Adventures in Kalman Filtering
— The “Prediction - Correction” World —

Prof. James Richard Forbes

McGill University, Department of Mechanical Engineering

May 13, 2021

- Sensors rarely measure states of interest directly. How do we "back out" states that are not measured directly?
  - Within an IMU there is a rate gyro, an accelerometer, and often a magnetometer.
  - The rate gyro measures $\underrightarrow{\omega}^{ba}$ (resolved in what frame?), *not* a set of Euler angles $\boldsymbol{\theta}^{ba}$, not a quaternion $\mathbf{q}^{ba}$, nor a DCM $\mathbf{C}_{ba}$, and *not* a set of Euler-angle rates $\dot{\boldsymbol{\theta}}^{ba}$, not a quaternion rate $\dot{\mathbf{q}}^{ba}$, nor a DCM rate $\dot{\mathbf{C}}_{ba}$.[1]
  - The accelerometer measures $\underrightarrow{a}$ (resolved in what frame?), *not* $\underrightarrow{v}$, and *not* $\underrightarrow{r}$.
  - A magnetometer measures $\underrightarrow{m}$ (resolved in what frame?), *not* $\boldsymbol{\theta}^{ba}$.
  - There's no such thing as an "attitude sensor".
- Sensor data is imperfect; noise corrupts all measurements, and some measurements are (significantly) biased.
- Because noise and bias are *random*, we rely on concepts from probability theory to describe the properties of noise and bias that we are interested in *filtering*.

---

[1] $\underrightarrow{\omega}^{ba}$ is the angular velocity of frame b relative to frame a. A rate gyro measures $\underrightarrow{\omega}^{ba}$ (resolved in what frame?), and not a set of Euler angles, nor a set of Euler angle rates, nor a quaternion, nor a quaternion rate.

The Gaussian Distribution

- A continuous random variable is said to have a *normal* or *Gaussian* distribution if the pdf associated with the random variable x is given by

$$p(x; \bar{x}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right).$$

- $p(x; \bar{x}, \sigma^2)$ being a pdf means that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right) \mathrm{d}x = 1,$$

  where the mean is

$$\bar{x} = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right) \mathrm{d}x,$$

  and the variance is

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right) \mathrm{d}x.$$

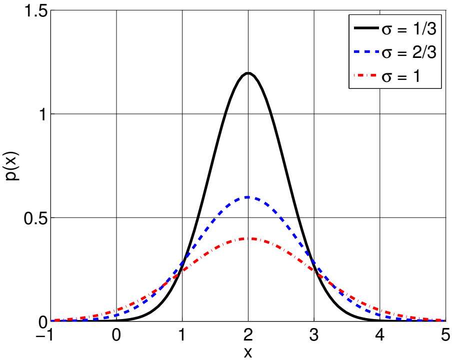

Figure: Gaussian pdfs where $\bar{x} = 2$ and σ takes on values of $1/3$, $2/3$, and 1.

Shown in Figure 1 are three normal distributions. The mean of each is distribution is $\bar{x} = 2$, while the standard deviation of each are $1/3$, $2/3$, and 1, respectively.

A short-hand notation for indicating x is normally distributed is $x \sim \mathcal{N}(\bar{x}, \sigma^2)$.

The Multidimensional Case

- In the N-dimensional case, a continuous random column matrix $\mathbf{x} \in \mathbb{R}^N$ is said to have a normal or Gaussian distribution if the pdf associated with $\mathbf{x}$ is given by

$$p(\mathbf{x}; \bar{\mathbf{x}}, \mathbf{Q}) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{Q}}} \exp\left(-\tfrac{1}{2}\left(\mathbf{x}-\bar{\mathbf{x}}\right)^{\mathsf{T}} \mathbf{Q}^{-1}\left(\mathbf{x}-\bar{\mathbf{x}}\right)\right),$$

  where $\bar{\mathbf{x}}$ is the mean and $\mathbf{Q}$ is the covariance matrix.
- The covariance matrix is symmetric and positive definite (thus ensuring $\mathbf{Q}$ is not singular, and thus $\mathbf{Q}^{-1}$ exists).
- Being a pdf, it can be shown that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^N \det \mathbf{Q}}} \exp\left(-\tfrac{1}{2}\left(\mathbf{x}-\bar{\mathbf{x}}\right)^{\mathsf{T}} \mathbf{Q}^{-1}\left(\mathbf{x}-\bar{\mathbf{x}}\right)\right) \mathrm{d}\mathbf{x} = 1,$$

  the mean is

$$\bar{\mathbf{x}} = \int_{-\infty}^{\infty} \mathbf{x} \frac{1}{\sqrt{(2\pi)^N \det \mathbf{Q}}} \exp\left(-\tfrac{1}{2}\left(\mathbf{x}-\bar{\mathbf{x}}\right)^{\mathsf{T}} \mathbf{Q}^{-1}\left(\mathbf{x}-\bar{\mathbf{x}}\right)\right) \mathrm{d}\mathbf{x},$$

  and the covariance is

$$\mathbf{Q} = \int_{-\infty}^{\infty} (\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^{\mathsf{T}} \frac{1}{\sqrt{(2\pi)^N \det \mathbf{Q}}} \exp\left(-\tfrac{1}{2}\left(\mathbf{x}-\bar{\mathbf{x}}\right)^{\mathsf{T}} \mathbf{Q}^{-1}\left(\mathbf{x}-\bar{\mathbf{x}}\right)\right) \mathrm{d}\mathbf{x}.$$

- A short-hand notation for indicating $\mathbf{x}$ is normally distributed is $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{Q})$.

The Static Case

- Consider
$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{xy}^{\mathsf{T}} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} \right). \tag{1}$$
- Consider the *affine* estimator
$$\hat{\mathbf{x}} = \mathbf{K}\mathbf{y} + \boldsymbol{\ell},$$
where $\hat{\mathbf{x}}$ is the estimate of the state $\mathbf{x}$ given the measurement $\mathbf{y}$.
- What form should $\mathbf{K}$ and $\boldsymbol{\ell}$ take?
- How can *a priori* information, such as that given in (1), be used to generate the estimated state $\hat{\mathbf{x}}$?
- Define the error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$.
- An **unbiased** estimate is desired, meaning $E\left[\mathbf{e}\right] = \mathbf{0}$.
- Using this definition,
$$\begin{aligned} \mathbf{0} &= E\left[\mathbf{x} - \hat{\mathbf{x}}\right] = E\left[\mathbf{x} - \mathbf{K}\mathbf{y} - \boldsymbol{\ell}\right] = E\left[\mathbf{x}\right] - E\left[\mathbf{K}\mathbf{y}\right] - \boldsymbol{\ell} = \boldsymbol{\mu}_x - \mathbf{K}\boldsymbol{\mu}_y - \boldsymbol{\ell}, \\ \boldsymbol{\ell} &= \boldsymbol{\mu}_x - \mathbf{K}\boldsymbol{\mu}_y. \end{aligned}$$
- Thus, an unbiased estimator is of the form
$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{K}\mathbf{y} + \boldsymbol{\ell} \\ &= \mathbf{K}\mathbf{y} + \boldsymbol{\mu}_x - \mathbf{K}\boldsymbol{\mu}_y \\ &= \boldsymbol{\mu}_x + \mathbf{K}(\mathbf{y} - \boldsymbol{\mu}_y). \end{aligned}$$

- How should we pick $\mathbf{K}$ to provide a **best** estimate?
- Consider

$$\begin{aligned}
\mathbf{P} &= E\left[\mathbf{e}\mathbf{e}^\mathsf{T}\right] \\
&= E\left[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^\mathsf{T}\right] \\
&= E\left[(\mathbf{x}-\boldsymbol{\mu}_x-\mathbf{K}(\mathbf{y}-\boldsymbol{\mu}_y))(\mathbf{x}-\boldsymbol{\mu}_x-\mathbf{K}(\mathbf{y}-\boldsymbol{\mu}_y))^\mathsf{T}\right] \\
&= E\left[(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^\mathsf{T}\right] - E\left[(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{y}-\boldsymbol{\mu}_y)^\mathsf{T}\right]\mathbf{K}^\mathsf{T} \\
&- \mathbf{K}E\left[(\mathbf{y}-\boldsymbol{\mu}_y)(\mathbf{x}-\boldsymbol{\mu}_x)\right] + \mathbf{K}E\left[(\mathbf{y}-\boldsymbol{\mu}_y)(\mathbf{y}-\boldsymbol{\mu}_y)^\mathsf{T}\right]\mathbf{K}^\mathsf{T} \\
&= \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\mathbf{K}^\mathsf{T} - \mathbf{K}\boldsymbol{\Sigma}_{xy}^\mathsf{T} + \mathbf{K}\boldsymbol{\Sigma}_{yy}\mathbf{K}^\mathsf{T}
\end{aligned}$$

- Recall that $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}^\mathsf{T})$, $\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$ and that $\mathrm{tr}(\mathbf{CD}) = \mathrm{tr}(\mathbf{DC})$ for all $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, $\mathbf{C} \in \mathbb{R}^{n\times m}$, $\mathbf{D} \in \mathbb{R}^{m\times n}$.
- Write $J(\mathbf{K}) = \mathrm{tr}(\mathbf{P})$ as

$$\begin{aligned}
J(\mathbf{K}) &= \mathrm{tr}(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\mathbf{K}^\mathsf{T} - \mathbf{K}\boldsymbol{\Sigma}_{xy}^\mathsf{T} + \mathbf{K}\boldsymbol{\Sigma}_{yy}\mathbf{K}^\mathsf{T}) \\
&= \mathrm{tr}(\boldsymbol{\Sigma}_{xx}) - \mathrm{tr}(\boldsymbol{\Sigma}_{xy}\mathbf{K}^\mathsf{T}) - \mathrm{tr}(\mathbf{K}\boldsymbol{\Sigma}_{xy}^\mathsf{T}) + \mathrm{tr}(\mathbf{K}\boldsymbol{\Sigma}_{yy}\mathbf{K}^\mathsf{T}) \\
&= \mathrm{tr}(\boldsymbol{\Sigma}_{xx}) - 2\mathrm{tr}(\boldsymbol{\Sigma}_{xy}\mathbf{K}^\mathsf{T}) + \mathrm{tr}(\mathbf{K}\boldsymbol{\Sigma}_{yy}\mathbf{K}^\mathsf{T})
\end{aligned}$$

- Consider a Taylor series expansion of a general function $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$, that is

$$f(\bar{\mathbf{x}} + \delta\mathbf{x}) = f(\bar{\mathbf{x}}) + \left[\left. \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\bar{\mathbf{x}}} \right] \delta\mathbf{x} + \tfrac{1}{2}\delta\mathbf{x}^{\mathsf{T}} \left[\left. \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}^{\mathsf{T}} \right) \right|_{\mathbf{x}=\bar{\mathbf{x}}} \right] \delta\mathbf{x} + \text{H.O.T.}$$

  where "H.O.T." means "higher-order terms", and

$$\left. \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\bar{\mathbf{x}}}, \qquad \left. \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}^{\mathsf{T}} \right) \right|_{\mathbf{x}=\bar{\mathbf{x}}}$$

  are the Jacobain and Hessian of $f(\cdot)$ evaluated at $\mathbf{x} = \bar{\mathbf{x}}$, respectfully.
- A necessary condition for $\bar{\mathbf{x}}$ to be an extremum (either a maximum or a minimum) is

$$\left. \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\bar{\mathbf{x}}} = \mathbf{0}.$$

- When $\mathbf{H} > 0$ then $\bar{\mathbf{x}}$ corresponds to a minimum.

- Consider $\mathbf{K} = \bar{\mathbf{K}} + \delta\mathbf{K}$ and a Taylor series expansion of $J(\cdot)$. To this end,

$$\begin{aligned} J(\bar{\mathbf{K}} + \delta\mathbf{K}) &= \mathrm{tr}(\mathbf{\Sigma}_{xx}) - 2\mathrm{tr}(\mathbf{\Sigma}_{xy}(\bar{\mathbf{K}} + \delta\mathbf{K})^{\mathsf{T}}) + \mathrm{tr}((\bar{\mathbf{K}} + \delta\mathbf{K})\mathbf{\Sigma}_{yy}(\bar{\mathbf{K}} + \delta\mathbf{K})^{\mathsf{T}}) \\ &= \mathrm{tr}(\mathbf{\Sigma}_{xx}) - 2\mathrm{tr}(\mathbf{\Sigma}_{xy}\bar{\mathbf{K}}^{\mathsf{T}}) - 2\mathrm{tr}(\mathbf{\Sigma}_{xy}\delta\mathbf{K}^{\mathsf{T}}) \\ &\quad + \mathrm{tr}(\bar{\mathbf{K}}\mathbf{\Sigma}_{yy}\bar{\mathbf{K}}^{\mathsf{T}}) + \mathrm{tr}(\bar{\mathbf{K}}\mathbf{\Sigma}_{yy}\delta\mathbf{K}^{\mathsf{T}}) + \mathrm{tr}(\delta\mathbf{K}\mathbf{\Sigma}_{yy}\bar{\mathbf{K}}^{\mathsf{T}}) + \mathrm{tr}(\delta\mathbf{K}\mathbf{\Sigma}_{yy}\delta\mathbf{K}^{\mathsf{T}}) \\ &= \underbrace{\mathrm{tr}(\mathbf{\Sigma}_{xx}) - 2\mathrm{tr}(\mathbf{\Sigma}_{xy}\bar{\mathbf{K}}^{\mathsf{T}}) + \mathrm{tr}(\bar{\mathbf{K}}\mathbf{\Sigma}_{yy}\bar{\mathbf{K}}^{\mathsf{T}})}_{J(\bar{\mathbf{K}})} \\ &\quad - 2\mathrm{tr}(\mathbf{\Sigma}_{xy}\delta\mathbf{K}^{\mathsf{T}}) + 2\mathrm{tr}(\bar{\mathbf{K}}\mathbf{\Sigma}_{yy}\delta\mathbf{K}^{\mathsf{T}}) + \mathrm{tr}(\delta\mathbf{K}\mathbf{\Sigma}_{yy}\delta\mathbf{K}^{\mathsf{T}}) \\ &= J(\bar{\mathbf{K}}) - 2\mathrm{tr}(\mathbf{\Sigma}_{xy}\delta\mathbf{K}^{\mathsf{T}} - \bar{\mathbf{K}}\mathbf{\Sigma}_{yy}\delta\mathbf{K}^{\mathsf{T}}) + \mathrm{tr}(\delta\mathbf{K}\mathbf{\Sigma}_{yy}\delta\mathbf{K}^{\mathsf{T}}) \\ &= J(\bar{\mathbf{K}}) - 2\mathrm{tr}((\mathbf{\Sigma}_{xy} - \bar{\mathbf{K}}\mathbf{\Sigma}_{yy})\delta\mathbf{K}^{\mathsf{T}}) + \mathrm{tr}(\delta\mathbf{K}\mathbf{\Sigma}_{yy}\delta\mathbf{K}^{\mathsf{T}}) \end{aligned}$$

- Thus,

$$\left.\frac{\partial J(\mathbf{K})}{\partial \mathbf{K}}\right|_{\mathbf{K}=\bar{\mathbf{K}}} = \mathbf{\Sigma}_{xy} - \bar{\mathbf{K}}\mathbf{\Sigma}_{yy}, \qquad \left.\frac{\partial}{\partial \mathbf{K}}\left(\frac{\partial J(\mathbf{K})}{\partial \mathbf{K}}^{\mathsf{T}}\right)\right|_{\mathbf{K}=\bar{\mathbf{K}}} = \mathbf{\Sigma}_{yy}$$

- Note, from the above derivation it follows that

$$\frac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X}^{\mathsf{T}})}{\partial \mathbf{X}} = \mathbf{A}, \qquad \frac{\partial \mathrm{tr}(\mathbf{X}\mathbf{A}\mathbf{X}^{\mathsf{T}})}{\partial \mathbf{X}} = 2\mathbf{X}\mathbf{A}.$$

  Don't memorize the above derivative definitions . . . understand the fundamentals, the bigger picture . . . that being, perturbing the independent variable, a Taylor series expansion, etc.

- For $\bar{\mathbf{K}}$ to be an extremum,

$$\begin{aligned}
\left.\frac{\partial J(\mathbf{K})}{\partial \mathbf{K}}\right|_{\mathbf{K}=\bar{\mathbf{K}}} &= \mathbf{0}, \\
\mathbf{\Sigma}_{xy} - \bar{\mathbf{K}}\mathbf{\Sigma}_{yy} &= \mathbf{0}, \\
\bar{\mathbf{K}}\mathbf{\Sigma}_{yy} &= \mathbf{\Sigma}_{xy}, \\
\bar{\mathbf{K}} &= \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}.
\end{aligned}$$

- The Hessian is $\mathbf{\Sigma}_{yy} > 0$. Thus, $\bar{\mathbf{K}} = \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}$ corresponds to a minimum of $J(\mathbf{K}) = \mathrm{tr}(\mathbf{P})$.
- In fact, because $J(\cdot)$ is convex, this minimum is a global minimum, and thus an unique minimum.
- Thus,

$$\begin{aligned}
\hat{\mathbf{x}} &= \boldsymbol{\mu}_x + \bar{\mathbf{K}}(\mathbf{y} - \boldsymbol{\mu}_y) \\
&= \boldsymbol{\mu}_x + \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)
\end{aligned}$$

  provides a best, unbiased, estimate of $\mathbf{x}$ given the measurement (or realization) $\mathbf{y}$ and the *a priori* information given in (1).
- Often we drop the “bar” and just write $\mathbf{K} = \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}$.

The Dynamic Case

- Consider a discrete-time system described by linear process (a.k.a. motion) and measurement (a.k.a. observation) models,

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1}, \qquad \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k),$$
$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{M}_k\mathbf{v}_k, \qquad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k).$$

- Let $\hat{\mathbf{x}}_k$ denote a state estimate. Can $\hat{\mathbf{x}}_k$ be found
  1. in an unbiased manner, and
  2. in an optimal manner?
- What does the word “unbiased” mean? It means

$$E\left[\hat{\mathbf{e}}_k\right] = \mathbf{0}, \qquad \forall k = 0, \ldots, K,$$

  where $\hat{\mathbf{e}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$.
- What does the word “optimal” mean? It means an objective function is extremized (either minimized or maximized).
- BLUE — “best, linear, unbiased, estimator”.

- Consider the predict-correct estimator structure,

$$\check{\mathbf{x}}_k = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1},$$
$$\hat{\mathbf{x}}_k = \check{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{y}_k - \check{\mathbf{y}}_k),$$

  where
  - $\check{\mathbf{x}}_k$ is the *a priori*, or predicted, state estimate,
  - $\check{\mathbf{y}}_k = \mathbf{H}_k\check{\mathbf{x}}_k$ is the predicted measurement, and
  - $\hat{\mathbf{x}}_k$ is the *a posteriori*, or corrected, state estimate.
- Define
  - $\check{\mathbf{e}}_k = \mathbf{x}_k - \check{\mathbf{x}}_k$, the *a priori*, or predicted, error,
  - $\check{\mathbf{P}}_k = E\left[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^{\mathsf{T}}\right]$, the *a priori*, or predicted, covariance,
  - $\hat{\mathbf{e}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$, the *a posteriori*, or corrected, error,
  - $\hat{\mathbf{P}}_k = E\left[\hat{\mathbf{e}}_k\hat{\mathbf{e}}_k^{\mathsf{T}}\right]$, the *a posteriori*, or corrected, covariance,
  - $\check{\boldsymbol{\rho}}_k = \mathbf{y}_k - \check{\mathbf{y}}_k$ the innovation, or the residual,
  - $\check{\mathbf{P}}_k^{\mathbf{y}_k\mathbf{y}_k} = E\left[\check{\boldsymbol{\rho}}_k\check{\boldsymbol{\rho}}_k^{\mathsf{T}}\right]$, the covariance associated with the innovation, and
  - $\check{\mathbf{P}}_k^{\mathbf{x}_k\mathbf{y}_k} = E\left[\check{\mathbf{e}}_k\check{\boldsymbol{\rho}}_k^{\mathsf{T}}\right]$, the cross covariance.

- Given $\hat{\mathbf{x}}_{k-1}$, $\hat{\mathbf{P}}_{k-1}$, and $\mathbf{u}_{k-1}$, the predicted state is

$$\check{\mathbf{x}}_k = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1}.$$

- The predicted covariance is

$$\begin{aligned}
\check{\mathbf{P}}_k &= E\left[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^\mathsf{T}\right] \\
&= E\left[(\mathbf{x}_k - \check{\mathbf{x}}_k)\check{\mathbf{e}}_k^\mathsf{T}\right] \\
&= E\left[(\mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1} - \mathbf{F}_{k-1}\check{\mathbf{x}}_k - \mathbf{G}_{k-1}\mathbf{u}_{k-1})\check{\mathbf{e}}_k^\mathsf{T}\right] \\
&= E\left[(\mathbf{F}_{k-1}\hat{\mathbf{e}}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1})(\hat{\mathbf{e}}_{k-1}^\mathsf{T}\mathbf{F}_{k-1}^\mathsf{T} + \mathbf{w}_{k-1}^\mathsf{T}\mathbf{L}_{k-1}^\mathsf{T})\right] \\
&= \mathbf{F}_{k-1}E\left[\hat{\mathbf{e}}_{k-1}\hat{\mathbf{e}}_{k-1}^\mathsf{T}\right]\mathbf{F}_{k-1}^\mathsf{T} + \mathbf{F}_{k-1}E\left[\hat{\mathbf{e}}_{k-1}\mathbf{w}_{k-1}^\mathsf{T}\right]\mathbf{L}_{k-1}^\mathsf{T} \\
&+ \mathbf{L}_{k-1}E\left[\mathbf{w}_{k-1}\hat{\mathbf{e}}_{k-1}^\mathsf{T}\right]\mathbf{F}_{k-1}^\mathsf{T} + \mathbf{L}_{k-1}E\left[\mathbf{w}_{k-1}\mathbf{w}_{k-1}^\mathsf{T}\right]\mathbf{L}_{k-1}^\mathsf{T} \\
&= \mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^\mathsf{T} + \mathbf{L}_{k-1}\mathbf{Q}_k\mathbf{L}_{k-1}^\mathsf{T}
\end{aligned}$$

where $E\left[\mathbf{w}_{k-1}\hat{\mathbf{e}}_{k-1}^\mathsf{T}\right] = \mathbf{0}$, $\hat{\mathbf{P}}_{k-1} = E\left[\hat{\mathbf{e}}_{k-1}\hat{\mathbf{e}}_{k-1}^\mathsf{T}\right]$, and
$\mathbf{Q}_{k-1} = E\left[\mathbf{w}_{k-1}\mathbf{w}_{k-1}^\mathsf{T}\right]$.

- Given the prediction, $\check{\mathbf{x}}_k$, a gain matrix $\mathbf{K} \in \mathbb{R}^{n_x \times n_y}$, and the measurement $\mathbf{y}_k$, is the correction $\hat{\mathbf{x}}_k = \check{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{y}_k - \check{\mathbf{y}}_k)$ unbiased?
- Unbiased means $E\left[\hat{\mathbf{e}}_k\right] = \mathbf{0}$. Using this definition,

$$\begin{aligned} E\left[\hat{\mathbf{e}}_k\right] &= E\left[\mathbf{x}_k - \hat{\mathbf{x}}_k\right] = E\left[\mathbf{x}_k - \check{\mathbf{x}}_k - \mathbf{K}_k(\mathbf{y}_k - \check{\mathbf{y}}_k)\right] \\ &= E\left[\mathbf{x}_k - \check{\mathbf{x}}_k\right] - \mathbf{K}_k E\left[\mathbf{H}_k\mathbf{x}_k + \mathbf{M}_k\mathbf{v}_k - \mathbf{H}_k\check{\mathbf{x}}_k\right] \\ &= E\left[\mathbf{x}_k - \check{\mathbf{x}}_k\right] - \mathbf{K}_k\mathbf{H}_k E\left[\mathbf{x}_k - \check{\mathbf{x}}_k\right] - \mathbf{K}_k\mathbf{M}_k E\left[\mathbf{v}_k\right] = (\mathbf{1} - \mathbf{K}_k\mathbf{H}_k)E\left[\check{\mathbf{e}}_k\right]. \end{aligned} \quad (2)$$

- Next, note that

$$\begin{aligned} E\left[\check{\mathbf{e}}_k\right] &= E\left[\mathbf{x}_k - \check{\mathbf{x}}_k\right] \\ &= E\left[\mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1} - \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} - \mathbf{G}_{k-1}\mathbf{u}_{k-1}\right] \\ &= \mathbf{F}_{k-1}E\left[\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right] + \mathbf{L}_{k-1}E\left[\mathbf{w}_{k-1}\right] = \mathbf{F}_{k-1}E\left[\hat{\mathbf{e}}_{k-1}\right]. \end{aligned} \quad (3)$$

- Provided $\hat{\mathbf{e}}_0 \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{P}}_0)$,[2]
  - then $E\left[\check{\mathbf{e}}_1\right] = \mathbf{0}$ from (3),
  - then $E\left[\hat{\mathbf{e}}_1\right] = \mathbf{0}$ from (2),
  - then $E\left[\check{\mathbf{e}}_2\right] = \mathbf{0}$ from (3),
  - then $E\left[\hat{\mathbf{e}}_2\right] = \mathbf{0}$ from (2),
  - ...
  - then $E\left[\hat{\mathbf{e}}_k\right] = \mathbf{0}$ from (2), ...
- In turn, the estimate $\hat{\mathbf{x}}_k$ is unbiased.

[2] $\hat{\mathbf{e}}_0 \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{P}}_0)$ does *not* mean that $\hat{\mathbf{e}}_0 = \mathbf{0}$; it means the pdf associated with $\hat{\mathbf{e}}_0$ has zero mean and covariance $\hat{\mathbf{P}}_0$.

An Optimization Problem

- Consider the cost function

$$J_k(\mathbf{K}_k) = \mathrm{tr}(\hat{\mathbf{P}}_k),$$

where $\hat{\mathbf{P}}_k = E\left[\hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^{\mathsf{T}}\right]$, $\hat{\mathbf{e}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$.

Q. Why minimize this cost function as a function of $\mathbf{K}_k$?

A. Doing so minimizes the error covariance, which in turn means minimizing the uncertainty in the state-estimation error.

- First, what is $\hat{\mathbf{P}}_k = E\left[\hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^{\mathsf{T}}\right]$? Using

$$\begin{aligned}
\hat{\mathbf{e}}_k &= \mathbf{x}_k - \hat{\mathbf{x}}_k \\
&= \mathbf{x}_k - \check{\mathbf{x}}_k - \mathbf{K}_k(\mathbf{y}_k - \check{\mathbf{y}}_k) \\
&= \check{\mathbf{e}}_k - \mathbf{K}_k\mathbf{H}_k(\mathbf{x}_k - \check{\mathbf{x}}_k) - \mathbf{K}_k\mathbf{M}_k\mathbf{v}_k \\
&= (\mathbf{1} - \mathbf{K}_k\mathbf{H}_k)\check{\mathbf{e}}_k - \mathbf{K}_k\mathbf{M}_k\mathbf{v}_k \quad \ldots
\end{aligned}$$

- …it follows that

$$
\begin{aligned}
\hat{\mathbf{P}}_k &= E\left[\hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^\mathsf{T}\right] \\
&= E\left[\left((\mathbf{1} - \mathbf{K}_k \mathbf{H}_k)\check{\mathbf{e}}_k - \mathbf{K}_k \mathbf{M}_k \mathbf{v}_k\right)\left(\check{\mathbf{e}}_k^\mathsf{T}(\mathbf{1} - \mathbf{H}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}) - \mathbf{v}_k^\mathsf{T}\mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}\right)\right] \\
&= E\left[(\mathbf{1} - \mathbf{K}_k \mathbf{H}_k)\check{\mathbf{e}}_k \check{\mathbf{e}}_k^\mathsf{T}(\mathbf{1} - \mathbf{H}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}) - (\mathbf{1} - \mathbf{K}_k \mathbf{H}_k)\check{\mathbf{e}}_k \mathbf{v}_k^\mathsf{T}\mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}\right. \\
&\quad \left. - \mathbf{K}_k \mathbf{M}_k \mathbf{v}_k \check{\mathbf{e}}_k^\mathsf{T}(\mathbf{1} - \mathbf{H}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}) + \mathbf{K}_k \mathbf{M}_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T}\mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}\right] \\
&= (\mathbf{1} - \mathbf{K}_k \mathbf{H}_k) E\left[\check{\mathbf{e}}_k \check{\mathbf{e}}_k^\mathsf{T}\right](\mathbf{1} - \mathbf{H}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}) - (\mathbf{1} - \mathbf{K}_k \mathbf{H}_k) E\left[\check{\mathbf{e}}_k \mathbf{v}_k^\mathsf{T}\right]\mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T} \\
&\quad - \mathbf{K}_k \mathbf{M}_k E\left[\mathbf{v}_k \check{\mathbf{e}}_k^\mathsf{T}\right](\mathbf{1} - \mathbf{H}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}) + \mathbf{K}_k \mathbf{M}_k E\left[\mathbf{v}_k \mathbf{v}_k^\mathsf{T}\right]\mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T} \\
&= (\mathbf{1} - \mathbf{K}_k \mathbf{H}_k)\check{\mathbf{P}}_k(\mathbf{1} - \mathbf{K}_k \mathbf{H}_k)^\mathsf{T} + \mathbf{K}_k \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T},
\end{aligned}
$$

where $E\left[\check{\mathbf{e}}_k \mathbf{v}_k^\mathsf{T}\right] = \mathbf{0}$.

- Using a slightly different form of $\hat{\mathbf{P}}_k$,

$$\hat{\mathbf{P}}_k = \check{\mathbf{P}}_k - \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \mathbf{K}_k^\mathsf{T} - \mathbf{K}_k \mathbf{H}_k \check{\mathbf{P}}_k + \mathbf{K}_k \left(\mathbf{H}_k \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} + \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T} \right) \mathbf{K}_k^\mathsf{T},$$

  then computing $\frac{\partial J_k(\mathbf{K})}{\partial \mathbf{K}}$ and setting the result to zero gives

$$\frac{\partial J_k(\mathbf{K})}{\partial \mathbf{K}} = -2\check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} + 2\mathbf{K}_k \left(\mathbf{H}_k \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} + \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T} \right) = \mathbf{0}.$$

- Rearranging, and solving for $\mathbf{K}_k$, results in

$$\mathbf{K}_k \left(\mathbf{H}_k \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} + \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T} \right) = \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T},$$

$$\mathbf{K}_k = \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \left(\mathbf{H}_k \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} + \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T} \right)^{-1}. \quad (4)$$

- $\mathbf{K}_k$ is called the *Kalman gain*.
- The inverse in (4) always exists. Why?

An alternate form of the Kalman Gain

- The filter innovation is $\check{\boldsymbol{\rho}}_k = \mathbf{y}_k - \check{\mathbf{y}}_k = \mathbf{H}_k(\mathbf{x}_k - \check{\mathbf{x}}_k) + \mathbf{M}_k\mathbf{v}_k = \mathbf{H}_k\check{\mathbf{e}}_k + \mathbf{M}_k\mathbf{v}_k$.
- Consider

$$\begin{aligned}
\check{\mathbf{P}}_k^{\mathbf{x}_k\mathbf{y}_k} &= E\left[\check{\mathbf{e}}_k\check{\boldsymbol{\rho}}_k^{\mathsf{T}}\right] \\
&= E\left[\check{\mathbf{e}}_k(\mathbf{H}_k\check{\mathbf{e}}_k + \mathbf{M}_k\mathbf{v}_k)^{\mathsf{T}}\right] \\
&= E\left[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^{\mathsf{T}}\right]\mathbf{H}_k^{\mathsf{T}} + E\left[\check{\mathbf{e}}_k\mathbf{v}_k^{\mathsf{T}}\right]\mathbf{M}_k^{\mathsf{T}} \\
&= \check{\mathbf{P}}_k\mathbf{H}_k^{\mathsf{T}},
\end{aligned}$$

  where $E\left[\check{\mathbf{e}}_k\mathbf{v}_k^{\mathsf{T}}\right] = \mathbf{0}$ and $E\left[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^{\mathsf{T}}\right] = \check{\mathbf{P}}_k$.
- Similarly,

$$\begin{aligned}
\check{\mathbf{P}}_k^{\mathbf{y}_k\mathbf{y}_k} &= E\left[\check{\boldsymbol{\rho}}_k\check{\boldsymbol{\rho}}_k^{\mathsf{T}}\right] \\
&= E\left[(\mathbf{H}_k\check{\mathbf{e}}_k + \mathbf{M}_k\mathbf{v}_k)(\mathbf{H}_k\check{\mathbf{e}}_k + \mathbf{M}_k\mathbf{v}_k)^{\mathsf{T}}\right] \\
&= \mathbf{H}_k E\left[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^{\mathsf{T}}\right]\mathbf{H}_k^{\mathsf{T}} + \mathbf{H}_k E\left[\check{\mathbf{e}}_k\mathbf{v}_k^{\mathsf{T}}\right]\mathbf{M}_k^{\mathsf{T}} + \mathbf{M}_k E\left[\mathbf{v}_k\check{\mathbf{e}}_k^{\mathsf{T}}\right]\mathbf{H}_k^{\mathsf{T}} + \mathbf{M}_k E\left[\mathbf{v}_k\mathbf{v}_k^{\mathsf{T}}\right]\mathbf{M}_k^{\mathsf{T}} \\
&= \mathbf{H}_k\check{\mathbf{P}}_k\mathbf{H}_k^{\mathsf{T}} + \mathbf{M}_k\mathbf{R}_k\mathbf{M}_k^{\mathsf{T}},
\end{aligned}$$

  where $E\left[\check{\mathbf{e}}_k\mathbf{v}_k^{\mathsf{T}}\right] = \mathbf{0}$, $\check{\mathbf{P}}_k = E\left[\check{\mathbf{e}}_k\check{\mathbf{e}}_k^{\mathsf{T}}\right]$, and $\mathbf{R}_k = E\left[\mathbf{v}_k\mathbf{v}_k^{\mathsf{T}}\right]$.

- It follows that

$$\begin{aligned}\mathbf{K}_k &= \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \left(\mathbf{H}_k \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} + \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T} \right)^{-1} \\ &= \check{\mathbf{P}}_k^{\mathbf{x}_k \mathbf{y}_k} \check{\mathbf{P}}_k^{\mathbf{y}_k \mathbf{y}_k \, -1}.\end{aligned}$$

- The *a posteriori* covariance can be written as

$$\begin{aligned}\hat{\mathbf{P}}_k &= \check{\mathbf{P}}_k - \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \mathbf{K}_k^\mathsf{T} - \mathbf{K}_k \mathbf{H}_k \check{\mathbf{P}}_k + \mathbf{K}_k \left(\mathbf{H}_k \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} + \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T} \right) \mathbf{K}_k^\mathsf{T} \\ &= \check{\mathbf{P}}_k - \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \mathbf{K}_k^\mathsf{T} - \mathbf{K}_k \mathbf{H}_k \check{\mathbf{P}}_k + \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \mathbf{K}_k^\mathsf{T} \\ &= \check{\mathbf{P}}_k - \mathbf{K}_k \mathbf{H}_k \check{\mathbf{P}}_k \\ &= \left(\mathbf{1} - \mathbf{K}_k \mathbf{H}_k \right) \check{\mathbf{P}}_k.\end{aligned}$$

- Also, note that

$$\begin{aligned}\hat{\mathbf{P}}_k &= \check{\mathbf{P}}_k - \mathbf{K}_k \left(\check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \right)^\mathsf{T} \\ &= \check{\mathbf{P}}_k - \check{\mathbf{P}}_k^{\mathbf{x}_k \mathbf{y}_k} \check{\mathbf{P}}_k^{\mathbf{y}_k \mathbf{y}_k \, -1} \check{\mathbf{P}}_k^{\mathbf{x}_k \mathbf{y}_k^\mathsf{T}}.\end{aligned}$$

- A comment on $E\left[\mathbf{w}_{k-1}\mathbf{e}_{k-1}^{\mathsf{T}}\right] = \mathbf{0}$.
  - Note that $\mathbf{w}_{k-1}$ impacts $\mathbf{x}_k$ via $\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1}$, and $\mathbf{w}_{k-2}$ impacts $\mathbf{x}_{k-1}$ via $\mathbf{x}_{k-1} = \mathbf{F}_{k-2}\mathbf{x}_{k-2} + \mathbf{L}_{k-2}\mathbf{w}_{k-2}$.
  - $\mathbf{w}_{k-1}$ does *not* impact $\mathbf{x}_{k-1}$; $\mathbf{L}_{k-1}\mathbf{w}_{k-1}$ is added to $\mathbf{F}_{k-1}\mathbf{x}_{k-1}$.
  - Because $\mathbf{e}_{k-1} = \mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}$, and $\mathbf{x}_{k-1}$ is impacted by $\mathbf{w}_{k-2}$ and *not* $\mathbf{w}_{k-1}$, it follows that $\mathbf{e}_{k-1}$ and $\mathbf{w}_{k-1}$ are uncorrelated.
  - Thus, $E\left[\mathbf{w}_{k-1}\mathbf{e}_{k-1}^{\mathsf{T}}\right] = \mathbf{0}$.
- A comment on $E\left[\check{\mathbf{e}}_k\mathbf{v}_k^{\mathsf{T}}\right] = \mathbf{0}$.
  - $\check{\mathbf{e}}_k = \mathbf{x}_k - \check{\mathbf{x}}_k = \mathbf{F}_{k-1}\mathbf{e}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1}$ where $\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1}$ and $\check{\mathbf{x}}_k = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1}$.
  - $\check{\mathbf{e}}_k$ is impacted by $\mathbf{w}_{k-1}$, but not by $\mathbf{v}_k$.
  - $\mathbf{w}_{k-1}$ and $\mathbf{v}_k^{\mathsf{T}}$ are uncorrelated, thus $\check{\mathbf{e}}_k$ and $\mathbf{v}_k$ are uncorrelated.
  - It follows that $E\left[\check{\mathbf{e}}_k\mathbf{v}_k^{\mathsf{T}}\right] = \mathbf{0}$.

Summary of the Kalman Filter

System:

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1} \\
\mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{M}_k\mathbf{v}_k \\
\mathbf{w}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k) \\
\mathbf{v}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)
\end{aligned}$$

Initialization:

$$\begin{aligned}
\hat{\mathbf{x}}_0 &= E\left[\mathbf{x}_0\right] \\
\hat{\mathbf{P}}_0 &= E\left[\left(\mathbf{x}_0 - \hat{\mathbf{x}}_0\right)\left(\mathbf{x}_0 - \hat{\mathbf{x}}_0\right)^\mathsf{T}\right]
\end{aligned}$$

Prediction:

$$\begin{aligned}
\check{\mathbf{x}}_k &= \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} \\
\check{\mathbf{P}}_k &= \mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^\mathsf{T} + \mathbf{L}_{k-1}\mathbf{Q}_{k-1}\mathbf{L}_{k-1}^\mathsf{T}
\end{aligned}$$

Correction:

$$\begin{aligned}
\mathbf{V}_k &= \mathbf{H}_k\check{\mathbf{P}}_k\mathbf{H}_k^\mathsf{T} + \mathbf{M}_k\mathbf{R}_k\mathbf{M}_k^\mathsf{T} \\
\mathbf{K}_k &= \check{\mathbf{P}}_k\mathbf{H}_k^\mathsf{T}\mathbf{V}_k^{-1} \\
\hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{y}_k - \check{\mathbf{y}}_k) \\
\hat{\mathbf{P}}_k &= (\mathbf{1} - \mathbf{K}_k\mathbf{H}_k)\check{\mathbf{P}}_k(\mathbf{1} - \mathbf{K}_k\mathbf{H}_k)^\mathsf{T} + \mathbf{K}_k\mathbf{M}_k\mathbf{R}_k\mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T} \\
&= \check{\mathbf{P}}_k - \mathbf{K}_k\mathbf{H}_k\check{\mathbf{P}}_k
\end{aligned}$$

Derivation of the Extended Kalman Filter (EKF)

- Consider a discrete-time system described by nonlinear process and measurement (observation) models,

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}), & \mathbf{w}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \\ \mathbf{y}_k &= \mathbf{g}_k(\mathbf{x}_k, \mathbf{v}_k), & \mathbf{v}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k). \end{aligned}$$

- To derive the EKF the nonlinear discrete-time system is linearized.
- Perform a Taylor series expansion in $\mathbf{x}_k$, $\mathbf{w}_k$, and $\mathbf{v}_k$ about some nominal $\bar{\mathbf{x}}_k$, $\bar{\mathbf{w}}_k$, $\bar{\mathbf{v}}_k$ such that

$$\begin{aligned} \mathbf{x}_k &= \bar{\mathbf{x}}_k + \delta \mathbf{x}_k, \\ \mathbf{w}_k &= \bar{\mathbf{w}}_k + \delta \mathbf{w}_k, \\ \mathbf{v}_k &= \bar{\mathbf{v}}_k + \delta \mathbf{v}_k, \end{aligned}$$

  where $\delta \mathbf{x}_k$, $\delta \mathbf{w}_k$, and $\delta \mathbf{v}_k$ are perturbations.
- To be consistent with the assumed disturbance and noise (i.e., the expected value of the disturbance and noise), $\bar{\mathbf{w}}_k$ and $\bar{\mathbf{v}}_k$ are both zero, that is, $\bar{\mathbf{w}}_k = \mathbf{0}$ and $\bar{\mathbf{v}}_k = \mathbf{0}$.

- Perturbing the process model,

$$\mathbf{x}_k = \bar{\mathbf{x}}_k + \delta\mathbf{x}_k = \mathbf{f}_{k-1}(\bar{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \bar{\mathbf{w}}_{k-1}) + \mathbf{F}_{k-1}\delta\mathbf{x}_{k-1} + \mathbf{L}_{k-1}\delta\mathbf{w}_{k-1} + \text{H.O.T.},$$

where

$$\mathbf{F}_{k-1} = \left.\frac{\partial \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1})}{\partial \mathbf{x}_{k-1}}\right|_{\bar{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \bar{\mathbf{w}}_{k-1}},$$

$$\mathbf{L}_{k-1} = \left.\frac{\partial \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1})}{\partial \mathbf{w}_{k-1}}\right|_{\bar{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \bar{\mathbf{w}}_{k-1}}.$$

- Perturbing the measurement model,

$$\mathbf{y}_k = \bar{\mathbf{y}}_k + \delta\mathbf{y}_k = \mathbf{g}_k(\bar{\mathbf{x}}_k, \bar{\mathbf{v}}_k) + \mathbf{H}_k\delta\mathbf{x}_k + \mathbf{M}_k\delta\mathbf{v}_k + \text{H.O.T.},$$

where

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{g}_k(\mathbf{x}_k, \mathbf{v}_k)}{\partial \mathbf{x}_k}\right|_{\bar{\mathbf{x}}_k, \bar{\mathbf{v}}_k},$$

$$\mathbf{M}_k = \left.\frac{\partial \mathbf{g}_k(\mathbf{x}_k, \mathbf{v}_k)}{\partial \mathbf{v}_k}\right|_{\bar{\mathbf{x}}_k, \bar{\mathbf{v}}_k}.$$

- Note $\mathbf{L}_k$ and $\mathbf{M}_k$ must be full column and row rank, respectively.

- Using $\mathbf{x}_k = \bar{\mathbf{x}}_k + \delta\mathbf{x}_k$ and $\mathbf{w}_k = \bar{\mathbf{w}}_k + \delta\mathbf{w}_k = \mathbf{0} + \delta\mathbf{w}_k$, and dropping H.O.T., rewrite the linearized process model as

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{f}_{k-1}(\bar{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \mathbf{0}) + \mathbf{F}_{k-1}\delta\mathbf{x}_{k-1} + \mathbf{L}_{k-1}\delta\mathbf{w}_{k-1} \\
&= \mathbf{f}_{k-1}(\bar{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \mathbf{0}) + \mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \bar{\mathbf{x}}_{k-1}) + \mathbf{L}_{k-1}\mathbf{w}_{k-1} \\
&= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \underbrace{\mathbf{f}_{k-1}(\bar{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \mathbf{0}) - \mathbf{F}_{k-1}\bar{\mathbf{x}}_{k-1}}_{\boldsymbol{u}_{k-1}} + \mathbf{L}_{k-1}\mathbf{w}_{k-1} \\
&= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{u}_{k-1} + \mathbf{L}_{k-1}\mathbf{w}_{k-1},
\end{aligned}$$

  where $\boldsymbol{u}_{k-1}$ is *known*.
- In a similar fashion, using $\mathbf{x}_k = \bar{\mathbf{x}}_k + \delta\mathbf{x}_k$ and $\mathbf{v}_k = \bar{\mathbf{v}}_k + \delta\mathbf{v}_k = \mathbf{0} + \delta\mathbf{v}_k$, and dropping H.O.T., rewrite the linearized measurement model as

$$\begin{aligned}
\mathbf{y}_k &= \mathbf{g}_k(\bar{\mathbf{x}}_k, \mathbf{0}) + \mathbf{H}_k\delta\mathbf{x}_k + \mathbf{M}_k\delta\mathbf{v}_k \\
&= \mathbf{g}_k(\bar{\mathbf{x}}_k, \mathbf{0}) + \mathbf{H}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k) + \mathbf{M}_k\mathbf{v}_k \\
&= \mathbf{H}_k\mathbf{x}_k + \underbrace{\mathbf{g}_k(\bar{\mathbf{x}}_k, \mathbf{0}) - \mathbf{H}_k\bar{\mathbf{x}}_k}_{\boldsymbol{\beta}_k} + \mathbf{M}_k\mathbf{v}_k \\
&= \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\beta}_k + \mathbf{M}_k\mathbf{v}_k,
\end{aligned}$$

  where $\boldsymbol{\beta}_k$ is *known*.

The Prediction Step

- The prediction step is

$$\check{\mathbf{x}}_k = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \boldsymbol{u}_{k-1},$$
$$\check{\mathbf{P}}_k = \mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^{\mathsf{T}} + \mathbf{L}_{k-1}\mathbf{Q}_{k-1}\mathbf{L}_{k-1}^{\mathsf{T}},$$

  where $\mathbf{F}_{k-1}$, $\boldsymbol{u}_{k-1}$, and $\mathbf{L}_{k-1}$ are evaluated at the best prior estimate of the state, $\hat{\mathbf{x}}_{k-1}$ (i.e., $\hat{\mathbf{x}}_{k-1}$ replaces $\bar{\mathbf{x}}_{k-1}$ in $\mathbf{F}_{k-1}$, $\boldsymbol{u}_{k-1}$, and $\mathbf{L}_{k-1}$).
- The computation of $\check{\mathbf{x}}_k$ above is equivalent to

$$\begin{aligned}\check{\mathbf{x}}_k &= \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \boldsymbol{u}_{k-1} \\ &= \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \left(\mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \mathbf{0}) - \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1}\right) \\ &= \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \mathbf{0})\end{aligned}$$

  which is just the nonlinear discrete time process model evaluated at $\hat{\mathbf{x}}_{k-1}$, $\mathbf{u}_{k-1}$, and $\mathbf{w}_{k-1} = \mathbf{0}$.
  - As with the Kalman filter, we perform a prediction step using the expected value of the disturbance, $\mathbf{w}_{k-1} = \mathbf{0}$.
  - It appears we are ignoring the disturbance, but we are not; if $\mathbf{w}_{k-1} \sim \mathcal{N}(\tilde{\mathbf{w}}_{k-1}, \mathbf{Q}_{k-1})$ then the prediction would be $\check{\mathbf{x}}_k = \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \tilde{\mathbf{w}}_{k-1})$.

The Correction Step

- The correction is given by

$$\begin{aligned}
\mathbf{V}_k &= \mathbf{H}_k \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} + \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T}, \\
\mathbf{K}_k &= \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \mathbf{V}_k^{-1}, \\
\hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \mathbf{K}_k (\mathbf{y}_k - \check{\mathbf{y}}_k), \\
\hat{\mathbf{P}}_k &= (\mathbf{1} - \mathbf{K}_k \mathbf{H}_k) \check{\mathbf{P}}_k (\mathbf{1} - \mathbf{K}_k \mathbf{H}_k)^\mathsf{T} + \mathbf{K}_k \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^\mathsf{T} \mathbf{K}_k^\mathsf{T} \\
&= \check{\mathbf{P}}_k - \mathbf{K}_k \mathbf{H}_k \check{\mathbf{P}}_k - \check{\mathbf{P}}_k \mathbf{H}_k^\mathsf{T} \mathbf{K}_k^\mathsf{T} + \mathbf{K}_k \mathbf{V}_k \mathbf{K}_k^\mathsf{T},
\end{aligned}$$

  where $\mathbf{H}_k$ and $\mathbf{M}_k$ are evaluated at $\check{\mathbf{x}}_k$ (i.e., $\check{\mathbf{x}}_k$ replaces $\bar{\mathbf{x}}_k$ in $\mathbf{H}_k$ and $\mathbf{M}_k$).
- The predicted measurement $\check{\mathbf{y}}_k$ is

$$\check{\mathbf{y}}_k = \mathbf{H}_k \check{\mathbf{x}}_k + \check{\boldsymbol{\beta}}_k,$$

  where $\mathbf{H}_k$ and $\check{\boldsymbol{\beta}}_k$ are evaluated at $\check{\mathbf{x}}_k$.

- The prediction measurement is equivalent to

$$\begin{aligned} \check{\mathbf{y}}_k &= \mathbf{H}_k \check{\mathbf{x}}_k + \check{\boldsymbol{\beta}}_k \\ &= \mathbf{H}_k \check{\mathbf{x}}_k + (\mathbf{g}_k(\check{\mathbf{x}}_k, \mathbf{0}) - \mathbf{H}_k \check{\mathbf{x}}_k) \\ &= \mathbf{g}_k(\check{\mathbf{x}}_k, \mathbf{0}), \end{aligned}$$

  the nonlinear discrete-time measurement model evaluated at $\check{\mathbf{x}}_k$, the a priori state estimate.
  - Again, we perform the correction step using the expected value of the noise, $\mathbf{v}_k = \mathbf{0}$.
  - It appears we are ignoring the noise, but we are not; if $\mathbf{v}_k \sim \mathcal{N}(\tilde{\mathbf{v}}_k, \mathbf{R}_k)$ then the correction would be $\check{\mathbf{y}}_k = \mathbf{g}_k(\check{\mathbf{x}}_k, \tilde{\mathbf{v}}_k)$.
- The correction is then also given by

$$\begin{aligned} \hat{\mathbf{x}}_k &= \check{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{y}_k - \check{\mathbf{y}}_k), \\ &= \check{\mathbf{x}}_k + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{g}_k(\check{\mathbf{x}}_k, \mathbf{0})\right). \end{aligned}$$

Summary of the Extended Kalman Filter

$$
\begin{array}{rrcl}
\text{System:} & \mathbf{x}_k & = & \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \\
 & \mathbf{y}_k & = & \mathbf{g}_k(\mathbf{x}_k, \mathbf{v}_k) \\
 & \mathbf{w}_k & \sim & \mathcal{N}(\mathbf{0}, \mathbf{Q}_k) \\
 & \mathbf{v}_k & \sim & \mathcal{N}(\mathbf{0}, \mathbf{R}_k) \\
\text{Initialization:} & \hat{\mathbf{x}}_0 & = & E\left[\mathbf{x}_0\right] \\
 & \hat{\mathbf{P}}_0 & = & E\left[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^\mathsf{T}\right] \\
\text{Prediction:} & \check{\mathbf{x}}_k & = & \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \mathbf{0}) \\
 & \check{\mathbf{P}}_k & = & \mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^\mathsf{T} + \mathbf{L}_{k-1}\mathbf{Q}_{k-1}\mathbf{L}_{k-1}^\mathsf{T} \\
\text{Correction:} & \mathbf{V}_k & = & \mathbf{H}_k\check{\mathbf{P}}_k\mathbf{H}_k^\mathsf{T} + \mathbf{M}_k\mathbf{R}_k\mathbf{M}_k^\mathsf{T} \\
 & \mathbf{K}_k & = & \check{\mathbf{P}}_k\mathbf{H}_k^\mathsf{T}\mathbf{V}_k^{-1} \\
 & \hat{\mathbf{x}}_k & = & \check{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{y}_k - \mathbf{g}_k(\check{\mathbf{x}}_k, \mathbf{0})) \\
 & \hat{\mathbf{P}}_k & = & (\mathbf{1} - \mathbf{K}_k\mathbf{H}_k)\check{\mathbf{P}}_k(\mathbf{1} - \mathbf{K}_k\mathbf{H}_k)^\mathsf{T} + \mathbf{K}_k\mathbf{M}_k\mathbf{R}_k\mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T} \\
 & & = & \check{\mathbf{P}}_k - \mathbf{K}_k\mathbf{H}_k\check{\mathbf{P}}_k
\end{array}
$$

The Iterative EKF

- Recall that

$$\mathbf{H}_k = \left. \frac{\partial \mathbf{g}_k(\mathbf{x}_k, \mathbf{v}_k)}{\partial \mathbf{x}_k} \right|_{\check{\mathbf{x}}_k, \mathbf{0}},$$

$$\mathbf{M}_k = \left. \frac{\partial \mathbf{g}_k(\mathbf{x}_k, \mathbf{v}_k)}{\partial \mathbf{v}_k} \right|_{\check{\mathbf{x}}_k, \mathbf{0}},$$

  which is to say that $\mathbf{H}_k$ and $\mathbf{M}_k$ are computed using $\check{\mathbf{x}}_k$ after the prediction step.
- Well, after the correction step we have a better estimate of the state, namely $\hat{\mathbf{x}}_k$.
- The idea behind the iterative EKF is to recompute $\mathbf{H}_k$ and $\mathbf{M}_k$ using a better estimate of the state, then recompute $\mathbf{K}_k$, and then finally recompute $\hat{\mathbf{x}}_k$ and $\hat{\mathbf{P}}_k$.
- This process is repeated until convergence.

Step-by-Step Details

1. Execute the prediction step normally, that is,

$$\check{\mathbf{x}}_k = \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, \mathbf{0}),$$
$$\check{\mathbf{P}}_k = \mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^\mathsf{T} + \mathbf{L}_{k-1}\mathbf{Q}_{k-1}\mathbf{L}_{k-1}^\mathsf{T},$$

   and set the linearization point to $\hat{\mathbf{x}}_{k,\mathrm{lin}} = \check{\mathbf{x}}_k$.

2. Compute

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{g}_k(\mathbf{x}_k, \mathbf{v}_k)}{\partial \mathbf{x}_k}\right|_{\hat{\mathbf{x}}_{k,\mathrm{lin}}, \mathbf{0}},$$
$$\mathbf{M}_k = \left.\frac{\partial \mathbf{g}_k(\mathbf{x}_k, \mathbf{v}_k)}{\partial \mathbf{v}_k}\right|_{\hat{\mathbf{x}}_{k,\mathrm{lin}}, \mathbf{0}}.$$

3. Compute

$$\mathbf{K}_k = \check{\mathbf{P}}_k\mathbf{H}_k^\mathsf{T}\left(\mathbf{H}_k\check{\mathbf{P}}_k\mathbf{H}_k^\mathsf{T} + \mathbf{M}_k\mathbf{R}_k\mathbf{M}_k^\mathsf{T}\right)^{-1},$$
$$\hat{\mathbf{x}}_k = \check{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{y}_k - (\mathbf{g}(\hat{\mathbf{x}}_{k,\mathrm{lin}}, \mathbf{0}) + \mathbf{H}_k(\check{\mathbf{x}}_k - \hat{\mathbf{x}}_{k,\mathrm{lin}}))).$$

4. - If $\|\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{k,\mathrm{lin}}\|_2 \geq \epsilon$ set $\hat{\mathbf{x}}_{k,\mathrm{lin}} = \hat{\mathbf{x}}_k$ and go back to Step 2.
   - If $\|\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{k,\mathrm{lin}}\|_2 < \epsilon$ go to time step $k+1$.

5. Compute

$$\hat{\mathbf{P}}_k = (\mathbf{1} - \mathbf{K}_k\mathbf{H}_k)\check{\mathbf{P}}_k(\mathbf{1} - \mathbf{K}_k\mathbf{H}_k)^\mathsf{T} + \mathbf{K}_k\mathbf{M}_k\mathbf{R}_k\mathbf{M}_k^\mathsf{T}\mathbf{K}_k^\mathsf{T}$$

Questions

Thank you for your attention.

Questions?

james.richard.forbes@mcgill.ca

Presentation created using LaTeX and Beamer.

References

Material herein is based on [1, 2, 3, 4, 5, 6, 7].

[1] T. D. Barfoot, “State estimation for aerospace vehicles; AER1513 course notes and experimental dataset,” *University of Toronto Institute for Aerospace Studies,* AER1513-002, Rev: 1.0, December 2009.

[2] R. Eustice, “NA568/EECS568 mobile robotics: Methods and algorithms,” *University of Michigan*, 2015.

[3] T. D. Barfoot, *State Estimation for Robotics*. New York, NY: Cambridge University Press, 2017.

[4] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York, NY: Academic Press, 1970.

[5] D. Simon, *Optimal State Estimation*. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

[6] R. F. Stengel, *Optimal Control and Estimation*. New York, NY: Dover, 1994.

[7] C. E. Rasmussen and C. K. Williams, *Gaussian processes for Machine Learning*. MIT press Cambridge, MA, 2006.