{bmatrix} 0.65 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$, $\mathbf{C} = \begin{bmatrix} 1.1 & 0.2 \\ 0.5 & 0.95 \end{bmatrix}$,

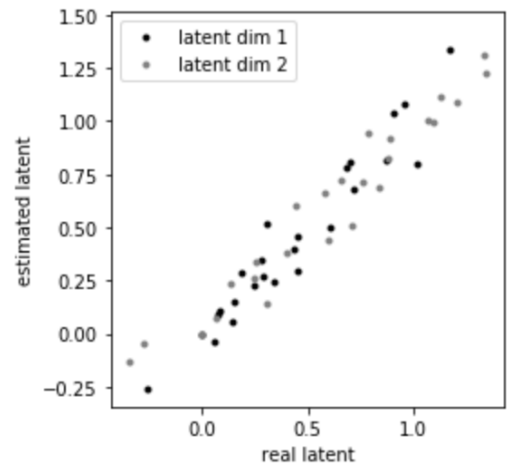
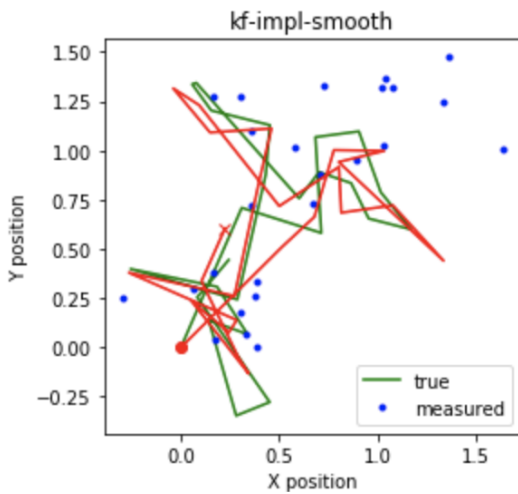
$\mathbf{Q} = 0.1\mathbf{I}_2$, $\mathbf{R} = 0.01\mathbf{I}_2$ with initial condition parameters $\mu_0 = [0 \ 0]^t$, $\Sigma_0 = 0.5\mathbf{I}$. Generate 25 samples from the model and plot them. Use these observation for inference (filtering and smoothing). How much does the Kalman gain \mathbf{K}_i , 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.

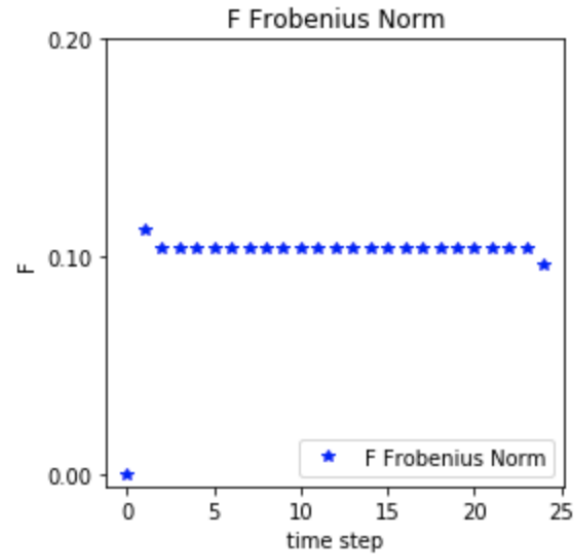
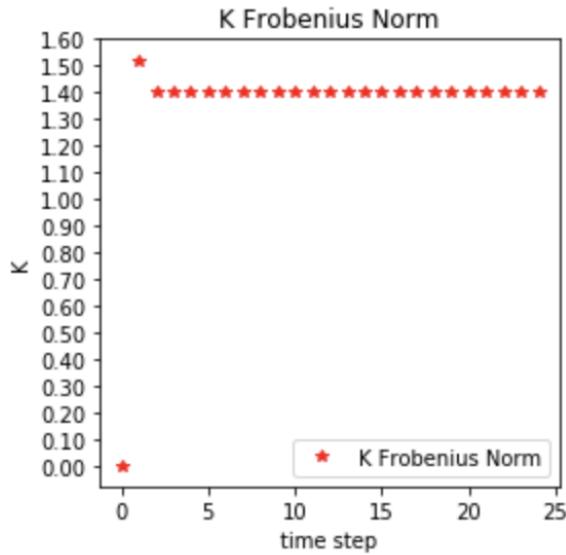
With the given parameters, using the code from the lab, the sampling of 25 observations gives us the plots below:



After inference (filtering and smoothing) we obtain the estimates in red:



At iteration i , Kalman gain is the matrix $\mathbf{K}_i = \Sigma_{i|i-1}\mathbf{C}^t(\mathbf{C}\Sigma_{i|i-1}\mathbf{C}^t + \mathbf{R})^{-1}$ and the "smooth" gain $\mathbf{F}_i = \Sigma_{i|i}\mathbf{A}^t\Sigma_{i+1|i}^{-1}$. At each time step, we compute the frobenius norm: $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{trace}(\mathbf{A}^*\mathbf{A})} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(\mathbf{A})}$ of the Kalman gain and F matrices and we plot them. On the plots below, we can see that once established since the system is linear, the Kalman matrix does not change.

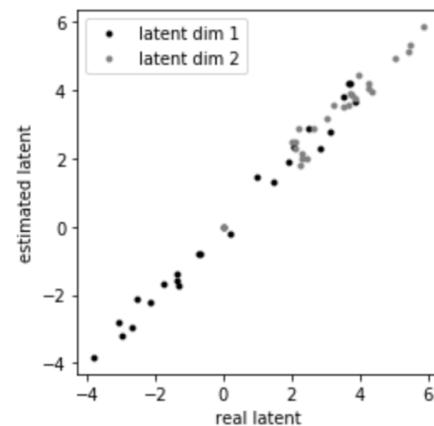
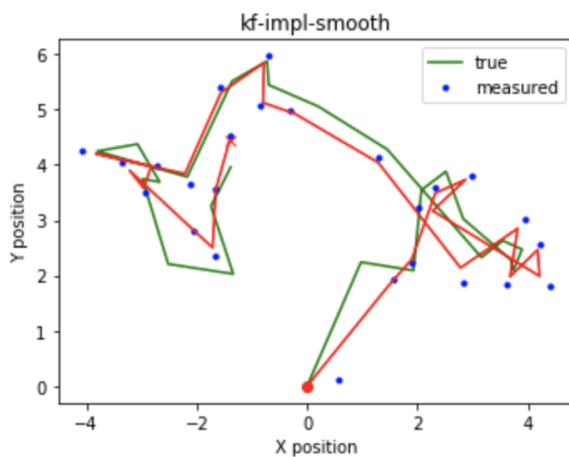


The intuition behind the Kalman gain are:

1. The more noisy the measurement is, the less precise it is. Hence, the innovation that represents the bias in the measurement may not be real innovation but rather an artifact of the measurement noise. Noise or uncertainty is directly captured by variance. Thus the larger the measurement variance noise, the lower the Kalman gain should be.
2. The more noisy the process state is, the more important the innovation should be taken into account. Hence, the larger the process state variance, the larger the Kalman gain should be. Summarizing these two intuitions, we would expect the Kalman gain to be:

$$K_i = \frac{\text{Process Noise}}{\text{Measurement Noise}}$$

Using a smoother dynamic system, we confirm the intuition experimentally, as described below, we vary different parameters and observe the results listed below:



As we reduce the observation noise the Kalman Gain increases and the F "Gain" decreases.

df	K-Norm-Mean	K-Norm-Var	F-Norm-Mean	F-Norm-Var
0	1.24374	0.06445	0.11908	0.00000
1	1.24374	0.06445	0.11908	0.00000
2	1.34438	0.07531	0.01845	0.00051
3	1.35630	0.07665	0.00650	0.00062

As we increase the temporal dynamics (transition matrix) the Kalman Gain and F increase.

```
4]: df
```

	K-Norm-Mean	K-Norm-Var	F-Norm-Mean	F-Norm-Var
0	1.24374	0.06445	0.11394	0.00054
1	1.24374	0.06445	0.11394	0.00054
2	1.34532	0.07541	0.16918	0.05159
3	1.35751	0.07678	0.52721	6.33260

As we increase the observation dynamics the Kalman Gain and F decrease.

```
26]: df
```

	K-Norm-Mean	K-Norm-Var	F-Norm-Mean	F-Norm-Var
0	1.24374	0.06445	0.11394	0.00054
1	1.24374	0.06445	0.11394	0.00054
2	0.13563	0.00077	0.00135	0.00000
3	0.01358	0.00001	0.00001	0.00000

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.

Assume we are given K samples from $\mathbf{z}_i^{(k)} \sim P(\mathbf{z}_i | \mathbf{x}_{1:i-1})$.

For $i > 1$ similarly to the LDS model, we start by computing the prior for \mathbf{z}_i :

$$\begin{aligned}\mu_{i|i-1} &= \mathbf{A}\mu_{i-1|i-1} \\ \Sigma_{i|i-1} &= \mathbf{A}\Sigma_{i-1|i-1}\mathbf{A}^T + \mathbf{Q}\end{aligned}$$

and samples from $\mathcal{N}(\mu_{i|i-1}, \Sigma_{i|i-1})$ which gives us $\mathbf{z}_i^{(k)} \sim P(\mathbf{z}_i | \mathbf{x}_{1:i-1})$. We then project into the observational space, equivalent to the step $(\mathbf{C}\mu_{i|i-1}, \mathbf{C}\Sigma_{i|i-1}\mathbf{C}^T + \Sigma)$ for the Kalman filter, and compute weights which incorporate the observations \mathbf{x}_i .

$$w_i^{(k)} = \frac{P(\mathbf{x}_i | \mathbf{z}_i^{(k)})}{\sum_{l=1, K} P(\mathbf{x}_i | \mathbf{z}_i^{(l)})}$$

We have that:

$$\begin{aligned}\mathbb{E}_{\mathbf{z}_i | \mathbf{x}_{1:i}}[\mathbf{z}_i] &= \mu_{i|i} = \int \mathbf{z}_i P(\mathbf{z}_i | \mathbf{x}_{1:i}) d\mathbf{z}_i \\ &= \int \mathbf{z}_i P(\mathbf{z}_i | \mathbf{x}_{1:i-1}, \mathbf{x}_i) d\mathbf{z}_i \\ &= \int \mathbf{z}_i \frac{P(\mathbf{x}_i | \mathbf{z}_i) P(\mathbf{z}_i | \mathbf{x}_{1:i-1})}{\int P(\mathbf{x}_i | \mathbf{z}_i) P(\mathbf{z}_i | \mathbf{x}_{1:i-1}) d\mathbf{x}_i \mathbf{z}_i} d\mathbf{z}_i \\ &\sim \sum_{k=1, N} w_i^{(k)} \mathbf{z}_i^{(k)} \text{ where } w_i^{(k)} = \frac{P(\mathbf{x}_i | \mathbf{z}_i^{(k)})}{\sum_{l=1, K} P(\mathbf{x}_i | \mathbf{z}_i^{(l)})} \\ 0 &\leq w_i^{(k)} \leq 1 \quad \sum_{k=1, K} w_i^{(k)} = 1 \\ \Sigma_{i|i} &= \hat{V}[\mathbf{z}_i] = \mathbb{E}[\mathbf{z}_i^2] - \mathbb{E}[\mathbf{z}_i]\mathbb{E}[\mathbf{z}_i]\end{aligned}$$

where $\mathbb{E}_{\mathbf{z}_i|\mathbf{x}_{1:i}}[\mathbf{z}_i^2] \sim \sum_{k=1,K} w_i^{(k)} \mathbf{z}_i^{(k)^2}$.

At this step we compute the prior on \mathbf{z}_{i+1} using a 2-step sampling:

- (a) Draw K independent class assignments $c_i^{(k)}$ based on weights $w_i^{(k)}$.
- (b) Compute

$$\begin{aligned} P(\mathbf{z}_{i+1}|\mathbf{x}_{1:i}) &= \mu_{i+1|i} = \int P(\mathbf{z}_{i+1}|\mathbf{z}_i, \mathbf{x}_{1:i}) P(\mathbf{z}_i|\mathbf{x}_{1:i}) = \int P(\mathbf{z}_{i+1}|\mathbf{z}_i) P(\mathbf{z}_i|\mathbf{x}_{1:i}) d\mathbf{z}_i \\ &\sim \sum_k w_i^{(k)} P(\mathbf{z}_{i+1}|\mathbf{z}_i^{(k)}) \end{aligned}$$

where we used the conditional independence between \mathbf{z}_{i+1} and $\mathbf{x}_{1:i}$.

Algorithm 1 Pseudo-code for particle filter

INPUT:

K - Number of samples

\mathbf{A} - transition matrix

\mathbf{C} - observation matrix

$\mathbf{\Gamma}$ - transition covariance matrix

$\mathbf{\Sigma}$ - transition covariance matrix

Draw N samples from $P(\mathbf{z}_1) = \mathcal{N}(\mu_0, \Sigma_0)$

for $i \leftarrow 1$ to T **do**

Sample from $P(\mathbf{x}_i|\mathbf{z}_i^{(k)}) \leftarrow \mathcal{N}(\mathbf{x}_i, \mathbf{z}_i^{(k)} * \mathbf{\Gamma}, \mathbf{\Sigma})$

Compute importance weights based on observation \mathbf{x}_i where $w_i^{(k)} \leftarrow \frac{P(\mathbf{x}_i|\mathbf{z}_i^{(k)})}{\sum_{l=1,K} P(\mathbf{x}_i|\mathbf{z}_i^{(l)})}$

Draw N independent class assignments $\{c_i^1, \dots, c_i^K\} \leftarrow \text{Multinomial}\{w_i^1, \dots, w_i^K\}$

Resample $\mathbf{z}_i^{c_k}$ based on class assignment

$\mathbf{z}_{i+1} \leftarrow \mathbf{A}\mathbf{z}_i^{c_k} + \mathcal{N}(0, \mathbf{Q})$

Let compute $\mathbb{E}[\mathbf{z}_i\mathbf{z}_{i+1}|\mathbf{x}_{1:i+1}]$. Using Bayes' rule:

$$\begin{aligned} \mathbb{E}[\mathbf{z}_i\mathbf{z}_{i+1}|\mathbf{x}_{1:i+1}] &= \sum_{\mathbf{z}_i, \mathbf{z}_{i+1}} \xi(\mathbf{z}_i, \mathbf{z}_{i+1}) \mathbf{z}_i\mathbf{z}_{i+1} \\ \xi(\mathbf{z}_i, \mathbf{z}_{i+1}) &= P(\mathbf{z}_i, \mathbf{z}_{i+1}|\mathbf{x}_{1:i+1}) \\ P(\mathbf{z}_i, \mathbf{z}_{i+1}|\mathbf{x}_{1:i+1}) &= \frac{P(\mathbf{x}_{1:i+1}|\mathbf{z}_i, \mathbf{z}_{i+1}) P(\mathbf{z}_i, \mathbf{z}_{i+1})}{P(\mathbf{x}_{1:i+1})} \\ P(\mathbf{z}_i, \mathbf{z}_{i+1}|\mathbf{x}_{1:i+1}) &= \frac{P(\mathbf{x}_{1:i}|\mathbf{z}_i) P(\mathbf{z}_i) P(\mathbf{z}_{i+1}|\mathbf{z}_i) P(\mathbf{x}_{i+1}|\mathbf{z}_{i+1})}{P(\mathbf{x}_{1:i+1})} \end{aligned}$$

The numerator is constant and given by the distribution of the data. We want to determine $P(\mathbf{x}_{1:i}|\mathbf{z}_i)$, $P(\mathbf{z}_i)$, $P(\mathbf{z}_{i+1}|\mathbf{z}_i)$ and $P(\mathbf{x}_{i+1}|\mathbf{z}_{i+1})$. Starting with $P(\mathbf{x}_{1:i}|\mathbf{z}_i)$, using Bayes' rule:

$$\begin{aligned} P(\mathbf{x}_{1:i}|\mathbf{z}_i) &= \frac{P(\mathbf{z}_i|\mathbf{x}_{1:i}) P(\mathbf{x}_{1:i})}{\int P(\mathbf{x}_{1:i}) d\mathbf{x}_i} \\ P(\mathbf{z}_i|\mathbf{x}_{1:i}) &= \text{unnormalized weights} \left(\sum_{l=1,K} P(\mathbf{x}_i|\mathbf{z}_i^{(k)}) \right) \cdot w_i^{(k)} \end{aligned}$$

$P(\mathbf{x}_{1:i}), \int P(\mathbf{x}_{1:i}) d\mathbf{x}_i =$ probability distributions given by the data

$P(\mathbf{z}_i) \sim P(\mathbf{z}_i|\mathbf{x}_{1:i-1})$ computed at step i of the particle filter algorithm

$P(\mathbf{z}_{i+1}|\mathbf{z}_i) =$ probabilities computed above in the expressions of $P(\mathbf{z}_{i+1}|\mathbf{x}_{1:i})$

$$P(\mathbf{x}_{i+1}|\mathbf{z}_{i+1}) = \text{unnormalized weights} \left(\sum_{l=1,K} P(\mathbf{x}_{i+1}|\mathbf{z}_{i+1}^{(k)}) \right) \cdot w_{i+1}^{(k)}$$