

Linear Dynamical Systems Notes

Joaquin Rapela*

May 14, 2022

1 Linear dynamical systems (LDS) model

$$\begin{aligned} \mathbf{x}_n &= A\mathbf{x}_{n-1} + \mathbf{w}_n & \mathbf{w}_n &\sim N(\mathbf{w}_n|\mathbf{0}, Q) & \mathbf{x}_n &\in \mathbb{R}^m \\ \mathbf{y}_n &= C\mathbf{x}_n + \mathbf{v}_n & \mathbf{v}_n &\sim N(\mathbf{v}_n|\mathbf{0}, R) & \mathbf{y}_n &\in \mathbb{R}^n \quad n = 1 \dots N \\ \mathbf{x}_0 &\sim N(\mathbf{x}_0|\mathbf{m}_0, V_0) \end{aligned}$$

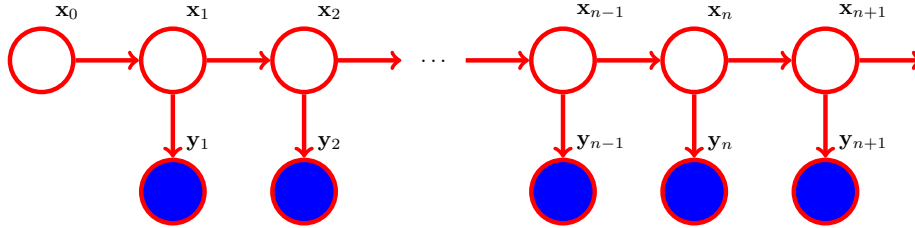


Figure 1: Graphical model for linear dynamical systems

2 Inference Problems

Prediction

$$P(\mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_{n-1}) = N(\mathbf{x}_n|\mathbf{x}_{n|n-1}, P_{n|n-1}) \quad (1)$$

Filtering

$$P(\mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_n) = N(\mathbf{x}_n|\mathbf{x}_{n|n}, P_{n|n}) \quad (2)$$

Smoothing

$$P(\mathbf{x}_n|\mathbf{y}_1, \dots, \mathbf{y}_N) = N(\mathbf{x}_n|\mathbf{x}_{n|N}, P_{n|N}) \quad (3)$$

*j.rapela@ucl.ac.uk

3 Kalman Filter

The Kalman filter algorithm addresses the prediction (Eq. 1) and filtering (Eq. 2) inference problems. It is an iterative algorithm, which alternates between computing the mean and covariance of the prediction distribution and computing the mean and covariance of the filtering distribution.

$\mathbf{x}_{1 0} = \mathbf{m}_0$	init prediction mean
$P_{1 0} = V_0$	init prediction covariance
$\mathbf{y}_{n n-1} = C\mathbf{x}_{n n-1}$	predicted observation
$\tilde{\mathbf{y}}_n = \mathbf{y}_n - \mathbf{y}_{n n-1}$	residual
$S_n = CP_{n n-1}C^\top + R$	residual covariance
$\mathbf{x}_{n n} = \mathbf{x}_{n n-1} + K_n\tilde{\mathbf{y}}_n$	filtering mean
$P_{n n} = (I_M - K_nC)P_{n n-1}$	filtering covariance
$K_n = P_{n n-1}C^\top S_n^{-1}$	Kalman gain
$\mathbf{x}_{n+1 n} = A\mathbf{x}_{n n}$	prediction mean
$P_{n+1 n} = AP_{n n}A^\top + Q$	prediction covariance

Inference of predicted and filtered mean and covariances proceeds in a forward fashion, inferring the prediction and filtering distributions from the first to the last state, as shown in the next figure.

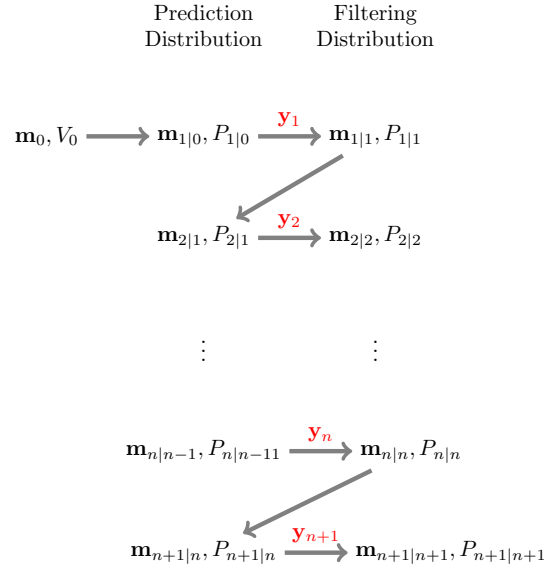


Figure 2: Order of calculation of the prediction and filtering distributions in the Kalman filtering algorithm.

4 Kalman Smoother

The Kalman smoother algorithm addresses the smoothing (Eq. 3) inference problem.

$$\begin{aligned}
x_{n|N} &= x_{n|n} + C_n(x_{n+1|N} - x_{n+1|n}) && \text{smoothed mean} \\
P_{n|N} &= P_{n|n} + C_n(P_{n+1|N} - P_{n+1|n})C_n^\top && \text{smoothed covariance} \\
C_n &= P_{n|n}A^\top P_{n+1|n}^{-1}
\end{aligned}$$

Inference of smoothed mean and covariances proceeds in a backward fashion: $x_{N|N}, P_{N|N} \rightarrow x_{N-1|N}, P_{N-1|N} \rightarrow \dots \rightarrow x_{1|N}, P_{1|N}$. The initial mean and covariances (i.e., $x_{N|N}, P_{N|N}$) are initialized from the last step of the Kalman filter.

5 Derivation of Kalman filter equations

Claim 1. $\mathbf{x}_{n|n-1} = A\mathbf{x}_{n-1|n-1}$

Proof.

$$\begin{aligned}
\mathbf{x}_{n|n-1} &= E\{\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} \\
&= E\{A\mathbf{x}_{n-1} + \mathbf{w}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} && (4) \\
&= AE\{\mathbf{x}_{n-1} | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} + E\{\mathbf{w}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} && (5) \\
&= A\mathbf{x}_{n-1|n-1} + E\{\mathbf{w}_n\} && (6) \\
&= A\mathbf{x}_{n-1|n-1} && (7)
\end{aligned}$$

□

Notes:

- Eq. 4 arises from the state equation of the LDS model in Section 1,
- Eq. 5 holds because the expectation distributes over sums
- Eq. 6 uses the definition of $\mathbf{x}_{n-1|n-1}$ and the fact that the state noise, \mathbf{w}_n , is independent of previous observations.
- Eq. 7 follows due to the zero mean of \mathbf{w}_n .

Claim 2. $P_{n|n-1} = AP_{n-1|n-1}A^\top + Q$

Proof.

$$\begin{aligned}
P_{n|n-1} &= \text{Cov}\{\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} \\
&= \text{Cov}\{A\mathbf{x}_{n-1} + \mathbf{w}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} && (8) \\
&= \text{Cov}\{A\mathbf{x}_{n-1} | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} + \text{Cov}\{\mathbf{w}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} && (9) \\
&= A \text{Cov}\{\mathbf{x}_{n-1} | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} A^\top + \text{Cov}\{\mathbf{w}_n\} && (10) \\
&= A P_{n-1|n-1} A^\top + Q && (11)
\end{aligned}$$

□

Notes:

1. Eq. 8 used the state equation of the LDS model in Section 1,

2. Eq. 9 is true because \mathbf{w}_n is independent from \mathbf{x}_{n-1} ,
3. Eq. 10 holds by the property $\text{Cov}\{A\mathbf{x}\} = A \text{Cov}\{\mathbf{x}\} A^\top$ and because \mathbf{w}_n is independent of previous observations.
4. Eq. 10 applied the definitions of $P_{n-1|n-1}$ and Q .

Claim 3. $\mathbf{x}_{n|n} = \mathbf{x}_{n|n-1} + K_n \tilde{\mathbf{y}}_n$ and $P_{n|n} = (I - K_n C) P_{n|n-1}$.

Proof. Define the random variables $\mathbf{x} = \mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}$ and $\mathbf{y} = \mathbf{y}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}$. Then $\mathbf{x} | \mathbf{y} = \mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_n$ and the mean and covariance that we want to find, $\mathbf{x}_{n|n}$ and $P_{n|n}$, are those of $\mathbf{x} | \mathbf{y}$. Thus, we want to compute the mean, $\mu_{\mathbf{x}|\mathbf{y}} = \mathbf{x}_{n|n}$, and covariance, $\Sigma_{\mathbf{x}|\mathbf{y}} = P_{n|n}$, of $\mathbf{x} | \mathbf{y}$.

Because \mathbf{x}_n and \mathbf{y}_n are jointly Gaussian, then \mathbf{x} and \mathbf{y} are also jointly Gaussian. Then, $\mu_{\mathbf{x}|\mathbf{y}}$ and $\Sigma_{\mathbf{x}|\mathbf{y}}$ are (Bishop, 2016, Chapter 2)

$$\mu_{\mathbf{x}|\mathbf{y}} = \mu_{\mathbf{x}} + \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} (\mathbf{y}_n - \mu_{\mathbf{y}}) \quad (12)$$

$$\Sigma_{\mathbf{x}|\mathbf{y}} = \Sigma_{\mathbf{xx}} - \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \quad (13)$$

Thus, to compute $\mu_{\mathbf{x}|\mathbf{y}}$ and $\Sigma_{\mathbf{x}|\mathbf{y}}$ we need to calculate $\mu_{\mathbf{x}}$, $\mu_{\mathbf{y}}$, $\Sigma_{\mathbf{xx}}$, $\Sigma_{\mathbf{xy}}$ and $\Sigma_{\mathbf{yy}}$.

$$\mu_{\mathbf{x}} = E\{\mathbf{x}\} = E\{\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} = \mathbf{x}_{n|n-1} \quad (14)$$

$$\begin{aligned} \mu_{\mathbf{y}} &= E\{\mathbf{y}\} = E\{\mathbf{y}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} = E\{C\mathbf{x}_n + \mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} = \\ &= CE\{\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} + E\{\mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} = C\mathbf{x}_{n|n-1} + E\{\mathbf{v}_n\} = C\mathbf{x}_{n|n-1} = \mathbf{y}_{n|n-1} \end{aligned} \quad (15)$$

Notes:

- The penultimate equality in Eq. 15 uses the definition of $\mathbf{x}_{n|n-1}$ and the fact that \mathbf{v}_n is independent of previous observations.
 - The last equality in Eq. 15 holds because the mean of \mathbf{v}_n is zero. Claim 1.
-

$$\begin{aligned} \Sigma_{\mathbf{yy}} &= \text{Cov}(\mathbf{y}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \\ &= \text{Cov}(C\mathbf{x}_n + \mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \\ &= \text{Cov}(C\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) + \text{Cov}(\mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \\ &= C \text{Cov}(\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) C^\top + \text{Cov}(\mathbf{v}_n) \end{aligned} \quad (16)$$

$$= CP_{n|n-1} C^\top + R \quad (17)$$

Notes:

- As in Eq. 9, Eq. 16 holds by the property $\text{Cov}\{A\mathbf{x}\} = A \text{Cov}\{\mathbf{x}\} A^\top$ and because \mathbf{v}_n is independent of previous observations.

$$\begin{aligned}
\Sigma_{\mathbf{xy}} &= \text{cCov}(\mathbf{x}_n, \mathbf{y}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \\
&= \text{cCov}(\mathbf{x}_n, C\mathbf{x}_n + \mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \\
&= \text{cCov}(\mathbf{x}_n, C\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) + \text{cCov}(\mathbf{x}_n, \mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) \\
&= \text{Cov}(\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) C^\top + 0 \\
&= P_{n|n-1} C^\top
\end{aligned}
\tag{18}$$

$$= \text{Cov}(\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) C^\top + 0 \tag{19}$$

$$= P_{n|n-1} C^\top \tag{20}$$

Notes:

- the first term in Eq. 19 arises from the first term of Eq. 18 since

$$\begin{aligned}
\text{cCov}(\mathbf{x}_n, C\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) &= E\{(\mathbf{x}_n - \mu_{\mathbf{x}})(C\mathbf{x}_n - C\mu_{\mathbf{x}})^\top | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} \\
&= E\{(\mathbf{x}_n - \mu_{\mathbf{x}})(\mathbf{x}_n - \mu_{\mathbf{x}})^\top C^\top | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} \\
&= E\{(\mathbf{x}_n - \mu_{\mathbf{x}})(\mathbf{x}_n - \mu_{\mathbf{x}})^\top | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} C^\top \\
&= \text{Cov}(\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) C^\top
\end{aligned}$$

- the second term in Eq. 19 arises from the second term of Eq. 18 since

$$\begin{aligned}
\text{cCov}(\mathbf{x}_n, \mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) &= E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1})\mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} \\
&= E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1}) | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} E\{\mathbf{v}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} \\
&= E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1}) | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} E\{\mathbf{v}_n\} \\
&= E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1}) | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}\} 0 \\
&= 0
\end{aligned}$$

the second line follows from the first one because \mathbf{v}_n is independent of \mathbf{x}_n , and the third line follows from the second one because \mathbf{v}_n is independent of previous observations.

$$\Sigma_{\mathbf{xx}} = \text{Cov}(\mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_{n-1}) = P_{n|n-1} \tag{21}$$

Having calculated $\mu_{\mathbf{x}}$, $\mu_{\mathbf{y}}$, $\Sigma_{\mathbf{xx}}$, $\Sigma_{\mathbf{xy}}$ and $\Sigma_{\mathbf{yy}}$ we now use Eqs. 12, 13, 14, 15, 17, 20, and 21 to obtain $\mathbf{x}_{n|n}$ and $P_{n|n}$.

$$\begin{aligned}
\mathbf{x}_{n|n} &= \mu_{\mathbf{x}|\mathbf{y}} = \mu_{\mathbf{x}} + \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} (\mathbf{y}_n - \mu_{\mathbf{y}}) \\
&= \mathbf{x}_{n|n-1} + P_{n|n-1} C^\top S_n^{-1} (\mathbf{y}_n - \mathbf{y}_{n|n-1}) \\
&= \mathbf{x}_{n|n-1} + K_n \tilde{\mathbf{y}}_n \\
P_{n|n} &= \Sigma_{\mathbf{x}|\mathbf{y}} = \Sigma_{\mathbf{xx}} - \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \\
&= P_{n|n-1} - P_{n|n-1} C^\top S_n^{-1} C P_{n|n-1} \\
&= (I - P_{n|n-1} C^\top S_n^{-1} C) P_{n|n-1} \\
&= (I - K_n C) P_{n|n-1}
\end{aligned}$$

□

References

Bishop, C. M. (2016). *Pattern recognition and machine learning*. Springer-Verlag New York.