Aalto University
School of Electrical Engineering

ELEC-E8740 — Filtering Problem and Kalman Filtering

Simo Särkkä

Aalto University

November 13, 2020

Contents

Intended Learning Outcomes

After this lecture, you will be able to:

- explain the relationship between the dynamic model, measurement model, and the filtering methodology,
- describe and employ the Kalman filter for linear state-space models.

Recap (1/3)

- The discretization of the linear ODE model

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_u \mathbf{u}(t)$$

is

$$\mathbf{x}_n = \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{L}_n \mathbf{u}_{n-1}$$

$$\mathbf{F}_n \triangleq e^{\mathbf{A}(t_n - t_{n-1})}, \; \mathbf{L}_n \triangleq \int_{t_{n-1}}^{t_n} e^{\mathbf{A}(t_n - t)} \mathbf{B}_u \mathrm{d}t$$

- The discretization of the linear SDE model

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_w \mathbf{w}(t)$$

is

$$\mathbf{x}_n = \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{q}_n, \; \mathbf{q}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$

$$\mathbf{Q}_n = \int_{t_{n-1}}^{t_n} e^{\mathbf{A}(t_n - \tau)} \mathbf{B}_w \mathbf{\Sigma}_w \mathbf{B}_w^{\mathsf{T}} e^{\mathbf{A}^{\mathsf{T}}(t_n - \tau)} \mathrm{d}\tau$$

Recap (2/3)

- Nonlinear dynamic model:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{B}_w(\mathbf{x}(t))\mathbf{w}(t)$$

- Discretization of the linearized model:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t)) + \mathbf{B}_w\mathbf{w}(t) \\ &\approx \mathbf{f}(\mathbf{x}_{n-1}) + \mathbf{A}_x(\mathbf{x}(t) - \mathbf{x}_{n-1}) + \mathbf{B}_w\mathbf{w}(t)\end{aligned}$$

$$\Downarrow$$

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \int_{t_{n-1}}^{t_n} e^{\mathbf{A}_x(t_n - t)} \mathrm{d}t \mathbf{f}(\mathbf{x}_{n-1}) + \mathbf{q}_n$$

with

$$\mathbf{q}_n \sim \mathcal{N}(0, \mathbf{Q}_n), \; \mathbf{Q}_n \approx \int_{t_{n-1}}^{t_n} e^{\mathbf{A}_x(t_n-\tau)} \mathbf{B}_w \mathbf{\Sigma}_w \mathbf{B}_w^\mathsf{T} e^{\mathbf{A}_x^\mathsf{T}(t_n-\tau)} \mathrm{d}\tau$$

Recap (3/3)

- Euler–Maruyama discretization:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{B}_w(\mathbf{x}(t))\mathbf{w}(t)$$
$$\Downarrow$$
$$\mathbf{x}_n = \mathbf{x}_{n-1} + \Delta t \mathbf{f}(\mathbf{x}_{n-1}) + \mathbf{q}_n$$

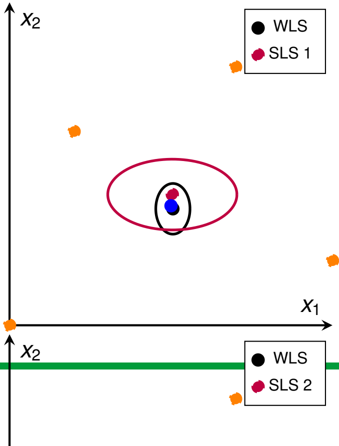
with

$$\begin{aligned}
\mathbf{q}_n &\sim \mathcal{N}(0, \mathbf{Q}_n), \\
\mathbf{Q}_n &\approx \Delta t \mathbf{B}_w(\mathbf{x}_{n-1}) \boldsymbol{\Sigma}_w \mathbf{B}_w(\mathbf{x}_{n-1})^\mathsf{T}.
\end{aligned}$$

Recall: Car Localization with Sequential Least Squares

- Goal: Estimate the position $\mathbf{x} = \begin{bmatrix} p^x & p^y \end{bmatrix}^\mathsf{T}$
- Measurements: Noisy position measurements $n = 1, 2, \ldots$:

$$\mathbf{y}_n = \mathbf{G}_n \mathbf{x}_n + \mathbf{r}_n$$



The Filtering Approach

- We can also allow the target to move between the measurements.
  - This movement can be described by a stochastic dynamic model.
- The measurement can be handled as in sequential least squares.
- Filter iterates the following two steps for all points in time:
  1. Prediction: Predict the current state using the dynamic model (also called *time update*)
  2. Measurement update: Estimate the current state using the prediction and the new measurement

The Filtering Approach: Prediction

- Objective: Predict the current state $\mathbf{x}_n$ at t_n given all previous data $\mathbf{y}_{1:n-1} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\}$
- Done by solving the dynamic model equation starting from the mean and covariance at the previous step.
- Notation:
  - $\hat{\mathbf{x}}_{n|n-1}$: Denotes the predicted value of $\mathbf{x}_n$ given the measurements $\mathbf{y}_{1:n-1} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}\}$
  - $\mathbf{P}_{n|n-1}$: Denotes the covariance of the predicted $\mathbf{x}_n$
- Prediction adds uncertainty

The Filtering Approach: Measurement Update

- Objective: Estimate the current value of $\mathbf{x}_n$ given the new measurement $\mathbf{y}_n$, taking the prediction into account
- Done by solving the regularized least squares problem with the predicted result as the regularization term.
- Notation:
  - $\hat{\mathbf{x}}_{n|n}$: Denotes the estimated value at t_n, given the measurements $\mathbf{y}_{1:n} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$
  - $\mathbf{P}_{n|n}$: Denotes the covariance at t_n
- The measurement update reduces uncertainty

Linear State-Space Model

- Linear state-space model:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{w}(t)$$
$$\mathbf{y}_n = \mathbf{G}_n \mathbf{x}(t_n) + \mathbf{r}_n$$

- Discrete-time equivalent:

$$\mathbf{x}_n = \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{q}_n$$
$$\mathbf{y}_n = \mathbf{G}_n \mathbf{x}_n + \mathbf{r}_n$$

with

$$\mathrm{E}\{\mathbf{q}_n\} = 0, \ \mathrm{Cov}\{\mathbf{q}_n\} = \mathbf{Q}_n,$$
$$\mathrm{E}\{\mathbf{r}_n\} = 0, \ \mathrm{Cov}\{\mathbf{r}_n\} = \mathbf{R}_n$$

- Initial conditions:

$$\mathrm{E}\{\mathbf{x}_0\} = \mathbf{m}_0$$
$$\mathrm{Cov}\{\mathbf{x}_0\} = \mathbf{P}_0$$

Linear Model: Prediction (1/2)

- Linear dynamic model:

$$\mathbf{x}_n = \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{q}_n, \quad \mathrm{Cov}\{\mathbf{q}_n\} = \mathbf{Q}_n$$

- Given:

$$\begin{aligned} \mathrm{E}\{\mathbf{x}_{n-1} \mid \mathbf{y}_{1:n-1}\} &= \hat{\mathbf{x}}_{n-1|n-1}, \\ \mathrm{Cov}\{\mathbf{x}_{n-1} \mid \mathbf{y}_{1:n-1}\} &= \mathbf{P}_{n-1|n-1}. \end{aligned}$$

- Predicted mean:

$$\begin{aligned} \hat{\mathbf{x}}_{n|n-1} &= \mathrm{E}\{\mathbf{x}_n \mid \mathbf{y}_{1:n-1}\} \\ &= \mathbf{F}_n \hat{\mathbf{x}}_{n-1|n-1} \end{aligned}$$

Linear Model: Prediction (2/2)

- Linear dynamic model:

$$\mathbf{x}_n = \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{q}_n, \quad \mathrm{Cov}\{\mathbf{q}_n\} = \mathbf{Q}_n$$

- Given:

$$\begin{aligned} \mathrm{E}\{\mathbf{x}_{n-1} \mid \mathbf{y}_{1:n-1}\} &= \hat{\mathbf{x}}_{n-1|n-1}, \\ \mathrm{Cov}\{\mathbf{x}_{n-1} \mid \mathbf{y}_{1:n-1}\} &= \mathbf{P}_{n-1|n-1}. \end{aligned}$$

- Covariance:

$$\begin{aligned} \mathbf{P}_{n|n-1} &= \mathrm{Cov}\{\mathbf{x}_n \mid \mathbf{y}_{1:n-1}\} \\ &= \mathrm{E}\{(\mathbf{x}_n - \mathrm{E}\{\mathbf{x}_n \mid \mathbf{y}_{1:n-1}\})(\mathbf{x}_n - \mathrm{E}\{\mathbf{x}_n \mid \mathbf{y}_{1:n-1}\})^\mathsf{T} \mid \mathbf{y}_{1:n-1}\} \\ &= \mathbf{F}_n \mathbf{P}_{n-1|n-1} \mathbf{F}_n^\mathsf{T} + \mathbf{Q}_n \end{aligned}$$

Linear Model: Measurement Update (1/3)

- Assume that the prediction yields the prior knowledge $\hat{\mathbf{x}}_{n|n-1}$, $\mathbf{P}_{n|n-1}$
- $\mathbf{y}_n$ provides the new information of the state
- We can use regularized least squares to estimate $\mathbf{x}_n$:

$$J_{\mathrm{ReLS}}(\mathbf{x}_n) = (\mathbf{y}_n - \mathbf{G}_n \mathbf{x}_n)^\mathsf{T} \mathbf{R}_n^{-1} (\mathbf{y}_n - \mathbf{G}_n \mathbf{x}_n) + (\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1})^\mathsf{T} \mathbf{P}_{n|n-1}^{-1} (\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1})$$

and solve

$$\hat{\mathbf{x}}_{n|n} = \operatorname*{argmin}_{\mathbf{x}_n} J_{\mathrm{ReLS}}(\mathbf{x}_n)$$

Linear Model: Measurement Update (2/3)

- Regularized least squares problem:

$$\begin{aligned}\hat{\mathbf{x}}_{n|n} = \operatorname*{argmin}_{\mathbf{x}_n} \, & (\mathbf{y}_n - \mathbf{G}_n \mathbf{x}_n)^\mathsf{T} \mathbf{R}_n^{-1} (\mathbf{y}_n - \mathbf{G}_n \mathbf{x}_n) \\ & + (\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1})^\mathsf{T} \mathbf{P}_{n|n-1}^{-1} (\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1})\end{aligned}$$

- Solution (see Lecture 3 / Chapter 3.4):

$$\begin{aligned}\mathbf{K}_n &= \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} + \mathbf{R}_n)^{-1} \\ \hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n (\mathbf{y}_n - \mathbf{G}_n \hat{\mathbf{x}}_{n|n-1})\end{aligned}$$

- Covariance of $\hat{\mathbf{x}}_{n|n}$:

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} + \mathbf{R}_n) \mathbf{K}_n^\mathsf{T}$$

- $\mathbf{K}_n$ is called the Kalman gain

Linear Model: Measurement Update (3/3)

- Measurement update:

$$
\begin{aligned}
\mathbf{K}_n &= \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} + \mathbf{R}_n)^{-1} \\
\hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n (\mathbf{y}_n - \mathbf{G}_n \hat{\mathbf{x}}_{n|n-1}) \\
\mathbf{P}_{n|n} &= \mathbf{P}_{n|n-1} - \mathbf{K}_n (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} + \mathbf{R}_n) \mathbf{K}_n^\mathsf{T}
\end{aligned}
$$

- It can be shown that it holds that:

$$
\begin{aligned}
\mathrm{E}\{\mathbf{x}_n \mid \mathbf{y}_{1:n}\} &= \hat{\mathbf{x}}_{n|n} \\
\mathrm{Cov}\{\mathbf{x}_n \mid \mathbf{y}_{1:n}\} &= \mathbf{P}_{n|n}
\end{aligned}
$$

Linear Model: Summary

- Prediction:

$$\hat{\mathbf{x}}_{n|n-1} = \mathbf{F}_n \hat{\mathbf{x}}_{n-1|n-1}$$
$$\mathbf{P}_{n|n-1} = \mathbf{F}_n \mathbf{P}_{n-1|n-1} \mathbf{F}_n^\mathsf{T} + \mathbf{Q}_n$$

- Measurement update:

$$\mathbf{K}_n = \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} + \mathbf{R}_n)^{-1}$$
$$\hat{\mathbf{x}}_{n|n} = \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n (\mathbf{y}_n - \mathbf{G}_n \hat{\mathbf{x}}_{n|n-1})$$
$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} + \mathbf{R}_n) \mathbf{K}_n^\mathsf{T}$$

- Initialization of the recursion:

$$\hat{\mathbf{x}}_{0|0} = \mathbf{m}_0$$
$$\mathbf{P}_{0|0} = \mathbf{P}_0$$

The Kalman Filter

Algorithm 1 Kalman Filter

1: Initialize $\hat{\mathbf{x}}_{0|0} = \mathbf{m}_0$, $\mathbf{P}_{0|0} = \mathbf{P}_0$

2: **for** $n = 1, 2, \ldots$ **do**

3: Prediction (time update):

$$
\begin{aligned}
\hat{\mathbf{x}}_{n|n-1} &= \mathbf{F}_n \hat{\mathbf{x}}_{n-1|n-1} \\
\mathbf{P}_{n|n-1} &= \mathbf{F}_n \mathbf{P}_{n-1|n-1} \mathbf{F}_n^{\mathsf{T}} + \mathbf{Q}_n
\end{aligned}
$$

4: Measurement update:

$$
\begin{aligned}
\mathbf{K}_n &= \mathbf{P}_{n|n-1} \mathbf{G}_n^{\mathsf{T}} (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n + \mathbf{R}_n)^{-1} \\
\hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n (\mathbf{y}_n - \mathbf{G}_n \hat{\mathbf{x}}_{n|n-1}) \\
\mathbf{P}_{n|n} &= \mathbf{P}_{n|n-1} - \mathbf{K}_n (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n + \mathbf{R}_n) \mathbf{K}_n^{\mathsf{T}}
\end{aligned}
$$

5: **end for**

Example: Car Localization (1/3)

- Goal: Estimate the kinematic state at each time t_n
- Dynamic model: 2D Wiener velocity model:

$$
\begin{bmatrix} \dot{p}^x(t) \\ \dot{p}^y(t) \\ \dot{v}^x(t) \\ \dot{v}^y(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p^x(t) \\ p^y(t) \\ v^x(t) \\ v^y(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}
$$

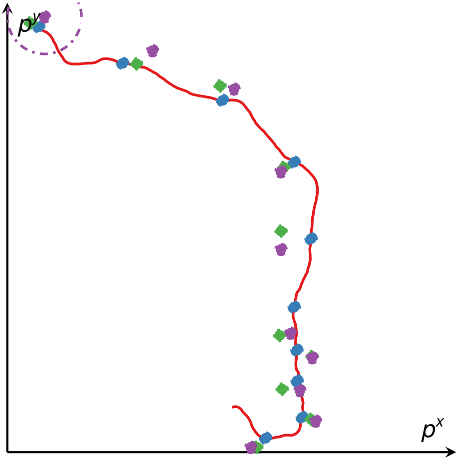
- Measurements: Noisy position ($\mathbf{G}_n = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}$):

$$
\mathbf{y}_n = \mathbf{G}_n \mathbf{x}_n + \mathbf{r}_n
$$

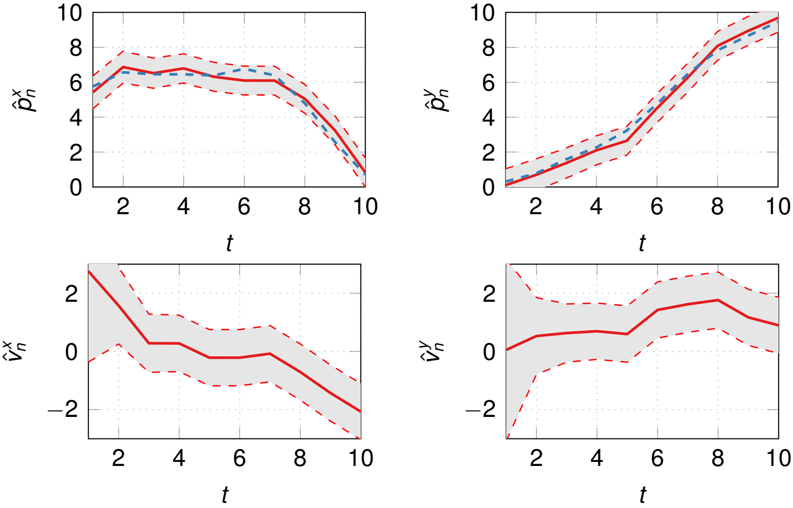
- Discrete-time linear state-space model:

$$
\begin{aligned}
\mathbf{x}_n &= \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{q}_n \\
\mathbf{y}_n &= \mathbf{G}_n \mathbf{x}_n + \mathbf{r}_n
\end{aligned}
$$

Example: Car Localization (2/3)



Example: Car Localization (3/3)



Performance Evaluation

A few questions:

- Why not just use the measurements $\mathbf{y}_n$ as the position estimate ($\hat{\mathbf{p}}_n = \mathbf{y}_n$)?
- How should we assess the performance of the algorithm?
- One possible criterion: The root mean squared error (RMSE):

$$e_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{\mathbf{x}}_{n|n} - \mathbf{x}_n)^{\mathsf{T}}(\hat{\mathbf{x}}_{n|n} - \mathbf{x}_n)}$$

Example: Car Localization

- RMSE for the primitive approach ($\hat{\mathbf{p}}_n = \mathbf{y}_n$):

$$e_{\text{RMSE}} = 0.41$$

- RMSE for Kalman filter:

$$e_{\text{RMSE}} = 0.29$$

- The prior knowledge imposed by the dynamic model significantly improves performance!

Measurement Update: Some Observations

- Measurement update:

$$\begin{aligned}
\mathbf{K}_n &= \mathbf{P}_{n|n-1}\mathbf{G}_n^\mathsf{T}(\mathbf{G}_n\mathbf{P}_{n|n-1}\mathbf{G}_n + \mathbf{R}_n)^{-1} \\
\hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n(\mathbf{y}_n - \mathbf{G}_n\hat{\mathbf{x}}_{n|n-1}) \\
\mathbf{P}_{n|n} &= \mathbf{P}_{n|n-1} - \mathbf{K}_n(\mathbf{G}_n\mathbf{P}_{n|n-1}\mathbf{G}_n + \mathbf{R}_n)\mathbf{K}_n^\mathsf{T}
\end{aligned}$$

- Prediction of the output and covariances:

$$\begin{aligned}
\mathrm{E}\{\mathbf{y}_n \mid \mathbf{y}_{1:n-1}\} &= \mathbf{G}_n\hat{\mathbf{x}}_{n|n-1} \\
\mathrm{Cov}\{\mathbf{y}_n \mid \mathbf{y}_{1:n-1}\} &= \mathbf{G}_n\mathbf{P}_{n|n-1}\mathbf{G}_n^\mathsf{T} + \mathbf{R}_n \\
\mathrm{Cov}\{\mathbf{x}_n, \mathbf{y}_n \mid \mathbf{y}_{1:n-1}\} &= \mathbf{P}_{n|n-1}\mathbf{G}_n^\mathsf{T}
\end{aligned}$$

Summary

- The filtering approach iterates between two steps:
  1. Prediction: $\hat{\mathbf{x}}_{n-1|n-1}, \mathbf{P}_{n-1|n-1} \Rightarrow \hat{\mathbf{x}}_{n|n-1}, \mathbf{P}_{n|n-1}$
  2. Measurement update: $\hat{\mathbf{x}}_{n|n-1}, \mathbf{P}_{n|n-1} \Rightarrow \hat{\mathbf{x}}_{n|n}, \mathbf{P}_{n|n}$
- The Kalman filter is the optimal filter for linear state-space models
  1. Prediction:

$$\begin{aligned}\hat{\mathbf{x}}_{n|n-1} &= \mathbf{F}_n \hat{\mathbf{x}}_{n-1|n-1} \\ \mathbf{P}_{n|n-1} &= \mathbf{F}_n \mathbf{P}_{n-1|n-1} \mathbf{F}_n^\mathsf{T} + \mathbf{Q}_n\end{aligned}$$

  2. Measurement update:

$$\begin{aligned}\mathbf{K}_n &= \mathbf{P}_{n|n-1} \mathbf{G}_n^\mathsf{T} (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n + \mathbf{R}_n)^{-1} \\ \hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n (\mathbf{y}_n - \mathbf{G}_n \hat{\mathbf{x}}_{n|n-1}) \\ \mathbf{P}_{n|n} &= \mathbf{P}_{n|n-1} - \mathbf{K}_n (\mathbf{G}_n \mathbf{P}_{n|n-1} \mathbf{G}_n + \mathbf{R}_n) \mathbf{K}_n^\mathsf{T}\end{aligned}$$