Inference in Linear Dynamical Systems

Joaquin Rapela[*]

June 27, 2025

1 Linear dynamical systems (LDS) model

$$
\begin{aligned}
\mathbf{x}_{n+1} &= A_n\mathbf{x}_n + \mathbf{w}_n \quad \text{with } \mathbf{w}_n \sim N(0, Q_n)\\
\mathbf{y}_n &= B_n\mathbf{x}_n + \mathbf{v}_n \quad \text{with } \mathbf{v}_n \sim N(0, R_n)\\
\mathbf{x}_0 &\sim N(\mathbf{m}_0, V_0)
\end{aligned}
$$

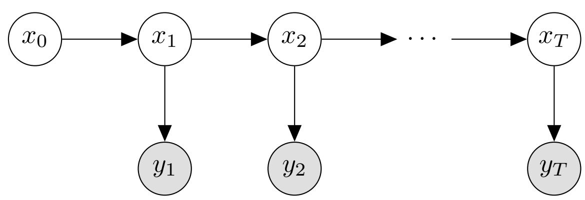


Figure 1: Graphical models for the linear dynamical system

2 Inference Problems

Prediction

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

Filtering

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

Smoothing

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

Forecasting

$$P(\mathbf{x}_{n+h}|\mathbf{y}_1,\ldots,\mathbf{y}_n) = N(\mathbf{x}_n|\mathbf{x}_{n+h|n}, P_{n+h|n}) \tag{4}$$

[*]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.

Inference of predicted and filtered means 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.

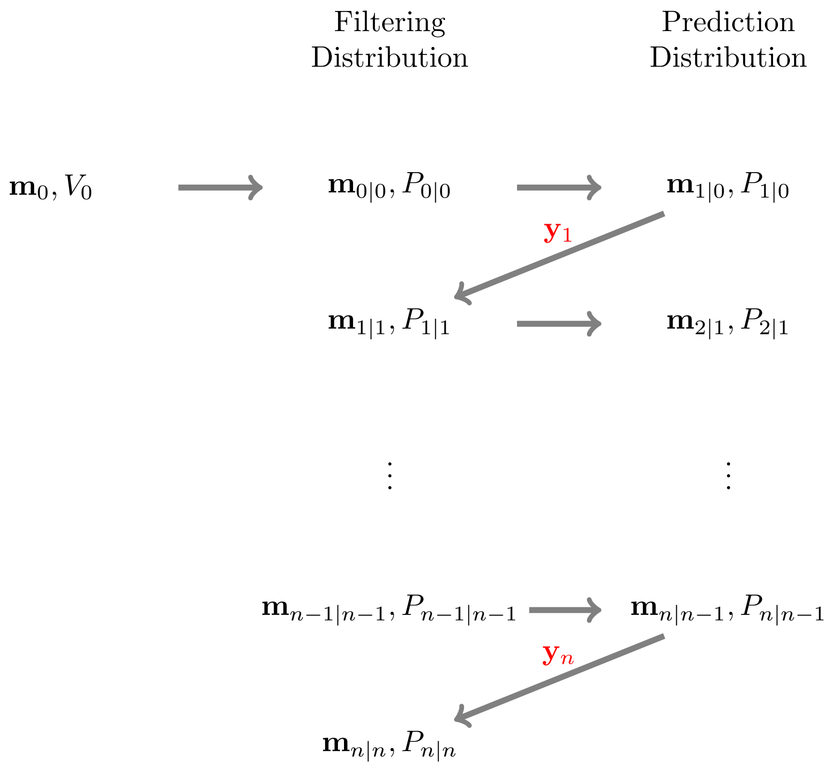


Figure 2: Order of calculation of the prediction and filtering distributions in the Kalman filtering algorithm.

4 Derivation of Kalman filter equations

Theorem 1. *Given the linear dynamical systems model given in Section 1 then the means and covariances of the prediction (Eq. 1) and filtering (Eq. 2) distributions are calculated iteratively as follows:*

$$
\begin{aligned}
\mathbf{x}_{0|0} &= \mathbf{m}_0 && \textit{init filtered mean} && (5)\\
P_{0|0} &= V_0 && \textit{init filtered covariance} && (6)\\
\mathbf{x}_{n|n-1} &= A_{n-1}\mathbf{x}_{n-1|n-1} && \textit{prediction mean} && (7)\\
P_{n|n-1} &= A_{n-1}P_{n-1|n-1}A_{n-1}^{\intercal} + Q_{n-1} && \textit{prediction covariance} && (8)\\
\hat{\mathbf{y}}_{n|n-1} &\triangleq E\{\mathbf{y}_n|\mathbf{y}_1,\ldots,\mathbf{y}_{n-1}\} = B_n\mathbf{x}_{n|n-1} && \textit{predicted observation} && (9)\\
\mathbf{z}_n &\triangleq \mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1} && \textit{residual} &&\\
S_n &\triangleq Cov\{\mathbf{z}_n|\mathbf{y}_1,\ldots,\mathbf{y}_{n-1}\} = B_nP_{n|n-1}B_n^{\intercal} + R_n && \textit{residual covariance} && (10)\\
\mathbf{K}_n &= P_{n|n-1}B_n^{\intercal}S_n^{-1} && \textit{Kalman gain} && (11)\\
\mathbf{x}_{n|n} &= \mathbf{x}_{n|n-1} + K_n\mathbf{z}_n && \textit{filtering mean} && (12)\\
\mathbf{P}_{n|n} &= (I - K_nB_n)P_{n|n-1} && \textit{filtering covariance} && (13)
\end{aligned}
$$

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

Proof. Call $Y_n = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$, then

$$
\begin{aligned}
\mathbf{x}_{n|n-1} &= E\{\mathbf{x}_n|Y_{n-1}\} = E\{A_{n-1}\mathbf{x}_{n-1} + \mathbf{w}_{n-1}|Y_{n-1}\}\\
&= A_{n-1}E\{\mathbf{x}_{n-1}|Y_{n-1}\} + E\{\mathbf{w}_{n-1}|Y_{n-1}\}\\
&= A_{n-1}\mathbf{x}_{n-1|n-1} + E\{\mathbf{w}_{n-1}\}^1 = A_{n-1}\mathbf{x}_{n-1|n-1}
\end{aligned}
$$

This proves Eq. 7.

$$
\begin{aligned}
\mathbf{P}_{n|n-1} &= \mathrm{Cov}\{\mathbf{x}_n|Y_{n-1}\} = E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1})(\mathbf{x}_n - \mathbf{x}_{n|n-1})^{\intercal}|Y_{n-1}\}\\
&= E\{\left(A_{n-1}\mathbf{x}_{n-1} + w_{n-1} - A_{n-1}\mathbf{x}_{n-1|n-1}\right)\left(A_{n-1}\mathbf{x}_{n-1} + w_{n-1} - A_{n-1}\mathbf{x}_{n-1|n-1}\right)^{\intercal}|Y_{n-1}\}\\
&= E\{\left(A_{n-1}(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1}) + w_{n-1}\right)\left(A_{n-1}(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1}) + w_{n-1}\right)^{\intercal}|Y_{n-1}\}\\
&= A_{n-1}E\{(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1})(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1})^{\intercal}|Y_{n-1}\}A_{n-1}^{\intercal}+\\
&\quad E\{w_{n-1}(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1})^{\intercal}|Y_{n-1}\}A_{n-1}^{\intercal}+\\
&\quad A_{n-1}E\{(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1})w_{n-1}^{\intercal}|Y_{n-1}\}+\\
&\quad E\{w_{n-1}w_{n-1}^{\intercal}|Y_{n-1}\}\\
&= A_{n-1}E\{(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1})(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1})^{\intercal}|Y_{n-1}\}A_{n-1}^{\intercal}+\\
&\quad E\{w_{n-1}|Y_{n-1}\}E\{(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1})^{\intercal}|Y_{n-1}\}A_{n-1}^{\intercal}{}^2+\\
&\quad A_{n-1}E\{(\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1})|Y_{n-1}\}E\{w_{n-1}^{\intercal}|Y_{n-1}\}+\\
&\quad E\{w_{n-1}w_{n-1}^{\intercal}\}\\
&= A_{n-1}P_{n-1|n-1}A_{n-1}^{\intercal} + Q_{n-1}{}^3
\end{aligned}
$$

This proves Eq. 8.

---

[1] $\mathbf{w}_{n-1}$ is independent of Y_{n-1}.

[2] $\mathbf{w}_{n-1}$ is independent of x_{n-1} given Y_{n-1}.

[3] $E\{\mathbf{x}_{n-1} - \mathbf{x}_{n-1|n-1}|Y_{n-1}\} = E\{\mathbf{x}_{n-1}|Y_{n-1}\} - \mathbf{x}_{n-1|n-1} = \mathbf{x}_{n-1|n-1} - \mathbf{x}_{n-1|n-1} = 0$.

$$
\begin{aligned}
\hat{\mathbf{y}}_{n|n-1} &= E\{\mathbf{y}_n|Y_{n-1}\} = E\{B_n\mathbf{x}_n + \mathbf{v}_n|Y_{n-1}\} = B_nE\{\mathbf{x}_n|Y_{n-1}\} + E\{\mathbf{v}_n|Y_{n-1}\} \\
&= B_n\mathbf{x}_{n|n-1} + E\{\mathbf{v}_n\} = B_n\mathbf{x}_{n|n-1}
\end{aligned}
$$

This proves Eq. 9.

Because

$$
\mathbf{z}_n = \mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1} = B_n\mathbf{x}_n + \mathbf{v}_n - B_n\mathbf{x}_{n|n-1} = B_n(\mathbf{x}_n - \mathbf{x}_{n|n-1}) + \mathbf{v}_n \tag{14}
$$

Y_{n-1} and $\mathbf{z}_n$ are fixed if and only if Y_n is fixed[4]. Then

.

$$
\begin{aligned}
\mathbf{x}_{n|n} &= E\{\mathbf{x}_n|Y_n\} = E\{\mathbf{x}_n|Y_{n-1},\mathbf{z}_n\}^5 \\
&= E\{\mathbf{x}_n|Y_{n-1}\} + \operatorname{Cov}(\mathbf{x}_n,\mathbf{z}_n|Y_{n-1})\operatorname{Cov}(\mathbf{z}_n|Y_{n-1})^{-1}\mathbf{z}_n{}^6 \qquad (15)\\
\operatorname{Cov}(\mathbf{x}_n,\mathbf{z}_n|Y_{n-1}) &= \operatorname{Cov}\left(\mathbf{x}_n, B_n(\mathbf{x}_n - \mathbf{x}_{n|n-1}) + \mathbf{v}_n|Y_{n-1}\right) \\
&= E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1})(B_n(\mathbf{x}_n - \mathbf{x}_{n|n-1}) + \mathbf{v}_n)^\intercal|Y_{n-1}\} \\
&= E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1})(\mathbf{x}_n - \mathbf{x}_{n|n-1})^\intercal|Y_{n-1}\}B_n^\intercal \\
&\quad + E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1})\mathbf{v}_n^\intercal|Y_{n-1}\} \\
&= P_{n|n-1}B_n^{\intercal 7} \qquad (16)\\
S_n = \operatorname{Cov}(\mathbf{z}_n|Y_{n-1}) &= E\{\mathbf{z}_n\mathbf{z}_n^\intercal|Y_{n-1}\} \\
&= E\{\left(B_n(\mathbf{x}_n - \mathbf{x}_{n|n-1}) + \mathbf{v}_n\right)\left(B_n(\mathbf{x}_n - \mathbf{x}_{n|n-1}) + \mathbf{v}_n\right)^\intercal|Y_{n-1}\}^8 \\
&= B_nE\{\left(\mathbf{x}_n - \mathbf{x}_{n|n-1}\right)\left(\mathbf{x}_n - \mathbf{x}_{n|n-1}\right)^\intercal|Y_{n-1}\}B_n^\intercal + B_nE\{\left(\mathbf{x}_n - \mathbf{x}_{n|n-1}\right)\mathbf{v}_n^\intercal|Y_{n-1}\} \\
&\quad + E\{\mathbf{v}_n\left(\mathbf{x}_n - \mathbf{x}_{n|n-1}\right)^\intercal\}B_n^\intercal + E\{\mathbf{v}_n\mathbf{v_n}^\intercal|Y_{n-1}\} \\
&= B_nE\{\left(\mathbf{x}_n - \mathbf{x}_{n|n-1}\right)\left(\mathbf{x}_n - \mathbf{x}_{n|n-1}\right)^\intercal|Y_{n-1}\}B_n^\intercal + \\
&\quad + E\{\mathbf{v}_n\mathbf{v_n}^\intercal\}^{79} \\
&= B_nP_{n|n-1}B_n^\intercal + R_n \qquad (17)
\end{aligned}
$$

This proves Eq. 10.

Combining Eqs. 15, 16 and 17 we obtain

$$
\begin{aligned}
\mathbf{x}_{n|n} &= \mathbf{x}_{n|n-1} + P_{n|n-1}B_n^\intercal S_n^{-1}\mathbf{z}_n \\
&= \mathbf{x}_{n|n-1} + K_n\mathbf{z}_n \quad \text{with } K_n = P_{n|n-1}B_n^\intercal S_n^{-1} \\
P_{n|n} &= \operatorname{Cov}(\mathbf{x}_n|Y_n) = \operatorname{Cov}(\mathbf{x}_n|Y_{n-1},\mathbf{z}_n) = P_{n|n-1} - P_{n|n-1}B_n^\intercal S_n^{-1}B_nP_{n|n-1}{}^{10} \\
&= \left(I - P_{n|n-1}B_n^\intercal S_n^{-1}B_n\right)P_{n|n-1} = \left(I - K_nB_n\right)P_{n|n-1}
\end{aligned}
$$

[4]If we now Y_{n-1} and $\mathbf{z}_n$, then we know $\hat{\mathbf{y}}_{n|n-1}$ and $\mathbf{z}_n$, then (by the first equality in Eq. 14) we know $\mathbf{y}_n$, thus we know Y_n. Also, if we know Y_n, we know $\hat{\mathbf{y}}_{n|n-1}$ and $\mathbf{y}_n$ and (by the first equality in Eq. 14) we know $\mathbf{z}_n$.

[5]The validity of the last equality follow from measure theory arguments (that I don't know).

[6]Refer to Eq. 18 in Lemma 1.

[7]$E\{(\mathbf{x}_n - \mathbf{x}_{n|n-1})\mathbf{v}_n^\intercal|Y_{n-1}\} = E\{\mathbf{x}_n - \mathbf{x}_{n|n-1}|Y_{n-1}\}E\{\mathbf{v}_n^\intercal|Y_{n-1}\} = (E\{\mathbf{x}_n|Y_{n-1}\} - \mathbf{x}_{n|n-1})E\{\mathbf{v}_n^\intercal|Y_{n-1}\} = (\mathbf{x}_{n|n-1} - \mathbf{x}_{n|n-1})E\{\mathbf{v}_n^\intercal|Y_{n-1}\} = 0$ because $\mathbf{x}_n$ is independent of $\mathbf{v}_n$ given Y_{n-1}.

[8]Eq. 14.

[9]$\mathbf{v}_n$ is independent from Y_{n-1}.

This proves Eqs. 11, 12 and 13.

Using Eqs. 5 and 6 in Eqs. 7 and 8 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. 5 and 6 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.

$$
\begin{aligned}
p(\mathbf{x}_1) &= \int p(\mathbf{x}_1, \mathbf{x}_0) d\mathbf{x}_0 = \int p(\mathbf{x}_1 | \mathbf{x}_0) p(\mathbf{x}_0) d\mathbf{x}_0 = \int \mathcal{N}(\mathbf{x}_1 | A_0 \mathbf{x}_0, Q_0) \mathcal{N}(\mathbf{x}_0 | \mathbf{m}_0, V_0) d\mathbf{x}_0 \\
&= \mathcal{N}(\mathbf{x}_1 | A_0 \mathbf{m}_0, A_0 V_0 A_0^{\intercal} + Q_0)^{11} = \mathcal{N}(\mathbf{x}_1 | \mathbf{x}_{1|0}, \mathbf{P}_{1|0})
\end{aligned}
$$

This proves Eqs. 5 and 6.

$\square$

---

[10]Refer to Eq. 19 in Lemma 1 with $\mathbf{x} = \mathbf{x}_n | Y_{n-1}$ and $\mathbf{y} = \mathbf{z}_n | Y_{n-1}$ giving

$$
\begin{aligned}
\Sigma_{x|y} &= \Sigma_{\mathbf{x}_n | \mathbf{z}_n, Y_{n-1}} = \Sigma_{\mathbf{x}_n | Y_n} = P_{n|n} \\
\Sigma_{xx} &= \Sigma_{\mathbf{x}_n | Y_{n-1}} = P_{n|n-1} \\
\Sigma_{xy} &= \Sigma_{\mathbf{x}_n \mathbf{z}_n | Y_{n-1}} = \mathrm{Cov}(\mathbf{x}_n, \mathbf{z}_n | Y_{n-1}) = P_{n|n-1} B_n^{\intercal} \\
\Sigma_{yy} &= \Sigma_{\mathbf{z}_n \mathbf{z}_n | Y_{n-1}} = \mathrm{Cov}(\mathbf{z}_n | Y_{n-1}) = S_n \\
&\text{thus} \\
P_{n|n} &= P_{n|n-1} - P_{n|n-1} B_n^{\intercal} S_n^{-1} B_n P_{n|n-1}
\end{aligned}
$$

[11]Lemma 2.

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

$$E\left\{\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\right\} = \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}$$
$$Cov\left\{\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\right\} = \begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix}$$

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

$$E\{\mathbf{x}|\mathbf{y}\} = \boldsymbol{\mu}_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) \tag{18}$$

and covariance matrix

$$Cov\{\mathbf{x}|\mathbf{y}\} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \tag{19}$$

Proof. Let

$$\mathbf{z} = \mathbf{x} - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) \tag{20}$$

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

$$\begin{aligned}
\mathrm{E}\{\mathbf{z}\} &= E\{\mathbf{x}\} = \boldsymbol{\mu}_x \\
\mathbf{z} - \boldsymbol{\mu}_z &= \left(\mathbf{x} - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right) - \boldsymbol{\mu}_x = (\mathbf{x} - \boldsymbol{\mu}_x) - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) \\
\mathrm{Cov}\{\mathbf{z}\} &= \mathrm{E}\{(\mathbf{z} - \boldsymbol{\mu}_z)(\mathbf{z} - \boldsymbol{\mu}_z)^\intercal\} \\
&= \mathrm{E}\{\left[(\mathbf{x} - \boldsymbol{\mu}_x) - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right]\left[(\mathbf{x} - \boldsymbol{\mu}_x) - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right]^\intercal\} \\
&= \mathrm{E}\{(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^\intercal\} - \mathrm{E}\{(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{y} - \boldsymbol{\mu}_y)^\intercal\}\Sigma_{yy}^{-1}\Sigma_{xy} - \Sigma_{xy}\Sigma_{yy}^{-1}\mathrm{E}\{(\mathbf{y} - \boldsymbol{\mu}_y))(\mathbf{x} - \boldsymbol{\mu}_x)^\intercal\} + \\
&\quad \Sigma_{xy}\Sigma_{yy}^{-1}\mathrm{E}\{(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)^\intercal\}\Sigma_{yy}^{-1}\Sigma_{xy} \\
&= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} + \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy}\Sigma_{yy}^{-1}\Sigma_{yx} \\
&= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \\
\mathrm{Cov}\{\mathbf{y}, \mathbf{z}\} &= \mathrm{E}\{(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{z} - \boldsymbol{\mu}_z)^\intercal\} \\
&= \mathbf{E}\{(\mathbf{y} - \boldsymbol{\mu}_y)((\mathbf{x} - \boldsymbol{\mu}_x) - \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y))^\intercal\} \\
&= \Sigma_{yx} - \Sigma_{yy}\Sigma_{yy}^{-1}\Sigma_{yx} = 0
\end{aligned} \tag{21}$$

Because $(\mathbf{y}, \mathbf{z})$ are uncorrelated (Eq. 21) 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. 20, $\mathbf{x} = \mathbf{z} + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$. Then

$$
\begin{aligned}
\mathrm{E}\{\mathbf{x}|\mathbf{y}\} &= \mathrm{E}\{\mathbf{z}|\mathbf{y}\} + \mathrm{E}\{\Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)|\mathbf{y}\} \\
&= \mathrm{E}\{\mathbf{z}\} + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y) \\
&= \boldsymbol{\mu}_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y) \\
\mathrm{Cov}\{\mathbf{x}|\mathbf{y}\} &= \mathrm{Cov}\{\mathbf{z} + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)|\mathbf{y}\} = \mathrm{Cov}\{\mathbf{z}|\mathbf{y}\}^{12} \\
&= \mathrm{Cov}\{\mathbf{z}\} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}
\end{aligned}
\tag{22}
$$

□

[12] 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*

$$
\begin{aligned}
p(\mathbf{y}|\mathbf{x}) &= \mathcal{N}(\mathbf{y}|A\mathbf{x}+\mathbf{b},\Sigma)\\
p(\mathbf{x}) &= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\Lambda)
\end{aligned}
$$

then

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

Proof.

$$
\begin{aligned}
\ln p(\mathbf{x},\mathbf{y}) &= \ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x})\\
&= -\frac{1}{2}\left(\mathbf{y}-(A\mathbf{x}+\mathbf{b})\right)^\intercal \Sigma^{-1}\left(\mathbf{y}-(A\mathbf{x}+\mathbf{b})\right) - \frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)^\intercal \Lambda^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right) + K_1\\
&= -\frac{1}{2}\mathbf{y}^\intercal\Sigma^{-1}\mathbf{y} + \frac{1}{2}\mathbf{y}^\intercal\Sigma^{-1}A\mathbf{x} + \frac{1}{2}\mathbf{x}^\intercal A^\intercal\Sigma^{-1}\mathbf{y} - \frac{1}{2}\mathbf{x}^\intercal\left(A^\intercal\Sigma^{-1}A+\Lambda^{-1}\right)\mathbf{x}\\
&\quad + \frac{1}{2}\mathbf{y}^\intercal\Sigma^{-1}\mathbf{b} + \frac{1}{2}\mathbf{x}^\intercal(-A^\intercal\Sigma^{-1}\mathbf{b}+\Lambda\boldsymbol{\mu}) + \frac{1}{2}\mathbf{b}^\intercal\Sigma^{-1}\mathbf{y} + \frac{1}{2}(-\mathbf{b}^\intercal\Sigma^{-1}A+\boldsymbol{\mu}^\intercal\Lambda)\mathbf{x} + K_2\\
&= -\frac{1}{2}[\mathbf{x}^\intercal,\mathbf{y}^\intercal]\begin{bmatrix} A^\intercal\Sigma^{-1}A+\Lambda^{-1} & -A^\intercal\Sigma^{-1}\\ -\Sigma^{-1}A & \Sigma^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\\
&\quad + \frac{1}{2}[\mathbf{x}^\intercal,\mathbf{y}^\intercal]\begin{bmatrix} -A^\intercal\Sigma^{-1}\mathbf{b}+\Lambda^{-1}\boldsymbol{\mu}\\ \Sigma^{-1}\mathbf{b}\end{bmatrix} + \frac{1}{2}[-\mathbf{b}^\intercal\Sigma^{-1}A+\boldsymbol{\mu}^\intercal\Lambda^{-1},\mathbf{b}^\intercal\Sigma^{-1}]\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} + K_2 \qquad (23)
\end{aligned}
$$

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

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

with

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

Next,

$$
\begin{aligned}
\ln p(\mathbf{x},\mathbf{y}) &= -\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}_x)^\intercal,(\mathbf{y}-\boldsymbol{\mu}_y)^\intercal\right]\Phi\left[(\mathbf{x}-\boldsymbol{\mu}_x),(\mathbf{y}-\boldsymbol{\mu}_y))\right] + K_1\\
&= -\frac{1}{2}[\mathbf{x}^\intercal,\mathbf{y}^\intercal]\Phi\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} + \frac{1}{2}[\mathbf{x}^\intercal,\mathbf{y}^\intercal]\Phi\begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix} + \frac{1}{2}[\boldsymbol{\mu}_x^\intercal,\boldsymbol{\mu}_y^\intercal]\Phi\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} + K_2 \qquad (24)
\end{aligned}
$$

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

From Eqs. 23 and 24 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

$$\Gamma = \begin{bmatrix} \Gamma_{xx} & \Gamma_{xy} \\ \Gamma_{yx} & \Gamma_{yy} \end{bmatrix} = \Phi^{-1} = \begin{bmatrix} A^\intercal \Sigma^{-1} A + \Lambda^{-1} & -A^\intercal \Sigma^{-1} \\ -\Sigma^{-1} A & \Sigma^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda & \Lambda A \\ A\Lambda & \Sigma + A \Lambda A^\intercal \end{bmatrix}^{13} \tag{25}$$

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

Thus,

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

□

[13] Lemma 3.

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.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix} =$$
$$\begin{bmatrix} AM - BD^{-1}CM & -AMBD^{-1} + BD^{-1} + BD^{-1}CMBD^{-1} \\ CM - CM & -CMBD^{-1} + I + CMBD^{-1} \end{bmatrix} =$$
$$\begin{bmatrix} (A - BD^{-1}C)M & (-A + M^{-1} + BD^{-1}C)MBD^{-1} \\ 0 & I \end{bmatrix} =$$
$$\begin{bmatrix} M^{-1}M & (-M^{-1} + M^{-1})MBD^{-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$
$$\begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} =$$
$$\begin{bmatrix} MA - MBD^{-1}C & MB - MB \\ -D^{-1}CMA + D^{-1}C + D^{-1}CMBD^{-1}C & -D^{-1}CMB + I - D^{-1}CMB \end{bmatrix} =$$
$$\begin{bmatrix} M(A - BD^{-1}C) & 0 \\ -D^{-1}CM(A - BD^{-1}C) + D^{-1}C & I \end{bmatrix} =$$
$$\begin{bmatrix} MM^{-1} & 0 \\ -D^{-1}CMM^{-1} + D^{-1}C & I \end{bmatrix} =$$
$$\begin{bmatrix} I & 0 \\ -D^{-1}C + D^{-1}C & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

□

5 Joint normality of the states and observations in the LDS

Lemma 4. *The state and observation random variables of an LDS* $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{y}_1, \ldots, \mathbf{y}_N\}$ *are jointly normal.*

Proof. Note that a set of real random variables $\mathcal{Z} = \{z_1, \ldots, z_N\}$ is jointly normal if and only if their joint probability distribution is a multivariate Normal distribution, if and only if the logarithm of this joint probability distribution is a quadratic function of the random variables in $\mathcal{Z}$ (i.e., $\ln P(z_1, \ldots, z_N) = k + \sum_{i=1}^N k_1(i) z_i + \sum_{i=1}^N \sum_{j=1}^N k_2(i,j) z_i z_j$, with $k_1(i)$ and $k_2(i,j)$ real numbers).

Thus, to prove this lemma it suffice to show that property P_n: "$\log P(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_n)$ is a quadratic function of the components of $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_n\}$" holds for any positive integer n. We show this by induction.

P_1 :

$$\begin{aligned}
\ln P(\mathbf{x}_0, \mathbf{x}_1, \mathbf{y}_1) =& \ln P(\mathbf{y}_1|\mathbf{x}_1) \ln P(\mathbf{x}_1|\mathbf{x}_0) \ln P(\mathbf{x}_0) \\
=& K - \frac{1}{2}(\mathbf{y}_1 - C\mathbf{x}_1)^\intercal R^{-1}(\mathbf{y}_1 - C\mathbf{x}_1) \\
& - \frac{1}{2}(\mathbf{x}_1 - A\mathbf{x}_0)^\intercal Q^{-1}(\mathbf{x}_1 - A\mathbf{x}_0) \\
& - \frac{1}{2}(\mathbf{x}_0 - \mathbf{m}_0)^\intercal Q^{-1}(\mathbf{x}_0 - \mathbf{m}_0)
\end{aligned} \tag{26}$$

P_1 follows from the observation that the components of $\mathbf{x}_0$, $\mathbf{x}_1$ and $\mathbf{y_1}$ are combined quadratically in Eq. 26.

$P_n \rightarrow P_{n+1}$:

$$\begin{aligned}
\ln P(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \mathbf{y}_1, \ldots, \mathbf{y}_n, \mathbf{y}_{n+1}) =& \ln P(\mathbf{y}_{n+1}|\mathbf{x}_{n+1}) + \\
& \ln P(\mathbf{x}_{n+1}|\mathbf{x}_n) + \\
& \ln P(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_n) \\
=& K - \frac{1}{2}(\mathbf{y}_{n+1} - C\mathbf{x}_{n+1})^\intercal R^{-1}(\mathbf{y}_{n+1} - C\mathbf{x}_{n+1}) \\
& - \frac{1}{2}(\mathbf{x}_{n+1} - A\mathbf{x}_n)^\intercal R^{-1}(\mathbf{x}_{n+1} - A\mathbf{x}_n) \\
& + \ln P(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_n)
\end{aligned} \tag{27}$$

P_{n+1} follows from the observation that the components of $\mathbf{x}_n$, $\mathbf{x}_{n+1}$ and $\mathbf{y}_{n+1}$ are combined quadratically in the first two lines of Eq. 27 and, by the inductive hypothesis P_n, the elements of $\mathbf{x}_0, \ldots, \mathbf{x}_n \mathbf{y}_1, \ldots \mathbf{y}_n$ are combined quadratically in the last line of Eq. 27.

□

6 Kalman Smoother

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

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

Inference of smoothed mean and covariances proceeds in a backward fashion: $x_{T|T}, P_{T|T} \rightarrow x_{T-1|T}, P_{T-1|T} \rightarrow x_{T-2|T}, P_{T-2|T} \rightarrow \ldots \rightarrow x_{1|T}, P_{1|T}$. The initial mean and covariances (i.e., $x_{T|T}, P_{T|T}$) are initialized from the last step of the Kalman filter.

7 Evaluation

7.1 Simulations

We compare the accuracy of the Kalman filter and smoother with that of the method of finite differences, to infer velocities and accelerations of a simulated object following the dynamics of the Discrete Wiener Process Acceleration (DWPA) model[1]. We used the following parameters in the simulations:

Name	*Value*
$\mathbf{x}_0$	$[0, 0]$
V_0	diag([0.001, 0.001])
$\gamma = \gamma_1 = \gamma_2$	1.0
$\sigma = \sigma_1 = \sigma_2$	varied

We simulated $N = 10,000$ samples with a step size $dt = 0.001$.

7.1.1 Lower noise ($\sigma = 1\mathrm{e}-10$)

We simulated measurements from the two-dimensional DWPA model with a standard deviation for the noise of the observations set to $\sigma = 1\mathrm{e}-10$. Figure 3 shows the state positions in blue and the noise-corrupted measurements in black.

We next inferred velocities and accelerations from the simulated measurements (Figure 4). For inference we used the Kalman filter, Kalman smoother and the finite difference method. For these low-noise simulations, all velocity and acceleration estimates were very accurate.

7.1.2 Medium noise ($\sigma = 1\mathrm{e}-3$)

We simulated measurements from the two-dimensional DWPA model with a standard deviation for the noise of the observations set to $\sigma = 1\mathrm{e}-3$. Figure 5 shows the state positions in blue and the noise-corrupted measurements in black.

We next inferred velocities and accelerations from the simulated measurements (Figure 6). For these medium-noise simulations, all velocity estimates were accurate. The Kalman filter and smoother estimates of acceleration were also accurate, but the finite difference estimates of acceleration were not.

7.1.3 Higher noise ($\sigma = 1\mathrm{e}-1$)

We simulated measurements from the two-dimensional DWPA model with a standard deviation for the noise of the observations set to $\sigma = 1\mathrm{e}-1$. Figure 7 shows the state positions in blue and the noise-corrupted measurements in black.

We next inferred velocities and accelerations from the simulated measurements (Figure 8). For these high-noise simulations, velocity and acceleration estimates by the Kalman filter and smoother were accurate, but those from the finite difference method were not.

[1] https://github.com/joacorapela/lds_python/blob/master/docs/tracking/tracking.pdf

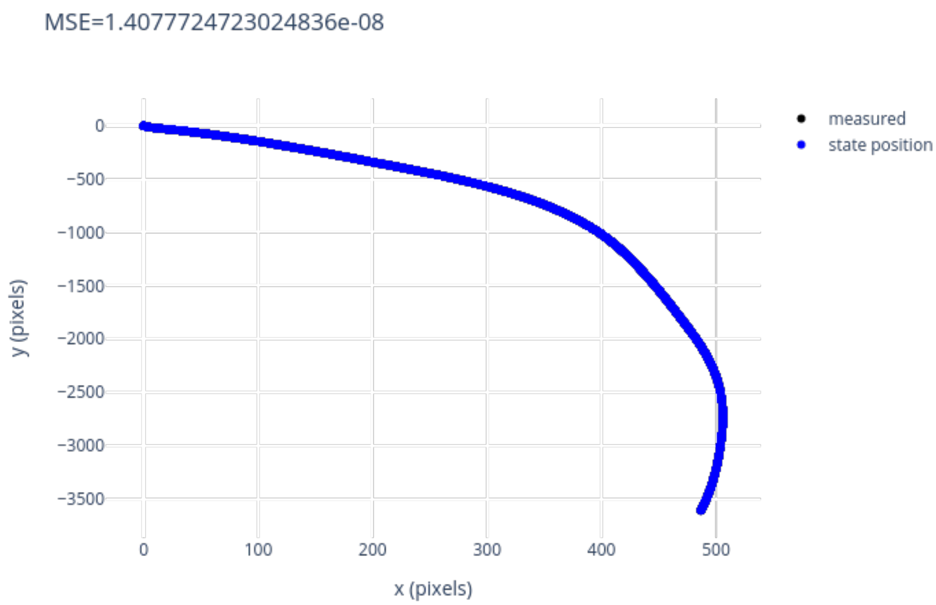
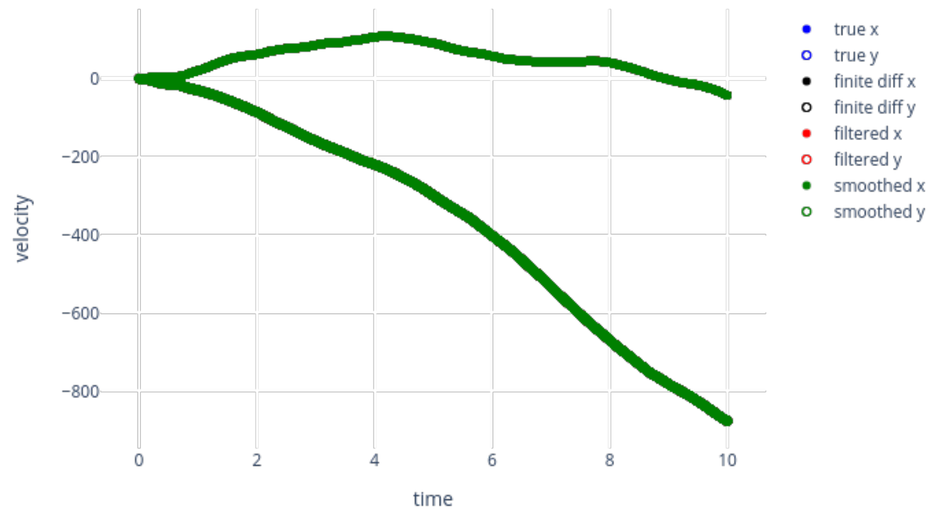
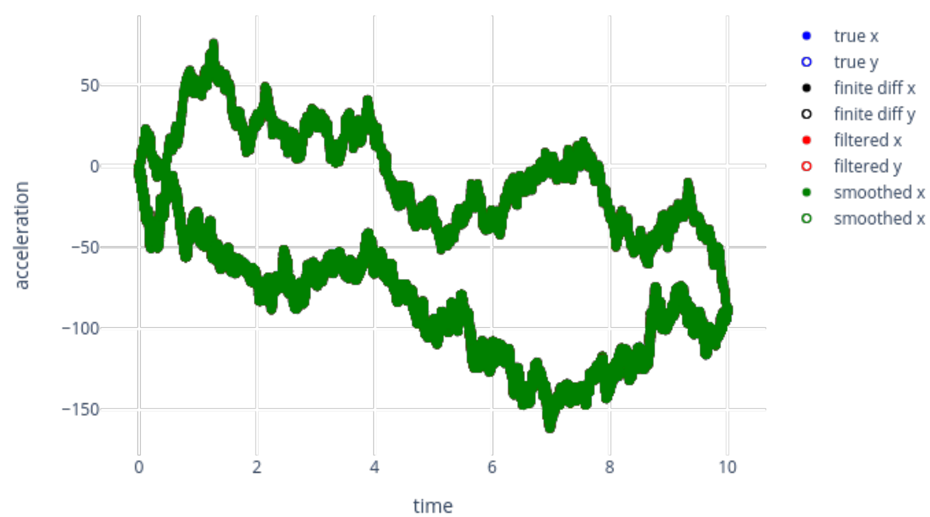


Figure 3: Noise corrupted measurements (black) and state positions (blue) simulated with low noise (standard deviation $\sigma = 1\text{e} - 10$) using a two-dimensional DWPA model. The noise was so low that the differences between measurements and state positions cannot be appreciated visually. The mean-squared error (MSE) between measurements and state positions is indcated in the title. Click on the image to view its interactive version.



(a) Velocity



(b) Acceleration

Figure 4: Estimated velocities (a) and accelerations (b) from low-noise simulations. Estimates were obtained using the finite differences method, the Kalman filter and the Kalman smoother. Velocity and Acceleration estimates using all methods were very accurate. Click on the image to view its interactive version.

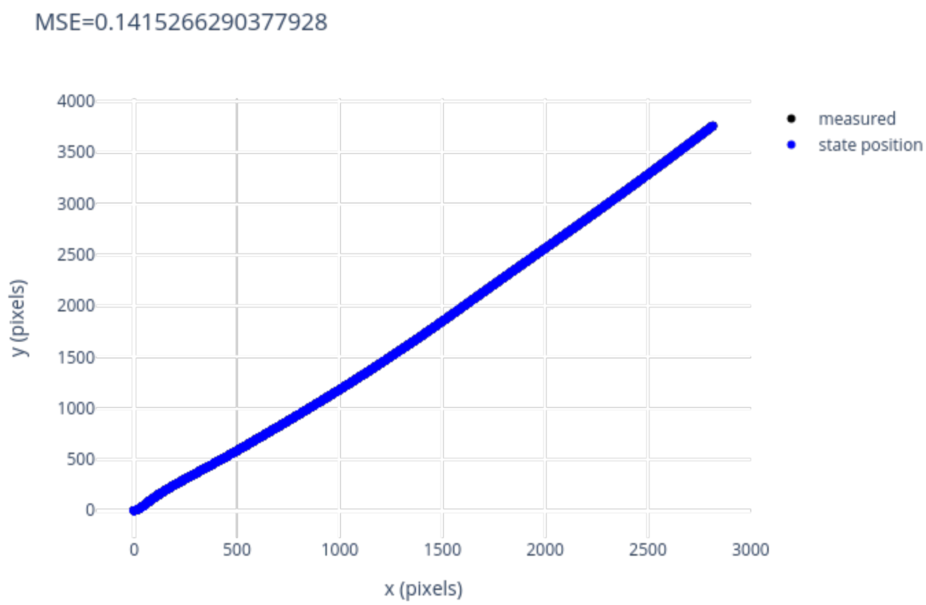
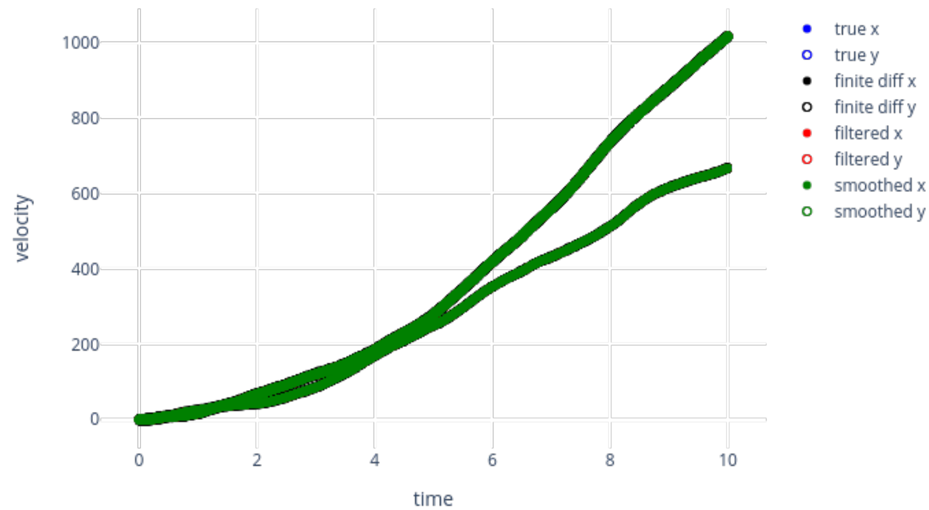
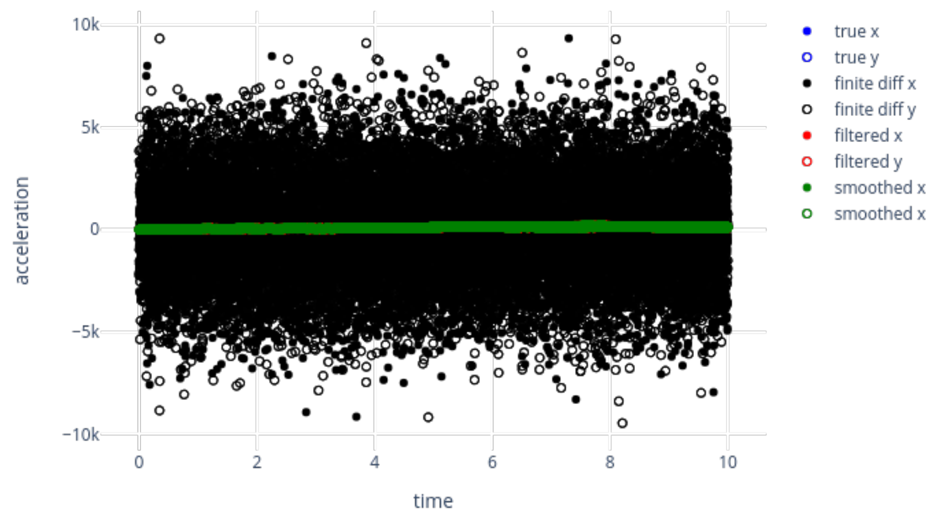


Figure 5: Noise corrupted measurements (black) and state positions (blue) simulated with medium noise (standard deviation $\sigma = 1\mathrm{e}{-3}$) using a two-dimensional DWPA model. The noise was still so low that the differences between measurements and state positions cannot be appreciated visually. The MSE between measurements and state positions is indicated in the title. Click on the image to view its interactive version.



(a) Velocity



(b) Acceleration

Figure 6: Estimated velocities (a) and accelerations (b) from medium-noise simulations. Estimates were obtained using the finite differences method, the Kalman filter and Kalman smoother. Estimates of velocity by all methods were accurate. Estimates of accelerations by the Kalman filter and smoother were also accurate, but not those by the finite difference method. Click on the image to view its interactive version.

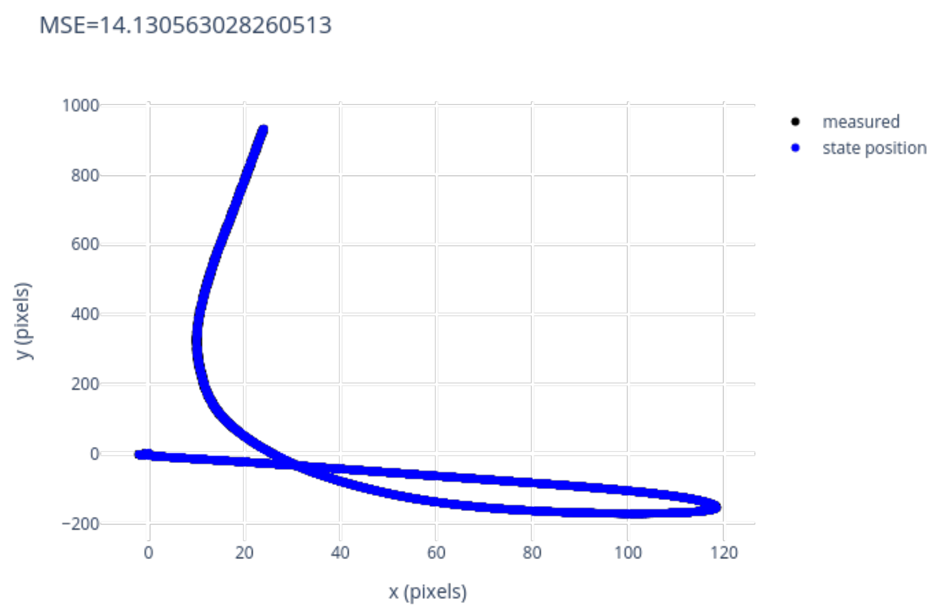


Figure 7: Noise corrupted measurements (black) and state positions (blue) simulated with high noise (standard deviation $\sigma = 1e-1$) using a two-dimensional DWPA model. The noise can now be appreciated visually. The MSE between measurements and state positions is indicated in the title. Click on the image to view its interactive version.

7.1.4 Conclusions

It is remarkable that the finite difference method breaks down when adding very little noise to the true measurements. As we increased the amount of noise, this break down happened earlier for accelerations than for velocities. The Kalman filter and smoother were robust to the amount of noise considered here, both for the estimation of velocities and accelerations.

7.2 Foraging mouse

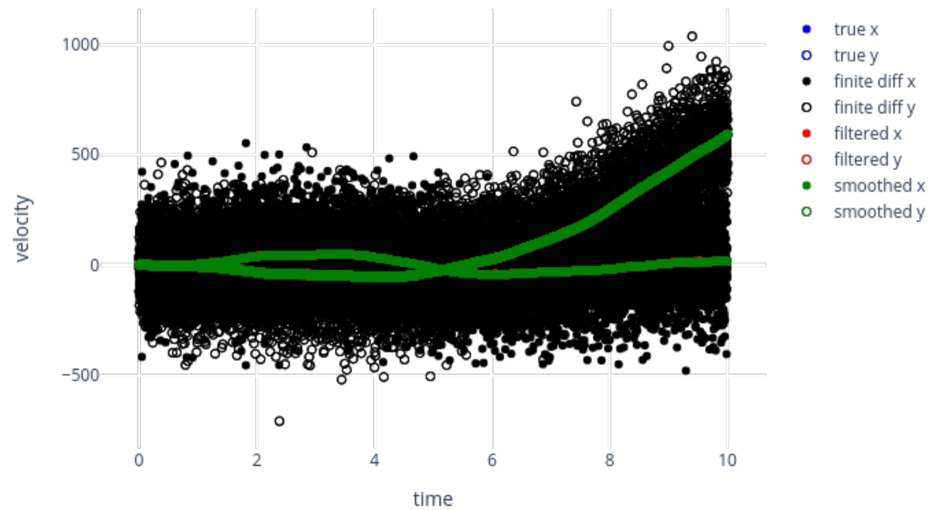
Figure 9 shows the measured and Kalman filtered and smoothed positions of a mouse foraging in a circular arena.

7.3 Mouse running on a maze

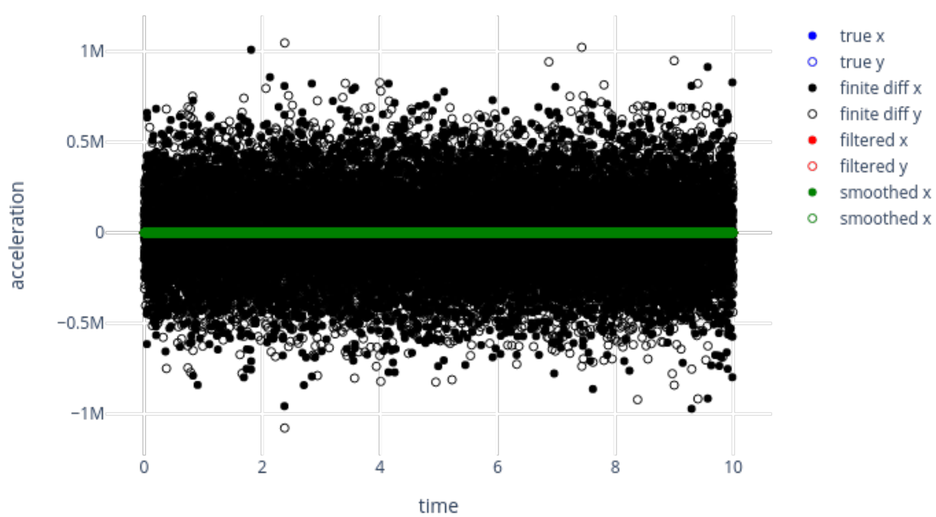
Figure 10 shows the measured and Kalman filtered and smoothed positions of a mouse running in a maze.

References

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



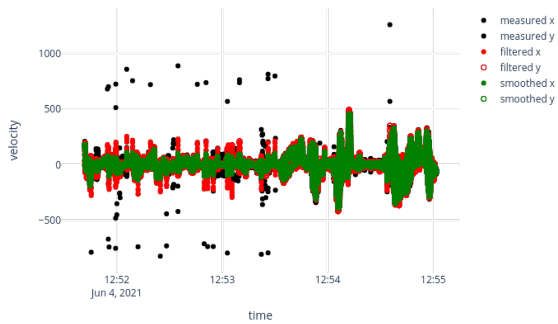
(a) Velocity



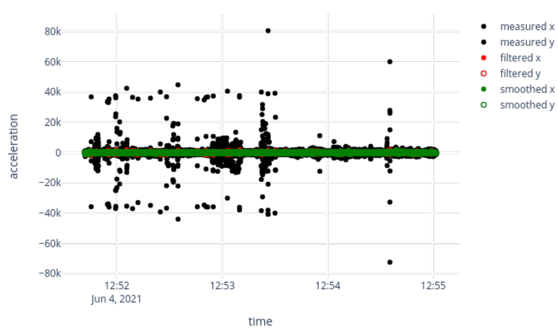
(b) Acceleration

Figure 8: Estimated velocities (a) and accelerations (b) from high-noise simulations. Estimates were obtained using the finite differences method, the Kalman filter and the Kalman smoother. Velocity and Acceleration estimates by the Kalman filter and smoother were accurate, but not those by the finite differences methods. Click on the image to view its interactive version.

(a) Positions

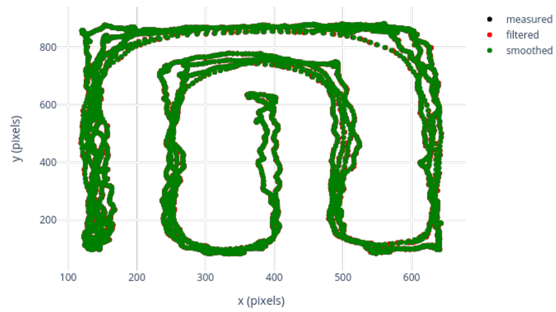


(b) Velocities

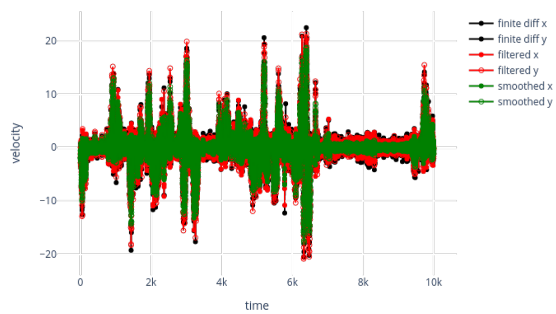


(c) Accelerations

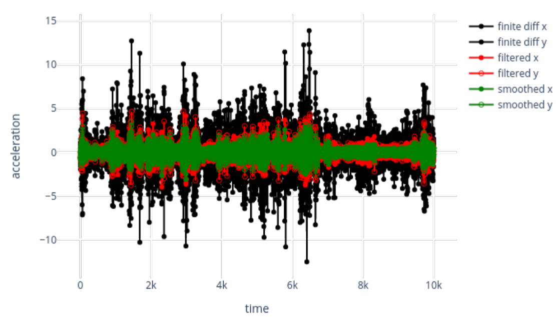
Figure 9: Kalman filtered and smoothed positions, velocities and accelerations of a mouse foraging in a circular arena. Click on the images to see their interactive versions.



(a) Positions



(b) Velocities



(c) Accelerations

Figure 10: Kalman filtered and smoothed positions, velocities and accelerations of a mouse running in a maze. Click on the images to see their interactive versions.