Derivation of the Kalman filter equations

Joaquín Rapela

June 12, 2025

Theorem 1. *Given the linear dynamical systems model*

$$
\begin{aligned}
\mathbf{x}_{t+1} &= A_t\mathbf{x}_t + \mathbf{w}_t \quad \textit{with } \mathbf{w}_t \sim N(0, Q_t)\\
\mathbf{y}_t &= B_t\mathbf{x}_t + \mathbf{v}_t \quad \textit{with } \mathbf{v}_t \sim N(0, R_t)\\
\mathbf{x}_0 &\sim N(\mathbf{m}_0, V_0)
\end{aligned}
$$

(represented in Fig. 1), then the predictive distribution, $p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})$*, and the filtering distribution,* $p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_t)$*, are*

$$
\begin{aligned}
p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}) &= N(\mathbf{x}_t|\mathbf{x}_{t|t-1}, P_{t|t-1})\\
p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_t) &= N(\mathbf{x}_t|\mathbf{x}_{t|t}, P_{t|t})
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{x}_{t|t-1} &= A_{t-1}\mathbf{x}_{t-1|t-1} && (1)\\
P_{t|t-1} &= A_{t-1}P_{t-1|t-1}A_{t-1}^\intercal + Q_{t-1} && (2)\\
\hat{\mathbf{y}}_{t|t-1} &\triangleq E\{\mathbf{y}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}\} = B_t\mathbf{x}_{t|t-1} && (3)\\
\mathbf{z}_t &\triangleq \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} &&\\
S_t &\triangleq Cov\{\mathbf{z}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}\} = B_tP_{t|t-1}B_t^\intercal + R_t && (4)\\
\mathbf{x}_{t|t} &= \mathbf{x}_{t|t-1} + K_t\mathbf{z}_t && (5)\\
\mathbf{P}_{t|t} &= (I - K_tB_t)P_{t|t-1} && (6)\\
\mathbf{K}_t &= P_{t|t-1}B_t^\intercal S_t^{-1} && (7)\\
\mathbf{x}_{0|0} &= \mathbf{m}_0 && (8)\\
P_{0|0} &= V_0 && (9)
\end{aligned}
$$

The following proof adds a few details to that given in Section 4.3.1 of Durbin and Koopman (2012).

Proof. Call $Y_t = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}$, then

$$
\mathbf{x}_{t|t-1} = E\{\mathbf{x}_t|Y_{t-1}\} = E\{A_{t-1}\mathbf{x}_{t-1} + \mathbf{w}_{t-1}|Y_{t-1}\}
$$

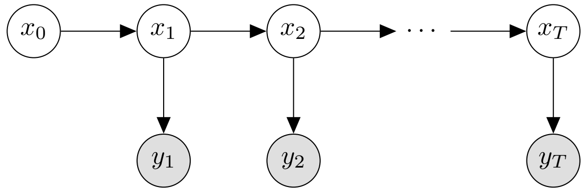

Figure 1: Graphical models for our linear dynamical system in Theorem 1.

$$
\begin{aligned}
&= A_{t-1}E\{\mathbf{x}_{t-1}|Y_{t-1}\} + E\{\mathbf{w}_{t-1}|Y_{t-1}\} \\
&= A_{t-1}\mathbf{x}_{t-1|t-1} + E\{\mathbf{w}_{t-1}\} = A_{t-1}\mathbf{x}_{t-1|t-1} \qquad (10)
\end{aligned}
$$

This proves Eq. 1.

$$
\begin{aligned}
\mathbf{P}_{t|t-1} &= \mathrm{Cov}\{\mathbf{x}_t|Y_{t-1}\} = E\{(\mathbf{x}_t - \mathbf{x}_{t|t-1})(\mathbf{x}_t - \mathbf{x}_{t|t-1})^\intercal|Y_{t-1}\} \\
&= E\{\left(A_{t-1}\mathbf{x}_{t-1} + w_{t-1} - A_{t-1}\mathbf{x}_{t-1|t-1}\right)\left(A_{t-1}\mathbf{x}_{t-1} + w_{t-1} - A_{t-1}\mathbf{x}_{t-1|t-1}\right)^\intercal|Y_{t-1}\} \\
&= E\{\left(A_{t-1}(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1}) + w_{t-1}\right)\left(A_{t-1}(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1}) + w_{t-1}\right)^\intercal|Y_{t-1}\} \\
&= A_{t-1}E\{(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})^\intercal|Y_{t-1}\}A_{t-1}^\intercal + \\
&\quad\; E\{w_{t-1}(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})^\intercal|Y_{t-1}\}A_{t-1}^\intercal + \\
&\quad\; A_{t-1}E\{(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})w_{t-1}^\intercal|Y_{t-1}\} + \\
&\quad\; E\{w_{t-1}w_{t-1}^\intercal|Y_{t-1}\} \\
&= A_{t-1}E\{(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})^\intercal|Y_{t-1}\}A_{t-1}^\intercal + \\
&\quad\; E\{w_{t-1}|Y_{t-1}\}E\{(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})^\intercal|Y_{t-1}\}A_{t-1}^\intercal + \\
&\quad\; A_{t-1}E\{(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})|Y_{t-1}\}E\{w_{t-1}^\intercal|Y_{t-1}\} + \\
&\quad\; E\{w_{t-1}w_{t-1}^\intercal\} \qquad (11) \\
&= A_{t-1}P_{t-1|t-1}A_{t-1}^\intercal + Q_{t-1} \qquad (12)
\end{aligned}
$$

This proves Eq. 2.

$$
\begin{aligned}
\hat{\mathbf{y}}_{t|t-1} &= E\{\mathbf{y}_t|Y_{t-1}\} = E\{B_t\mathbf{x}_t + \mathbf{v}_t|Y_{t-1}\} = B_tE\{\mathbf{x}_t|Y_{t-1}\} + E\{\mathbf{v}_t|Y_{t-1}\} \\
&= B_t\mathbf{x}_{t|t-1} + E\{\mathbf{v}_t\} = B_t\mathbf{x}_{t|t-1}
\end{aligned}
$$

This proves Eq. 3.

Because

$$
\mathbf{z}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} = B_t\mathbf{x}_t + \mathbf{v}_t - B_t\mathbf{x}_{t|t-1} = B_t(\mathbf{x}_t - \mathbf{x}_{t|t-1}) + \mathbf{v}_t \qquad (13)
$$

Y_{t-1} and $\mathbf{z}_t$ are fixed if and only if Y_t is fixed[1]. Then

$$
\mathbf{x}_{t|t} = E\{\mathbf{x}_t|Y_t\} = E\{\mathbf{x}_t|Y_{t-1}, \mathbf{z}_t\} \qquad (14)
$$

[1]If we now Y_{t-1} and $\mathbf{z}_t$, then we know $\hat{\mathbf{y}}_{t|t-1}$ and $\mathbf{z}_t$, then (by the first equality in Eq. 13) we know $\mathbf{y}_t$, thus we know Y_t. Also, if we know Y_t, we know $\hat{\mathbf{y}}_{t|t-1}$ and $\mathbf{y}_t$ and (by the first equality in Eq. 13) we know $\mathbf{z}_t$.

$$
\begin{aligned}
&= E\{\mathbf{x}_t|Y_{t-1}\} + \mathrm{Cov}\left(\mathbf{x}_t, \mathbf{z}_t|Y_{t-1}\right)\mathrm{Cov}\left(\mathbf{z}_t|Y_{t-1}\right)^{-1}\mathbf{z}_t && (15)\\
\mathrm{Cov}\left(\mathbf{x}_t, \mathbf{z}_t|Y_{t-1}\right) &= \mathrm{Cov}\left(\mathbf{x}_t, B_t(\mathbf{x}_t - \mathbf{x}_{t|t-1}) + \mathbf{v}_t|Y_{t-1}\right)\\
&= E\{(\mathbf{x}_t - \mathbf{x}_{t|t-1})(B_t(\mathbf{x}_t - \mathbf{x}_{t|t-1}) + \mathbf{v}_t)^\intercal|Y_{t-1}\}\\
&= E\{(\mathbf{x}_t - \mathbf{x}_{t|t-1})(\mathbf{x}_t - \mathbf{x}_{t|t-1})^\intercal|Y_{t-1}\}B_t^\intercal\\
&\quad + E\{(\mathbf{x}_t - \mathbf{x}_{t|t-1})\mathbf{v}_t^\intercal|Y_{t-1}\}\\
&= P_{t|t-1}B_t^\intercal && (16)\\
S_t = \mathrm{Cov}\left(\mathbf{z}_t|Y_{t-1}\right) &= E\{\mathbf{z}_t\mathbf{z}_t^\intercal|Y_{t-1}\}\\
&= E\{\left(B_t(\mathbf{x}_t - \mathbf{x}_{t|t-1}) + \mathbf{v}_t\right)\left(B_t(\mathbf{x}_t - \mathbf{x}_{t|t-1}) + \mathbf{v}_t\right)^\intercal|Y_{t-1}\}\\
&= B_t E\{\left(\mathbf{x}_t - \mathbf{x}_{t|t-1}\right)\left(\mathbf{x}_t - \mathbf{x}_{t|t-1}\right)^\intercal|Y_{t-1}\}B_t^\intercal\\
&\quad + E\{\mathbf{v}_t\mathbf{v_t}^\intercal|Y_{t-1}\}\\
&= B_t P_{t|t-1}B_t^\intercal + R_t && (17)
\end{aligned}
$$

This proves Eq. 4.

Combining Eqs. 15, 16 and 17 we obtain

$$
\begin{aligned}
\mathbf{x}_{t|t} &= \mathbf{x}_{t|t-1} + P_{t|t-1}B_t^\intercal S_t^{-1}\mathbf{z}_t\\
&= \mathbf{x}_{t|t-1} + K_t\mathbf{z}_t \quad \text{with } K_t = P_{t|t-1}B_t^\intercal S_t^{-1}\\
P_{t|t} &= \mathrm{Cov}(\mathbf{x}_t|Y_t) = \mathrm{Cov}(\mathbf{x}_t|Y_{t-1}, \mathbf{z}_t) = P_{t|t-1} - P_{t|t-1}B_t^\intercal S_t^{-1}B_t P_{t|t-1}\\
&= \left(I - P_{t|t-1}B_t^\intercal S_t^{-1}B_t\right)P_{t|t-1} = \left(I - K_t B_t\right)P_{t|t-1} && (18)
\end{aligned}
$$

This proves Eqs. 5, 6 and 7.

Using Eqs. 8 and 9 in Eqs. 1 and 2 we obtain

$$
\begin{aligned}
\mathbf{x}_{1|0} &= A_0\mathbf{x}_{0|0} = A_0\mathbf{m}_0\\
\mathbf{P}_{1|0} &= A_0 P_{0|0}A_0^\intercal + Q_0 = A_0 V_0 A_0^\intercal + Q_0
\end{aligned}
$$

If Eqs. 8 and 9 are correct, then the density of $\mathbf{x}_1$ should be $p(\mathbf{x}_1) = \mathcal{N}(\mathbf{x}_1|\mathbf{x}_{1|0}, P_{1|0})$. We now calculate this density using the linear dynamical system model in Theorem 1.

$$
\begin{aligned}
p(\mathbf{x}_1) &= \int p(\mathbf{x}_1, \mathbf{x}_0)d\mathbf{x}_0 = \int p(\mathbf{x}_1|\mathbf{x}_0)p(\mathbf{x}_0)d\mathbf{x}_0 = \int \mathcal{N}(\mathbf{x}_1|A_0\mathbf{x}_0, Q_0)\mathcal{N}(\mathbf{x}_0|\mathbf{m}_0, V_0)d\mathbf{x}_0\\
&= \mathcal{N}(\mathbf{x}_1|A_0\mathbf{m}_0, A_0V_0A_0^\intercal + Q_0) = \mathcal{N}(\mathbf{x}_1|\mathbf{x}_{1|0}, \mathbf{P}_{1|0}) && (19)
\end{aligned}
$$

This proves Eqs. 8 and 9.

□

Notes:

1. the first equality in Eq. 10 holds because $\mathbf{w}_{t-1}$ is independent of Y_{t-1}.
2. the second and third terms in Eq. 11 hold because w_{t-1} is independent of x_{t-1} given Y_{t-1}.
3. Eq. 12 holds because $E\{\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1}|Y_{t-1}\} = 0$.

4. the validity of Eq. 14 follows from measure theory concepts that I don't really know.

5. the last equality in Eq. 15 follows from Eq. **??** in Lemma 1

6. the last equality in Eq. 16 holds because $\mathbf{x}_t$ in independent (and therefore uncorrelated) of $\mathbf{v}_t$.

7. the second equality of Eq. 17 uses Eq. 13.

8. the third equality of Eq. 17 holds because $\mathbf{x}_t$ is independent of $\mathbf{v}_t$ given Y_{t-1}.

9. the last equality of Eq. 17 holds because $\mathbf{v}_t$ is independent of Y_{t-1}.

10. in the third equality of Eq. 18 we used Eq. 21 in Lemma 1 with $\mathbf{x} = \mathbf{x}_t|Y_{t-1}$ and $\mathbf{y} = \mathbf{z}_t|Y_{t-1}$ giving

$$
\begin{aligned}
\Sigma_{x|y} &= \Sigma_{\mathbf{x}_t|\mathbf{z}_t, Y_{t-1}} = \Sigma_{\mathbf{x}_t|Y_t} = P_{t|t} \\
\Sigma_{xx} &= \Sigma_{\mathbf{x}_t|Y_{t-1}} = P_{t|t-1} \\
\Sigma_{xy} &= \Sigma_{\mathbf{x}_t\mathbf{z}_t|Y_{t-1}} = \mathrm{Cov}(\mathbf{x}_t, \mathbf{z}_t|Y_{t-1}) = P_{t|t-1}B_t^\intercal \\
\Sigma_{yy} &= \Sigma_{\mathbf{z}_t\mathbf{z}_t|Y_{t-1}} = \mathrm{Cov}(\mathbf{z}_t|Y_{t-1}) = S_t \\
&\text{thus} \\
P_{t|t} &= P_{t|t-1} - P_{t|t-1}B_t^\intercal S_t^{-1} B_t P_{t|t-1}
\end{aligned}
$$

11. in the fourth equality of Eq. 19 we used Lemma **??**.

Lemma 1. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be jointly Gaussian distributed random vectors with*

$$E\left\{\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\right\} = \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix} \tag{20}$$

$$Cov\left\{\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\right\} = \begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix} \tag{21}$$

where Σ_{yy} *is assumed to be non-singular. Then the conditional distribution of* $\mathbf{x}$ *given* $\mathbf{y}$ *is Gaussian with mean vector*

$$E\{\mathbf{x}|\mathbf{y}\} = \boldsymbol{\mu}_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$$

and covariance matrix

$$Cov\{\mathbf{x}|\mathbf{y}\} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$$

Proof. Let

$$\mathbf{z} = \mathbf{x} - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) \tag{22}$$

Since $(\mathbf{x}, \mathbf{y})$ are jointly Gaussian, and $(\mathbf{z}, \mathbf{y})$ is an affine transformation of $(\mathbf{x}, \mathbf{y})$, then $(\mathbf{z}, \mathbf{y})$ are jointly Gaussian.

We have

$$\begin{aligned}
\mathrm{E}\{\mathbf{z}\} &= E\{\mathbf{x}\} = \boldsymbol{\mu}_x\\
\mathbf{z} - \boldsymbol{\mu}_z &= \left(\mathbf{x} - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right) - \boldsymbol{\mu}_x = (\mathbf{x} - \boldsymbol{\mu}_x) - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\\
\mathrm{Cov}\{\mathbf{z}\} &= \mathrm{E}\{(\mathbf{z} - \boldsymbol{\mu}_z)(\mathbf{z} - \boldsymbol{\mu}_z)^\intercal\}\\
&= \mathrm{E}\{\left[(\mathbf{x} - \boldsymbol{\mu}_x) - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right]\left[(\mathbf{x} - \boldsymbol{\mu}_x) - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right]^\intercal\}\\
&= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} + \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy}\Sigma_{yy}^{-1}\Sigma_{yx}\\
&= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\\
\mathrm{Cov}\{\mathbf{y}, \mathbf{z}\} &= \mathrm{E}\{(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{z} - \boldsymbol{\mu}_z)^\intercal\}\\
&= \mathbf{E}\{(\mathbf{y} - \boldsymbol{\mu}_y)((\mathbf{x} - \boldsymbol{\mu}_x) - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y))^\intercal\}\\
&= \Sigma_{yx} - \Sigma_{yy}\Sigma_{yy}^{-1}\Sigma_{yx} = 0
\end{aligned} \tag{23}$$

Because $(\mathbf{y}, \mathbf{z})$ are uncorrelated (Eq. 23) and jointly Gaussian, they are independent. Thus, $\mathrm{E}\{\mathbf{z}|\mathbf{y}\} = \mathrm{E}\{\mathbf{z}\}$ and $\mathrm{Cov}\{\mathbf{z}|\mathbf{y}\} = \mathrm{Cov}\{\mathbf{z}\}$.

From Eq. 22, $\mathbf{x} = \mathbf{z} + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$. Then

$$\begin{aligned}
\mathrm{E}\{\mathbf{x}|\mathbf{y}\} &= \mathrm{E}\{\mathbf{z}|\mathbf{y}\} + \mathrm{E}\{\Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)|\mathbf{y}\}\\
&= \mathrm{E}\{\mathbf{z}\} + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)
\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{\mu}_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) \\
\mathrm{Cov}\{\mathbf{x}|\mathbf{y}\} &= \mathrm{Cov}\{\mathbf{z} + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)|\mathbf{y}\} = \mathrm{Cov}\{\mathbf{z}|\mathbf{y}\} \\
&= \mathrm{Cov}\{\mathbf{z}\} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}
\end{aligned} \tag{24}
$$

□

Notes:

1. The last equality in Eq. 24 holds because, when conditioning on $\mathbf{y}$, the term $\Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$ is a constant, and constants are irrelevant when computing covariances.

Lemma 2. *Let*

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|A\mathbf{x}+\mathbf{b}, \Sigma) \tag{25}$$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \Lambda) \tag{26}$$

then

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|A\boldsymbol{\mu}+\mathbf{b}, A\Lambda A^\intercal + \Sigma) \tag{27}$$

Proof.

$$\begin{aligned}
\ln p(\mathbf{x},\mathbf{y}) &= \ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x}) \\
&= -\frac{1}{2}\left(\mathbf{y}-(A\mathbf{x}+\mathbf{b})\right)^\intercal \Sigma^{-1}\left(\mathbf{y}-(A\mathbf{x}+\mathbf{b})\right) - \frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)^\intercal \Lambda^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right) + K_1 \\
&= -\frac{1}{2}\mathbf{y}^\intercal\Sigma^{-1}\mathbf{y} + \frac{1}{2}\mathbf{y}^\intercal\Sigma^{-1}A\mathbf{x} + \frac{1}{2}\mathbf{x}^\intercal A^\intercal\Sigma^{-1}\mathbf{y} - \frac{1}{2}\mathbf{x}^\intercal\left(A^\intercal\Sigma^{-1}A+\Lambda^{-1}\right)\mathbf{x} \\
&\quad + \frac{1}{2}\mathbf{y}^\intercal\Lambda^{-1}\mathbf{b} + \frac{1}{2}\mathbf{x}^\intercal(-A^\intercal\Sigma^{-1}\mathbf{b}+\Lambda\boldsymbol{\mu}) + \frac{1}{2}\mathbf{b}^\intercal\Lambda^{-1}\mathbf{y} + \frac{1}{2}(-\mathbf{b}^\intercal\Sigma^{-1}A+\boldsymbol{\mu}^\intercal\Lambda)\mathbf{x} + K_2 \\
&= -\frac{1}{2}[\mathbf{x}^\intercal,\mathbf{y}^\intercal]\begin{bmatrix} A^\intercal\Sigma^{-1}A+\Lambda^{-1} & -A^\intercal\Sigma^{-1} \\ -\Sigma^{-1}A & \Sigma^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \\
&\quad + \frac{1}{2}[\mathbf{x}^\intercal,\mathbf{y}^\intercal]\begin{bmatrix} A^\intercal\Sigma^{-1}\mathbf{b}+\Lambda^{-1}\boldsymbol{\mu} \\ -\Sigma^{-1}\mathbf{b}\end{bmatrix} + \frac{1}{2}[\mathbf{b}^\intercal\Sigma^{-1}A+\boldsymbol{\mu}^\intercal\Lambda^{-1}, -\mathbf{b}^\intercal\Sigma^{-1}]\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} + K_2
\end{aligned} \tag{28}$$

where K_1 and K_2 are contants that does not depend on $\mathbf{x}$ or $\mathbf{y}$.

Because $\ln p(\mathbf{x},\mathbf{y})$ is a quadratic form, then $p(\mathbf{x},\mathbf{y})$ is a normal probability density function (pdf), thus its marginal $p(\mathbf{y})$ is also a normal pdf. Our aim is to derive the mean and covariance of $\mathbf{y}$, $\boldsymbol{\mu}_y$ and Γ_{yy}, respectively.

Call

$$p(\mathbf{x},\mathbf{y}) = \mathcal{N}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\middle|\middle|\begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}, \Gamma\right) \tag{29}$$

with

$$\Phi^{-1} = \Gamma = \begin{bmatrix}\Gamma_{xx} & \Gamma_{xy}\\ \Gamma_{yx} & \Gamma_{yy}\end{bmatrix}$$

Next,

$$\begin{aligned}
\ln p(\mathbf{x},\mathbf{y}) &= -\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}_x)^\intercal, (\mathbf{y}-\boldsymbol{\mu}_y)^\intercal\right]\Phi\left[(\mathbf{x}-\boldsymbol{\mu}_x), (\mathbf{y}-\boldsymbol{\mu}_y))\right] + K_1 \\
&= -\frac{1}{2}[\mathbf{x}^\intercal,\mathbf{y}^\intercal]\Phi\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} + \frac{1}{2}[\mathbf{x}^\intercal,\mathbf{y}^\intercal]\Phi\begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix} + \frac{1}{2}[\boldsymbol{\mu}_x^\intercal,\boldsymbol{\mu}_y^\intercal]\Phi\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} + K_2
\end{aligned} \tag{30}$$

where K_1 and K_2 are contants that does not depend on $\mathbf{x}$ or $\mathbf{y}$.

From Eqs. 28 and 30 it follows that

$$\Phi = \begin{bmatrix} A^\intercal \Sigma^{-1} A + \Lambda^{-1} & -A^\intercal \Sigma^{-1} \\ -\Sigma^{-1} A & \Sigma^{-1} \end{bmatrix}$$

and

$$\Phi \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} = \begin{bmatrix} A^\intercal \Sigma^{-1} \mathbf{b} + \Lambda^{-1} \boldsymbol{\mu} \\ -\Sigma^{-1} \mathbf{b} \end{bmatrix}$$

Then

$$\Gamma = \begin{bmatrix} \Gamma_{xx} & \Gamma_{xy} \\ \Gamma_{yx} & \Gamma_{yy} \end{bmatrix} = \Phi^{-1} = \begin{bmatrix} A^\intercal \Sigma^{-1} A + \Lambda^{-1} & -A^\intercal \Sigma^{-1} \\ -\Sigma^{-1} A & \Sigma^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda & \Lambda A \\ A\Lambda & \Sigma + A\Lambda A^\intercal \end{bmatrix} \quad (31)$$

$$\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix} = \Phi^{-1} \begin{bmatrix} A^\intercal \Sigma^{-1} \mathbf{b} + \Lambda^{-1} \boldsymbol{\mu} \\ -\Sigma^{-1} \mathbf{b} \end{bmatrix} = \Gamma \begin{bmatrix} A^\intercal \Sigma^{-1} \mathbf{b} + \Lambda^{-1} \boldsymbol{\mu} \\ -\Sigma^{-1} \mathbf{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ A\boldsymbol{\mu} + \mathbf{b} \end{bmatrix}$$

Thus,

$$\Gamma_{yy} = \Sigma + A\Lambda A^\intercal$$
$$\boldsymbol{\mu}_y = A\boldsymbol{\mu} + \mathbf{b}$$

□

Note:

1. The last equality in Eq. 31 follows from Lemma **??**.

Lemma 3.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & -D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix}$$

$$\textit{with}$$

$$M = (A - BD^{-1}C)^{-1}$$

Proof.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & -D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix} =$$

$$\begin{bmatrix} AM - BD^{-1}CM & -AMBD^{-1} + BD^{-1} + BD^{-1}CMBD^{-1} \\ CM - CM & -CMBD^{-1} + I + CMBD^{-1} \end{bmatrix} =$$

$$\begin{bmatrix} (A - BD^{-1}C)M & (-A + M^{-1} + BD^{-1}C)MBD^{-1} \\ 0 & I \end{bmatrix} =$$

$$\begin{bmatrix} M^{-1}M & (-M^{-1} + M^{-1})MBD^{-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

$$\begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & -D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} =$$

$$\begin{bmatrix} MA - MBD^{-1}C & MB - MB \\ -D^{-1}CMA - D^{-1}C + D^{-1}CMBD^{-1}C & -D^{-1}CMB + I - D^{-1}CMB \end{bmatrix} =$$

$$\begin{bmatrix} M(A - BD^{-1}C) & 0 \\ -D^{-1}CM(A - BD^{-1}C) + D^{-1}C & I \end{bmatrix} =$$

$$\begin{bmatrix} MM^{-1} & 0 \\ -D^{-1}CMM^{-1} + D^{-1}C & I \end{bmatrix} =$$

$$\begin{bmatrix} I & 0 \\ -D^{-1}C + D^{-1}C & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

□

References

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38. OUP Oxford.