# A Geometric Interpretation of Kalman Filters

Liam P.

2024-1-29

# Outline

# Topic

# Physical Problem and Pictures

Given:

- An initial random vector $x_0 \in \mathbb{R}^n$ with covariance $C_0 = E[x_0 x_0^T]$
- Measurements $\{z_i\}_{i=1}^k \subset \mathbb{R}^m$
- a linear relationship between $x_i$ and $z_i$, as well as $x_i$ and $x_{i+1}$:

$$z_i = M_i x_i + \delta_i$$
$$x_{i+1} = A_i x_i + \epsilon_i$$

where $\epsilon_i, \delta_i$ white Gaussian noise with covariance $Q_i$, $R_i$ resp.

- An estimate $x_{k+1|k}$ is called a best least-squares estimate of $x_{k+1}$ out of random vectors in $M$ if

$$\inf_{y \in M} \mathbb{E}[\|x_{k+1} - y\|^2] = \mathbb{E}[\|x_{k+1} - x_{k+1|k}\|^2]$$

# Physical Problem and Pictures

Given:

- An initial random vector $x_0 \in \mathbb{R}^n$ with covariance $C_0 = E[x_0 x_0^T]$
- Measurements $\{z_i\}_{i=1}^k \subset \mathbb{R}^m$
- a linear relationship between $x_i$ and $z_i$, as well as $x_i$ and $x_{i+1}$:

$$z_i = M_i x_i + \delta_i$$
$$x_{i+1} = A_i x_i + \epsilon_i$$

where $\epsilon_i, \delta_i$ white Gaussian noise with covariance $Q_i$, $R_i$ resp.

- An estimate $x_{k+1|k}$ is called a best least-squares estimate of $x_{k+1}$ out of random vectors in $M$ if

$$\inf_{y \in M} \mathbb{E}[\|x_{k+1} - y\|^2] = \mathbb{E}[\|x_{k+1} - x_{k+1|k}\|^2]$$

# Physical Problem and Pictures

Given:

- An initial random vector $\mathrm{x}_0 \in \mathbb{R}^n$ with covariance $C_0 = E[x_0 x_0^T]$
- Measurements $\{\mathrm{z}_i\}_{i=1}^k \subset \mathbb{R}^m$
- a linear relationship between $\mathrm{x}_i$ and $\mathrm{z}_i$, as well as $\mathrm{x}_i$ and $\mathrm{x}_{i+1}$:

$$\begin{aligned} \mathrm{z}_i &= M_i \mathrm{x}_i + \delta_i \\ \mathrm{x}_{i+1} &= A_i \mathrm{x}_i + \epsilon_i \end{aligned}$$

where $\epsilon_i, \delta_i$ white Gaussian noise with covariance $Q_i$, $R_i$ resp.

- An estimate $\mathrm{x}_{k+1|k}$ is called a best least-squares estimate of $\mathrm{x}_{k+1}$ out of random vectors in $M$ if

$$\inf_{y \in M} \mathbb{E}[\|x_{k+1} - y\|^2] = \mathbb{E}[\|x_{k+1} - x_{k+1|k}\|^2]$$

# The Optimal Least-Squares Solution (Kalman)

## Theorem

*(Kalman): Let $x_{0|-1} = x_0$. The best expected least-squares estimate of $x_{k+1|k}$ and $C_{k+1|k}$ is given recursively by*

1.) *(Update)*

$$x_{k|k} = x_{k|k-1} + C_{k|k-1} H_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} (z_k - M_k x_{k|k-1})$$

$$C_{k|k} = C_{k|k-1} - C_{k|k-1} M_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} M_k C_{k|k-1}^T$$

2.) *(Predict)*

$$x_{k+1|n} = A_k x_{k|k}$$

$$C_{k+1|n} = A_k C_{k|k} A_k^T + Q_k$$

# The Optimal Least-Squares Solution (Kalman)

## Theorem

*(Kalman): Let $x_{0|-1} = x_0$. The best expected least-squares estimate of $x_{k+1|k}$ and $C_{k+1|k}$ is given recursively by*

1.) *(Update)*

$$x_{k|k} = x_{k|k-1} + C_{k|k-1} H_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} \left( z_k - M_k x_{k|k-1} \right)$$

$$C_{k|k} = C_{k|k-1} - C_{k|k-1} M_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} M_k C_{k|k-1}^T$$

2.) *(Predict)*

$$x_{k+1|n} = A_k x_{k|k}$$

$$C_{k+1|n} = A_k C_{k|k} A_k^T + Q_k$$

# The Optimal Least-Squares Solution (Kalman)

## Theorem

*(Kalman): Let $x_{0|-1} = x_0$. The best expected least-squares estimate of $x_{k+1|k}$ and $C_{k+1|k}$ is given recursively by*

1.) *(Update)*

$$x_{k|k} = x_{k|k-1} + C_{k|k-1} H_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} \left( z_k - M_k x_{k|k-1} \right)$$
$$C_{k|k} = C_{k|k-1} - C_{k|k-1} M_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} M_k C_{k|k-1}^T$$

2.) *(Predict)*

$$x_{k+1|n} = A_k x_{k|k}$$
$$C_{k+1|n} = A_k C_{k|k} A_k^T + Q_k$$

# Topic

# Basic Definitions

**Definition**

$(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a *Hilbert Space* if it satisfies the following conditions:

- $\mathcal{H}$ is vector space
- $\mathcal{H}$ is equipped with an inner product $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$
- The norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ induces a complete metric space topology on $\mathcal{H}$

# Examples and Non-examples

Examples:

- $\mathbb{R}^n$ for every $n$ equipped with the usual dot-product $\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$
- The space $\ell^2$ of square-summable sequences $x = (x_1, x_2, \cdots, x_n, \cdots)$ equipped with the inner product $\langle x, y\rangle = \sum_{i=1}^{\infty} x_i y_i$
- The space $L^2(\Omega)$ of square $\mu$ -integrable measurable real-valued functions on $\Omega$, i.e., $f : \Omega \to \mathbb{R}$ measurable and $\int_\Omega |f|^2 d\mu$, with inner product $\langle f, g\rangle = \int_\Omega fg \, d\mu$

# Examples and Non-examples

Examples:

- $\mathbb{R}^n$ for every $n$ equipped with the usual dot-product $\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$
- The space $\ell^2$ of square-summable sequences $x = (x_1, x_2, \cdots, x_n, \cdots)$ equipped with the inner product $\langle x, y\rangle = \sum_{i=1}^{\infty} x_i y_i$
- The space $L^2(\Omega)$ of square $\mu$ -integrable measurable real-valued functions on $\Omega$, i.e., $f : \Omega \to \mathbb{R}$ measurable and $\int_\Omega |f|^2 d\mu$, with inner product $\langle f, g\rangle = \int_\Omega fg \, d\mu$

# Examples and Non-examples

Examples:

- $\mathbb{R}^n$ for every $n$ equipped with the usual dot-product $\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$
- The space $\ell^2$ of square-summable sequences $x = (x_1, x_2, \cdots, x_n, \cdots)$ equipped with the inner product $\langle x, y\rangle = \sum_{i=1}^{\infty} x_i y_i$
- The space $L^2(\Omega)$ of square $\mu$ -integrable measurable real-valued functions on $\Omega$, i.e., $f : \Omega \to \mathbb{R}$ measurable and $\int_\Omega |f|^2 d\mu$, with inner product $\langle f, g\rangle = \int_\Omega fg \, d\mu$

## Our Hilbert Space

We need to endow our problem with a vector-space structure, and an inner product:

- Objects: Random vectors $\mathrm{x} = [x_1\ x_2\ \cdots\ x_n]^T$ where $x_i$ are each random variables with $\mathbb{E}[x_i^2] < \infty$
- Inner product: $\langle \mathrm{x}, \mathrm{y} \rangle_{\mathbb{E}^n} = \mathbb{E}[x \cdot y] = \sum_{i=1}^{n} \mathbb{E}[x_i y_i]$

In particular: the minimum norm problem $\inf_{y \in M} \mathbb{E}[\|y - x\|^2]$ can be expressed as $\inf_{y \in M} \|y - x\|_{\mathbb{E}^n}^2$.

# Our Hilbert Space

We need to endow our problem with a vector-space structure, and an inner product:

- Objects: Random vectors $\mathrm{x} = [x_1 \; x_2 \; \cdots \; x_n]^T$ where $x_i$ are each random variables with $\mathbb{E}[x_i^2] < \infty$
- Inner product: $\langle \mathrm{x}, \mathrm{y} \rangle_{\mathbb{E}^n} = \mathbb{E}[x \cdot y] = \sum_{i=1}^{n} \mathbb{E}[x_i y_i]$

In particular: the minimum norm problem $\inf_{y \in M} \mathbb{E}[\|y - x\|^2]$ can be expressed as $\inf_{y \in M} \|y - x\|_{\mathbb{E}^n}^2$.

# Our Hilbert Space

We need to endow our problem with a vector-space structure, and an inner product:

- Objects: Random vectors $\mathrm{x} = [x_1 \; x_2 \; \cdots \; x_n]^T$ where $x_i$ are each random variables with $\mathbb{E}[x_i^2] < \infty$
- Inner product: $\langle \mathrm{x}, \mathrm{y} \rangle_{\mathbb{E}^n} = \mathbb{E}[x \cdot y] = \sum_{i=1}^{n} \mathbb{E}[x_i y_i]$

In particular: the minimum norm problem $\inf_{y \in M} \mathbb{E}[\|y - x\|^2]$ can be expressed as $\inf_{y \in M} \|y - x\|_{\mathbb{E}^n}^2$.

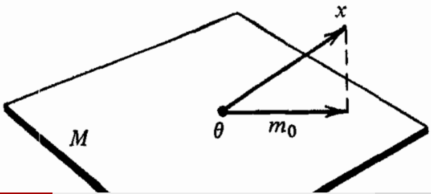
# What We'll Need: The Projection Theorem

## Theorem (Orthogonal Projections Exist)

*Let $M \subseteq \mathcal{H}$ be a closed subspace of $\mathcal{H}$. Let $x \in \mathcal{H}$ be any vector. Then there exists a unique $m_0 \in M$ which attains a minimum distance to $x$:*

$$\inf_{m \in M} \|x - m\| = \|x - m_0\|$$

*Furthermore, the error vector $x - m_0$ is perpendicular to $M$, i.e., for all $m \in M$, $\langle x - m_0, m \rangle = 0$.*



# Topic

# The Projection we Need

Let $y = Ax + \epsilon$ where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ are random vectors and

$$E[\epsilon\epsilon^T] = Q, \quad E[xx^T] = C, \quad E[\epsilon x^T] = 0.$$

- Idea: We know $y$ and $A$, but we want $x$
- Estimate components $x$ as linear combinations of components of $y$

$$x_i = \sum_{j=1}^{m} a_j y_j$$

- Then looking for closest $\hat{x} \in M = \{Ky \,:\, K \in \mathbb{R}^{n\times m}\}$
  - (can write $M$ as $\text{span}(y_i e_j)_{i\in\{1,\cdots,m\},j\in\{1,\cdots,n\}}$)

# The Projection we Need

Let $y = Ax + \epsilon$ where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ are random vectors and

$$E[\epsilon\epsilon^T] = Q, \quad E[xx^T] = C, \quad E[\epsilon x^T] = 0.$$

- Idea: We know $y$ and $A$, but we want $x$
- Estimate components $x$ as linear combinations of components of $y$

$$x_i = \sum_{j=1}^{m} a_j y_j$$

- Then looking for closest $\hat{x} \in M = \{Ky : K \in \mathbb{R}^{n\times m}\}$
  - (can write $M$ as $\text{span}(y_i e_j)_{i\in\{1,\cdots,m\},j\in\{1,\cdots,n\}}$)

# The Projection we Need

Let $y = Ax + \epsilon$ where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ are random vectors and

$$E[\epsilon\epsilon^T] = Q, \quad E[xx^T] = C, \quad E[\epsilon x^T] = 0.$$

- Idea: We know $y$ and $A$, but we want $x$
- Estimate components $x$ as linear combinations of components of $y$

$$x_i = \sum_{j=1}^{m} a_j y_j$$

- Then looking for closest $\hat{x} \in M = \{Ky : K \in \mathbb{R}^{n \times m}\}$
  - (can write $M$ as $\mathrm{span}(y_i e_j)_{i \in \{1,\cdots,m\}, j \in \{1,\cdots,n\}}$)

# The Projection we Need

Let $y = Ax + \epsilon$ where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ are random vectors and

$$E[\epsilon\epsilon^T] = Q, \quad E[xx^T] = C, \quad E[\epsilon x^T] = 0.$$

- Idea: We know $y$ and $A$, but we want $x$
- Estimate components $x$ as linear combinations of components of $y$

$$x_i = \sum_{j=1}^{m} a_j y_j$$

- Then looking for closest $\hat{x} \in M = \{Ky \,:\, K \in \mathbb{R}^{n\times m}\}$
  - (can write $M$ as $\text{span}(y_i e_j)_{i\in\{1,\cdots,m\}, j\in\{1,\cdots,n\}}$)

# Optimization on (sub)-Hilbert Spaces

## Theorem

*Let* $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ *be random vectors, with* $cov(y,y)_{ij} = \langle y_i, y_j \rangle_{\mathbb{E}} = \mathbb{E}[y_i y_j]$, $cov(x,y)_{ij} = \langle x_i, y_j \rangle_{\mathbb{E}}$ *and* $span(\{y_i e_j\}_{i,j=1}^{m,n}) = M \subseteq \mathbb{R}^n$. *Then the unique minimizing vector* $\hat{x} \in M$ *is given by*

$$\hat{x} = cov(x,y)cov(y,y)^{-1}y$$

$$\hat{C} = \mathbb{E}[(\hat{x} - x)(\hat{x} - x)^T] = cov(x,x) - cov(x,y)cov(y,y)^{-1}cov(x,y)^T$$

# Optimization on (sub)-Hilbert Spaces

Proof.

- Any vector $y_0 \in M$ can be written as $Ky$ for $K \in \mathbb{R}^{n \times m}$.
- The minimizing vector $Ky = \hat{x} \in M$ must have the property that $\langle Ky - x, y_i e_j \rangle = 0$ for all $i \in [1, \cdots, n]$, $j \in [1, \cdots, m]$. Therefore

$$\forall i, j : \ \langle Ky, y_i e_j \rangle = \langle x, y_i e_j \rangle$$

- Forming a system of equations gives $cov(y, y)K^T = cov(x, y)^T$. Then $K = cov(x, y)cov(y, y)^{-1}$
- Compute $\mathbb{E}[(x - \hat{x})(x - \hat{x})^T]$ by plugging in $\hat{x}$

□

# Optimization on (sub)-Hilbert Spaces

Proof.

- Any vector $y_0 \in M$ can be written as $Ky$ for $K \in \mathbb{R}^{n \times m}$.
- The minimizing vector $Ky = \hat{x} \in M$ must have the property that $\langle Ky - x, y_i e_j \rangle = 0$ for all $i \in [1, \cdots, n]$, $j \in [1, \cdots, m]$. Therefore

$$\forall i, j : \ \langle Ky, y_i e_j \rangle = \langle x, y_i e_j \rangle$$

- Forming a system of equations gives $cov(y, y)K^T = cov(x, y)^T$. Then $K = cov(x, y)cov(y, y)^{-1}$
- Compute $\mathbb{E}[(x - \hat{x})(x - \hat{x})^T]$ by plugging in $\hat{x}$

# Optimization on (sub)-Hilbert Spaces

Proof.

- Any vector $y_0 \in M$ can be written as $Ky$ for $K \in \mathbb{R}^{n \times m}$.
- The minimizing vector $Ky = \hat{x} \in M$ must have the property that $\langle Ky - x, y_i e_j \rangle = 0$ for all $i \in [1, \cdots, n]$, $j \in [1, \cdots, m]$. Therefore

$$\forall i, j : \ \langle Ky, y_i e_j \rangle = \langle x, y_i e_j \rangle$$

- Forming a system of equations gives $cov(y, y)K^T = cov(x, y)^T$. Then $K = cov(x, y)cov(y, y)^{-1}$
- Compute $\mathbb{E}[(x - \hat{x})(x - \hat{x})^T]$ by plugging in $\hat{x}$

□

# The Projection we Need

Let $y = Ax + \epsilon$ where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ are random vectors and

$$E[\epsilon\epsilon^T] = Q, \quad E[xx^T] = C, \quad E[\epsilon x^T] = 0.$$

**Theorem**

*Then the optimal linear estimate $\hat{x}$ is*

$$\hat{x} = CA^T(ACA^T + Q)^{-1}y$$
$$\hat{C} = C - CA^T(ACA^T + Q)^{-1}AC^T$$

# Optimal Estimator of Transformed Subspace

Let $L$ be any linear transformation of $\mathbb{R}^n$. Then if you transform the problem to finding the closest vector to $Lx$ in $\text{span}(\{y_i(Le_j)\}) = LM$, the solution is exactly $L\hat{x}$.

Example (Example)

If you're given $x_{k|k}$ and want to find $x_{k+1|k}$, just hit $x_{k|k}$ with your dynamics,

$$x_{k+1|k} = A_k x_{k|k}$$

# Optimal Estimator of Transformed Subspace

Let $L$ be any linear transformation of $\mathbb{R}^n$. Then if you transform the problem to finding the closest vector to $Lx$ in $\text{span}(\{y_i(Le_j)\}) = LM$, the solution is exactly $L\hat{x}$.

**Example (Example)**

If you're given $x_{k|k}$ and want to find $x_{k+1|k}$, just hit $x_{k|k}$ with your dynamics,

$$x_{k+1|k} = A_k x_{k|k}$$

# Direct Sums from a Sum

- Already have a best least-squares estimate $\hat{x} \in \mathbb{R}^n$ of $x \in \mathbb{R}^n$ from a vector $y \in \mathbb{R}^m$.
- What happens if new information arrives, $y' \in \mathbb{R}^\ell$?
- Linear algebra idea: form orthogonal subspaces $M$ and $\tilde{M}$ from $y$ and $y'$ resp.
  - "Take out from $y'$ what we know from $y$":

$$\tilde{y} = y' - \text{proj}_{\{y_i e_j\}_{j \in \{1, \cdots, \ell\}}}(y')$$

# Direct Sums from a Sum

- Already have a best least-squares estimate $\hat{x} \in \mathbb{R}^n$ of $x \in \mathbb{R}^n$ from a vector $y \in \mathbb{R}^m$.
- What happens if new information arrives, $y' \in \mathbb{R}^\ell$?
- Linear algebra idea: form orthogonal subspaces $M$ and $\tilde{M}$ from $y$ and $y'$ resp.
  - "Take out from $y'$ what we know from $y$":

$$\tilde{y} = y' - \text{proj}_{\{y_i e_j\}_{j \in \{1,\cdots,\ell\}}}(y')$$

# Direct Sums from a Sum

- Already have a best least-squares estimate $\hat{x} \in \mathbb{R}^n$ of $x \in \mathbb{R}^n$ from a vector $y \in \mathbb{R}^m$.
- What happens if new information arrives, $y' \in \mathbb{R}^\ell$?
- Linear algebra idea: form orthogonal subspaces $M$ and $\tilde{M}$ from $y$ and $y'$ resp.
  - "Take out from $y'$ what we know from $y$":

$$\tilde{y} = y' - \text{proj}_{\{y_i e_j\}_{j \in \{1,\cdots,\ell\}}}(y')$$

# Direct Sums from a Sum

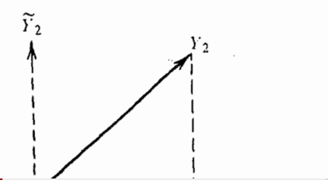
**Theorem ($P_{M+M'} = P_M \oplus P_{\tilde{M}}$)**

*The best estimate of the random vector $x \in \mathbb{R}^n$ built from both $y \in \mathbb{R}^m$ and $y' \in \mathbb{R}^\ell$ can be computed as*

$$\hat{\hat{x}} = \hat{x} + cov(x, \tilde{y}) cov(\tilde{y}, \tilde{y})^{-1} \tilde{y}$$

$$\hat{\hat{C}} = \hat{C} - cov(x, \tilde{y}) cov(\tilde{y}, \tilde{y})^{-1} cov(x, \tilde{y})^T$$

*where $\tilde{y} = y' - \mathrm{proj}_{\{y_i e_j\}_{j \in \{1, \cdots, \ell\}}}(y')$ and $\tilde{M}$ is the subspace of random vectors in $\mathbb{R}^n$ generated by vectors of the form $K\tilde{y}$ for $n \times \ell$ matrix $K$.*



# Topic

# The measurement problem

Given $x_{k|k-1}$, we'd like to update our estimate to $x_{k|k}$ subject to

$$z_k = M_k x_k + \delta_k$$

- Apply previous theorem and linear transformation property: set $\tilde{z} = z_k - M_k x_{k|k-1}$. Then

$$x_{k|k} = x_{k|k-1} + cov(x_k, \tilde{z}) cov(\tilde{z}, \tilde{z})^{-1} \tilde{z}$$
$$C_{k|k} = C_{k|k-1} - cov(x_k, \tilde{z}) cov(\tilde{z}, \tilde{z})^{-1} cov(x, \tilde{z})^T$$

(in more typical form:)

$$x_{k|k} = x_{k|k-1} + C_{k|k-1} M_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} \left( z_k - M_k x_{k|k-1} \right)$$

# The measurement problem

Given $x_{k|k-1}$, we'd like to update our estimate to $x_{k|k}$ subject to

$$z_k = M_k x_k + \delta_k$$

- Apply previous theorem and linear transformation property: set $\tilde{z} = z_k - M_k x_{k|k-1}$. Then

$$x_{k|k} = x_{k|k-1} + cov(x_k, \tilde{z}) cov(\tilde{z}, \tilde{z})^{-1} \tilde{z}$$
$$C_{k|k} = C_{k|k-1} - cov(x_k, \tilde{z}) cov(\tilde{z}, \tilde{z})^{-1} cov(x, \tilde{z})^T$$

(in more typical form:)

$$x_{k|k} = x_{k|k-1} + C_{k|k-1} M_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} \left( z_k - M_k x_{k|k-1} \right)$$

# The measurement problem

Given $x_{k|k-1}$, we'd like to update our estimate to $x_{k|k}$ subject to

$$z_k = M_k x_k + \delta_k$$

- Apply previous theorem and linear transformation property: set $\tilde{z} = z_k - M_k x_{k|k-1}$. Then

$$\begin{aligned} x_{k|k} &= x_{k|k-1} + cov(x_k, \tilde{z}) cov(\tilde{z}, \tilde{z})^{-1} \tilde{z} \\ C_{k|k} &= C_{k|k-1} - cov(x_k, \tilde{z}) cov(\tilde{z}, \tilde{z})^{-1} cov(x, \tilde{z})^T \end{aligned}$$

(in more typical form:)

$$x_{k|k} = x_{k|k-1} + C_{k|k-1} M_k^T \left[ M_k C_{k|k-1} M_k^T + R_k \right]^{-1} \left( z_k - M_k x_{k|k-1} \right)$$

# The dynamics problem

Given $x_{k|k}$, we'd like to update our estimate to $x_{k+1|k}$ subject to

$$x_{k+1} = A_k x_k + \epsilon_k$$

Solved in the example earlier, just hit $x_{k|k}$ with $A_k$ to get $x_{k+1|k}$

$$x_{k+1|k} = A_k x_{k|k}$$
$$C_{k+1|k} = A_k C_{k|k} A_k^T + cov(\epsilon, \epsilon)$$

# The dynamics problem

Given $x_{k|k}$, we'd like to update our estimate to $x_{k+1|k}$ subject to

$$x_{k+1} = A_k x_k + \epsilon_k$$

Solved in the example earlier, just hit $x_{k|k}$ with $A_k$ to get $x_{k+1|k}$

$$x_{k+1|k} = A_k x_{k|k}$$
$$C_{k+1|k} = A_k C_{k|k} A_k^T + cov(\epsilon, \epsilon)$$

# Putting it together

**Theorem (Kalman, recursive linear estimation)**

*Let $x_{0|-1} = x_0$ be known, and $\epsilon_i, \delta_i$ white noise with positive definite covariances. Then the best estimate of $x_{k+1|k}$ given measurements $\{z_k\}$ can be computed recursively by*

$$
\begin{aligned}
&(\text{update, set } \tilde{z} = z_k - M_k x_{k|k-1}):\\
&\quad x_{k|k} = x_{k|k-1} + cov(\tilde{z}, x_k) cov(\tilde{z}, \tilde{z})^{-1} \tilde{z}\\
&\quad C_{k|k} = C_{k|k-1} - cov(x, \tilde{z}) cov(\tilde{z}, \tilde{z})^{-1} cov(x, \tilde{z})^T\\
&(\text{predict})\\
&\quad x_{k+1|k} = A_k x_{k|k}\\
&\quad C_{k+1|k} = A_k C_{k|k} A_k^T + cov(\epsilon, \epsilon)
\end{aligned}
$$

# Putting it together

## Theorem (Kalman, recursive linear estimation)

*Let $x_{0|-1} = x_0$ be known, and $\epsilon_i, \delta_i$ white noise with positive definite covariances. Then the best estimate of $x_{k+1|k}$ given measurements $\{z_k\}$ can be computed recursively by*

$$
\begin{aligned}
&(update,\ set\ \ \tilde{z} = z_k - M_k x_{k|k-1}):\\
&\quad x_{k|k} = x_{k|k-1} + cov(\tilde{z}, x_k)cov(\tilde{z}, \tilde{z})^{-1}\tilde{z}\\
&\quad C_{k|k} = C_{k|k-1} - cov(x, \tilde{z})cov(\tilde{z}, \tilde{z})^{-1}cov(x, \tilde{z})^T\\
&(predict)\\
&\quad x_{k+1|k} = A_k x_{k|k}\\
&\quad C_{k+1|k} = A_k C_{k|k} A_k^T + cov(\epsilon, \epsilon)
\end{aligned}
$$

# Putting it together

**Theorem (Kalman, recursive linear estimation)**

*Let $x_{0|-1} = x_0$ be known, and $\epsilon_i, \delta_i$ white noise with positive definite covariances. Then the best estimate of $x_{k+1|k}$ given measurements $\{z_k\}$ can be computed recursively by*

$$
\begin{aligned}
&(update,\ set\ \ \tilde{z} = z_k - M_k x_{k|k-1}):\\
&\qquad x_{k|k} = x_{k|k-1} + cov(\tilde{z}, x_k) cov(\tilde{z}, \tilde{z})^{-1} \tilde{z}\\
&\qquad C_{k|k} = C_{k|k-1} - cov(x, \tilde{z}) cov(\tilde{z}, \tilde{z})^{-1} cov(x, \tilde{z})^T\\
&(predict)\\
&\quad x_{k+1|k} = A_k x_{k|k}\\
&\quad C_{k+1|k} = A_k C_{k|k} A_k^T + cov(\epsilon, \epsilon)
\end{aligned}
$$

# Topic

# Extensions

- You can extrapolate further than time $n$: $x_{n+k|n} = A^k x_{k|k}$
- If there are gaps in the measurements at times $t_i$, account for this theoretically by setting $M_i = 0$
- If your dynamics depend on a time elapsed, $\Delta t = t_{i+1} - t_i$, encode this in $A_i$ and $\epsilon_i$ and everything follows through, e.g., if $x_n = [p_n \ v_n]^T$.

$$A_n(\Delta t) = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$$

$$\epsilon_i(\Delta t) = (\Delta t)^\nu \epsilon$$

# Extensions

- You can extrapolate further than time $n$: $x_{n+k|n} = A^k x_{k|k}$
- If there are gaps in the measurements at times $t_i$, account for this theoretically by setting $M_i = 0$
- If your dynamics depend on a time elapsed, $\Delta t = t_{i+1} - t_i$, encode this in $A_i$ and $\epsilon_i$ and everything follows through, e.g., if $x_n = [p_n \; v_n]^T$.

$$A_n(\Delta t) = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$$

$$\epsilon_i(\Delta t) = (\Delta t)^\nu \epsilon$$

# Extensions

- You can extrapolate further than time $n$: $x_{n+k|n} = A^k x_{k|k}$
- If there are gaps in the measurements at times $t_i$, account for this theoretically by setting $M_i = 0$
- If your dynamics depend on a time elapsed, $\Delta t = t_{i+1} - t_i$, encode this in $A_i$ and $\epsilon_i$ and everything follows through, e.g., if $x_n = [p_n \ v_n]^T$.

$$A_n(\Delta t) = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$$

$$\epsilon_i(\Delta t) = (\Delta t)^\nu \epsilon$$

# Extended Kalman Filter (ish)

Suppose your measurements are instead given by a nonlinear differentiable function $g : \mathbb{R}^n \to \mathbb{R}^m$:

$$z_k = g(x_k) + \epsilon_k$$

Mantra: Use the function where you can, but use linearization of $g$ for projections. Then we pray.

**Example (EKF (half))**

Set $\tilde{z} = z_k - g(x_{k|k-1})$. Set $M_k = Jac_g(x_k)$. Update to $x_{k|k}$ by:

$$x_{k|k} = x_{k|k-1} + cov(x_k, \tilde{z})cov(\tilde{z}, \tilde{z})^{-1}\tilde{z}$$

where $cov(\tilde{z}, \tilde{z}) = M_k C_{k|k-1} M_k^T + cov(\delta_k, \delta_k)$ and $cov(x, \tilde{z}) = C_{k|k-1} M_k^T$.

# Extended Kalman Filter (ish)

Suppose your measurements are instead given by a nonlinear differentiable function $g : \mathbb{R}^n \to \mathbb{R}^m$:

$$z_k = g(x_k) + \epsilon_k$$

Mantra: Use the function where you can, but use linearization of $g$ for projections. Then we pray.

**Example (EKF (half))**

Set $\tilde{z} = z_k - g(x_{k|k-1})$. Set $M_k = Jac_g(x_k)$. Update to $x_{k|k}$ by:

$$x_{k|k} = x_{k|k-1} + cov(x_k, \tilde{z})cov(\tilde{z}, \tilde{z})^{-1}\tilde{z}$$

where $cov(\tilde{z}, \tilde{z}) = M_k C_{k|k-1} M_k^T + cov(\delta_k, \delta_k)$ and $cov(x, \tilde{z}) = C_{k|k-1} M_k^T$.

# Extended Kalman Filter (ish)

Suppose your measurements are instead given by a nonlinear differentiable function $g : \mathbb{R}^n \to \mathbb{R}^m$:

$$z_k = g(x_k) + \epsilon_k$$

Mantra: Use the function where you can, but use linearization of $g$ for projections. Then we pray.

## Example (EKF (half))

Set $\tilde{z} = z_k - g(x_{k|k-1})$. Set $M_k = Jac_g(x_k)$. Update to $x_{k|k}$ by:

$$x_{k|k} = x_{k|k-1} + cov(x_k, \tilde{z})cov(\tilde{z}, \tilde{z})^{-1}\tilde{z}$$

where $cov(\tilde{z}, \tilde{z}) = M_k C_{k|k-1} M_k^T + cov(\delta_k, \delta_k)$ and $cov(x, \tilde{z}) = C_{k|k-1} M_k^T$.