Probabilistic Modelling and Reasoning, Tutorial Answer Sheet 7 (for Week 9)

School of Informatics, University of Edinburgh

Instructor: Amos Storkey
Prob. Sheet. ID: 27114.1

1. The model is defined by the following distributions

$$\begin{aligned} \boldsymbol{z}_1 &\sim N(\boldsymbol{\mu}_0, P_0) && \text{(Initialisation)} \\ \boldsymbol{z}_{n+1}|\boldsymbol{z}_n &\sim N(A\boldsymbol{z}_n, \Gamma) && \text{(Dynamic model)} \\ \boldsymbol{x}_n|\boldsymbol{z}_n &\sim N(C\boldsymbol{z}_n, \Sigma) && \text{(Observation model)} \end{aligned}$$

giving the following update equations

$$\begin{aligned} P_{n-1} &= AV_{n-1}A^{\mathrm{T}} + \Gamma && \text{(time update of state covariance)} \\ \boldsymbol{\mu}_n &= A\boldsymbol{\mu}_{n-1} + K_n(\boldsymbol{x}_n - CA\boldsymbol{\mu}_{n-1}) && \text{(measurement update of state mean)} \\ V_n &= (I - K_nC)P_{n-1} && \text{(measurement update of state covariance)} \end{aligned}$$

where $K_n = P_{n-1}C^{\mathrm{T}}(CP_{n-1}C^{\mathrm{T}} + \Sigma)^{-1}$.

Our system has

$$\boldsymbol{\mu}_0 = 0 \qquad P_0 = \sigma_0^2 \qquad A = 1 \qquad \Gamma = 1 \qquad C = 1 \qquad \Sigma = 1$$

for $n = 1$,

$$\begin{aligned} K_1 &= \frac{\sigma_0^2}{\sigma_0^2 + 1} \\ \mu_1 &= \frac{\sigma_0^2}{\sigma_0^2 + 1}x_1 && \lim_{\sigma_0^2 \to \infty} \mu_1 = x_1 \\ V_1 &= \left(1 - \frac{\sigma_0^2}{\sigma_0^2 + 1}\right)\sigma_0^2 = \frac{\sigma_0^2}{\sigma_0^2 + 1} && \lim_{\sigma_0^2 \to \infty} V_1 = 1 \end{aligned}$$

for $n = 2$,

$$P_1 = \frac{\sigma_0^2}{\sigma_0^2 + 1} + 1 = \frac{2\sigma_0^2 + 1}{\sigma_0^2 + 1} \qquad \lim_{\sigma_0^2 \to \infty} P_1 = 2$$

$$K_2 = \frac{2\sigma_0^2 + 1}{\sigma_0^2 + 1} \left(\frac{2\sigma_0^2 + 1}{\sigma_0^2 + 1} + 1 \right)^{-1} = \frac{2\sigma_0^2 + 1}{3\sigma_0^2 + 2} \qquad \lim_{\sigma_0^2 \to \infty} K_2 = \frac{2}{3}$$

$$\mu_2 = \mu_1 + K_2(x_2 - \mu_1) = (1 - K_2)\mu_1 + K_2 x_2$$

$$= \frac{\sigma_0^2}{3\sigma_0^2 + 2} x_1 + \frac{2\sigma_0^2 + 1}{3\sigma_0^2 + 2} x_2 \qquad \lim_{\sigma_0^2 \to \infty} \mu_2 = \frac{1}{3} x_1 + \frac{2}{3} x_2$$

$$V_2 = \left(1 - \frac{2\sigma_0^2 + 1}{3\sigma_0^2 + 2} \right) \frac{2\sigma_0^2 + 1}{\sigma_0^2 + 1} = \frac{2\sigma_0^2 + 1}{3\sigma_0^2 + 2} \qquad \lim_{\sigma_0^2 \to \infty} V_2 = \frac{2}{3}$$

for $n = 3$,

$$P_2 = \frac{2\sigma_0^2 + 1}{3\sigma_0^2 + 2} + 1 = \frac{5\sigma_0^2 + 3}{3\sigma_0^2 + 2} \qquad \lim_{\sigma_0^2 \to \infty} P_2 = \frac{5}{3}$$

$$K_3 = \frac{5\sigma_0^2 + 3}{3\sigma_0^2 + 2} \left(\frac{5\sigma_0^2 + 3}{3\sigma_0^2 + 2} + 1 \right)^{-1} = \frac{5\sigma_0^2 + 3}{8\sigma_0^2 + 5} \qquad \lim_{\sigma_0^2 \to \infty} K_3 = \frac{5}{8}$$

$$\mu_3 = \mu_2 + K_3(x_3 - \mu_2) = (1 - K_3)\mu_2 + K_3 x_3$$

$$= \frac{(\sigma_0^2 x_1 + (2\sigma_0^2 + 1)x_2 + (5\sigma_0^2 + 3)x_3}{8\sigma_0^2 + 5} \qquad \lim_{\sigma_0^2 \to \infty} \mu_3 = \frac{x_1 + 2x_2 + 5x_3}{8}$$

$$V_3 = \left(1 - \frac{5\sigma_0^2 + 3}{8\sigma_0^2 + 5} \right) \frac{5\sigma_0^2 + 3}{3\sigma_0^2 + 2} = \frac{5\sigma_0^2 + 3}{8\sigma_0^2 + 5} \qquad \lim_{\sigma_0^2 \to \infty} V_3 = \frac{5}{8}$$

Note: Present is weighted more than the past
Note: Can work in the limits directly instead of using σ_0^2 all the way.

2. Draw pictures: sinusoidal variation in each component. If we add noise it adds a random walk on at each point in time. As this noise is small, one can approximately think of breaking the noise into a radial component, a transverse component in the direction of the rotation and a transverse component perpendicular to the rotation. Note this is an approximation, and thinking of the cumulative radial component as towards the centre is not accurate.

   If $\alpha < 1$ then the rotation contract space, and so the systems will tend to a Gaussian distribution that balances out the contraction with the additive Gaussian noise.

   If $alpha > 1$ then the system gets larger and the deterministic part eventually dominates the noise so the result is a path that is effectively a spiral. However the initial noise will affect the location of the spiral.

3. The wrong rotation matrix would mean that the noise component would have to compensate for the over or under-rotated system. A larger noise term would be needed, and the size of those noise terms would have to be larger, resulting in a poorer likelihood. A deterministic model in 2 dimensions would fail due to phase shifts as well as shifts in magnitude. In higher dimensions it is worse still, as the noise takes the points completely off the original plane of rotation. The deterministic model may be okay for small time periods, but would soon fail over longer time periods.