ECE 286: Bayesian Machine Perception

Class 3: The Kalman Filter

Florian Meyer

Electrical and Computer Engineering Department
University of California San Diego

Expectation and Covariance

- Expectation of a random vector $\boldsymbol{x}$

discrete case

$$\mathbb{E}\{\boldsymbol{x}\} = \sum_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}\, p(\boldsymbol{x})$$

continuous case

$$\mathbb{E}\{\boldsymbol{x}\} = \int \boldsymbol{x}\, f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

- Expectation of transformed random vector $g(\boldsymbol{x})$

$$\mathbb{E}\{g(\boldsymbol{x})\} = \sum_{\boldsymbol{x} \in \mathcal{X}} g(\boldsymbol{x})\, p(\boldsymbol{x})$$

$$\mathbb{E}\{g(\boldsymbol{x})\} = \int g(\boldsymbol{x})\, f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

- Covariance of a random vector $\boldsymbol{x}$

$$\mathbb{C}\{\boldsymbol{x}\} = \mathbb{E}\big\{(\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x}\})(\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x}\})^{\mathrm{T}}\big\} = \mathbb{E}\big\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\big\} - \mathbb{E}\{\boldsymbol{x}\}\mathbb{E}\{\boldsymbol{x}\}^{\mathrm{T}}$$

The Gaussian Distribution

- Gaussian distribution of continuous random vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_I]^{\mathrm{T}}$

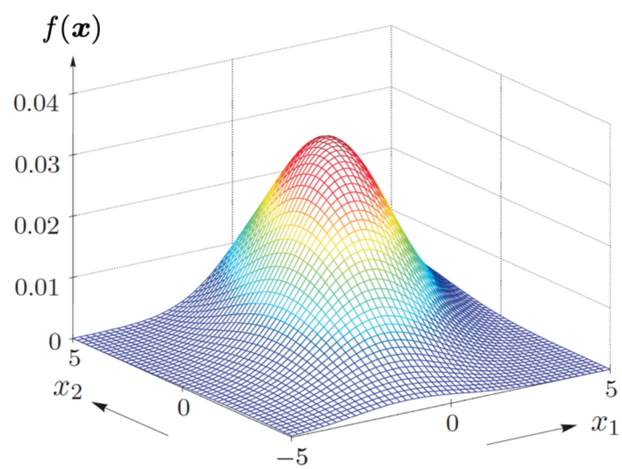
$$f(\boldsymbol{x}) = \det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{x}})^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})^{\mathrm{T}}\,\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})\Big)$$

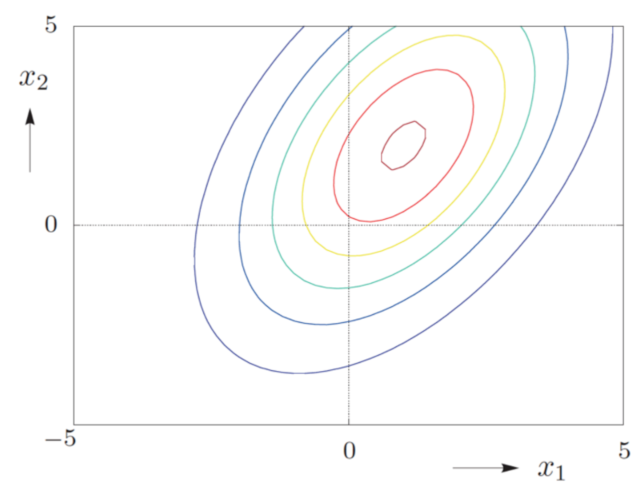
where $\boldsymbol{\mu}_{\boldsymbol{x}} = \mathbb{E}\{\boldsymbol{x}\}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \mathbb{C}\{\boldsymbol{x}\}$

- Example ($I = 2$):

$$\boldsymbol{\mu}_{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

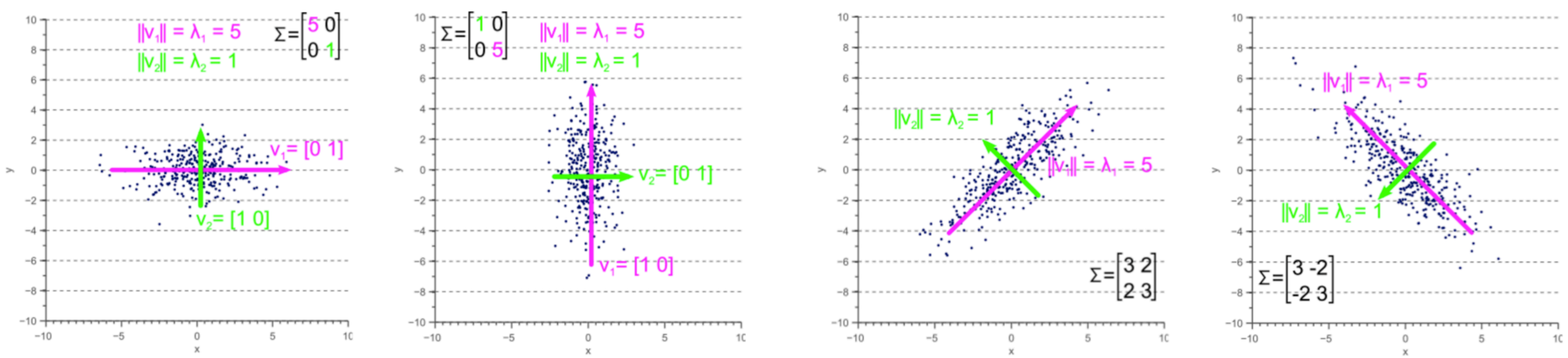
$$\boldsymbol{\Sigma}_{\boldsymbol{x}} = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$$





Geometric Interpretation of the Covariance Matrix

- Variances $[\sigma_{x_1}^2 \, \sigma_{x_2}^2 \ldots \sigma_{x_I}^2]^{\mathrm{T}} = \mathrm{diag}\, \boldsymbol{\Sigma_x}$ represent the spread of $\boldsymbol{x}$ along the axes
- The eigenvalues of $\boldsymbol{\Sigma_x}$ represent the variance of $\boldsymbol{x}$ along eigenvector directions
- Example ($\boldsymbol{\mu} = [0\,0]^{\mathrm{T}}$):



V. Spruyt, *A geometric interpretation of the covariance matrix*, 2014.

Recap: State-Space Model

- Consider a sequence of states $\boldsymbol{x}_n$ and a sequence of measurements $\boldsymbol{y}_n$

State-Transition Model:

State $\boldsymbol{x}_n$ evolves according to

$$\boldsymbol{x}_n = g_n(\boldsymbol{x}_{n-1}, \underbrace{\boldsymbol{u}_n}_{\text{driving noise (white)}})$$

This determines the joint prior

$$f(\boldsymbol{x}_{0:n}) = f(\boldsymbol{x}_0) \prod_{n'=1}^{n} f(\boldsymbol{x}_{n'}|\boldsymbol{x}_{n'-1})$$

Measurement Model:

Measurement $\boldsymbol{y}_n$ is generated as

$$\boldsymbol{y}_n = h_n(\boldsymbol{x}_n, \underbrace{\boldsymbol{v}_n}_{\text{measurement noise (white)}})$$

This determines the joint likelihood

$$f(\boldsymbol{y}_{1:n}|\boldsymbol{x}_{1:n}) = \prod_{n'=1}^{n} f(\boldsymbol{y}_{n'}|\boldsymbol{x}_{n'})$$

- By using Bayes' rule we obtain the marginal posterior pdf (for $\boldsymbol{z}_{1:n}$ fixed)

$$f(\boldsymbol{x}_{0:n}|\boldsymbol{z}_{1:n}) \propto f(\boldsymbol{x}_{0:n}) f(\boldsymbol{z}_{1:n}|\boldsymbol{x}_{1:n}) = f(\boldsymbol{x}_0) \prod_{n'=1}^{n} f(\boldsymbol{x}_{n'}|\boldsymbol{x}_{n'-1}) f(\boldsymbol{z}_{n'}|\boldsymbol{x}_{n'})$$

Linear-Gaussian State-Space Model

- Consider a sequence of states $\boldsymbol{x}_n$ and a sequence of measurements $\boldsymbol{y}_n$

State-Transition Model:

State $\boldsymbol{x}_n$ evolves according to

$$\boldsymbol{x}_n = \boldsymbol{G}_n \boldsymbol{x}_{n-1} + \underbrace{\boldsymbol{u}_n}_{\text{driving noise (white)}}$$

with Gaussian driving noise

$$\boldsymbol{u}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{u}_n})$$

Measurement Model:

Measurement $\boldsymbol{y}_n$ is generated as

$$\boldsymbol{y}_n = \boldsymbol{H}_n \boldsymbol{x}_n + \underbrace{\boldsymbol{v}_n}_{\text{measurement noise (white)}}$$

with Gaussian measurement noise

$$\boldsymbol{v}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{v}_n})$$

- Prior PDF at $n = 0$, $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{x}_0}, \boldsymbol{\Sigma}_{\boldsymbol{x}_0})$

Kalman Prediction Step

- Recall prediction step of sequential Bayesian estimation

$$\underbrace{f(\boldsymbol{x}_n|\boldsymbol{y}_{1:n-1})}_{\substack{\text{Predicted}\\ \text{posterior pdf}}} = \int \underbrace{f(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})}_{\substack{\text{State-transition}\\ \text{pdf}}} \underbrace{f(\boldsymbol{x}_{n-1}|\boldsymbol{y}_{1:n-1})}_{\substack{\text{Previous}\\ \text{posterior pdf}}} \mathrm{d}\boldsymbol{x}_{n-1}$$

- $f(\boldsymbol{x}_{n-1}|\boldsymbol{y}_{1:n-1})$ is Gaussian with mean $\boldsymbol{\mu}_{\boldsymbol{x}_{n-1}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}_{n-1}}$

- $f(\boldsymbol{x}_n|\boldsymbol{y}_{1:n-1})$ is Gaussian with mean $\boldsymbol{\mu}^-_{\boldsymbol{x}_n}$ and covariance $\boldsymbol{\Sigma}^-_{\boldsymbol{x}_n}$ given as

$$\boldsymbol{\mu}^-_{\boldsymbol{x}_n} = G_n \boldsymbol{\mu}_{\boldsymbol{x}_{n-1}} \qquad\qquad \boldsymbol{\Sigma}^-_{\boldsymbol{x}_n} = \boldsymbol{G}_n \boldsymbol{\Sigma}_{\boldsymbol{x}_{n-1}} \boldsymbol{G}_n^{\mathrm{T}} + \boldsymbol{\Sigma}_{\boldsymbol{u}_n}$$

Kalman Update Step

- Recall measurement update step of sequential Bayesian estimation

$$\underbrace{f(\boldsymbol{x}_n|\boldsymbol{y}_{1:n})}_{\text{Posterior pdf}} \propto \underbrace{f(\boldsymbol{y}_n|\boldsymbol{x}_n)}_{\substack{\text{Likelihood}\\ \text{function}}} \underbrace{f(\boldsymbol{x}_n|\boldsymbol{y}_{1:n-1})}_{\substack{\text{Predicted}\\ \text{posterior pdf}}}$$

- $f(\boldsymbol{x}_n|\boldsymbol{y}_{1:n-1})$ is Gaussian with mean $\boldsymbol{\mu}^-_{\boldsymbol{x}_n}$ and covariance $\boldsymbol{\Sigma}^-_{\boldsymbol{x}_n}$

- $f(\boldsymbol{x}_n|\boldsymbol{y}_{1:n})$ is Gaussian with mean $\boldsymbol{\mu}_{\boldsymbol{x}_n}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}_n}$ (both can be calculated in closed form)

Kalman Update Step

- Kalman gain

$$\boldsymbol{K}_n = \boldsymbol{\Sigma}_{\boldsymbol{x}_n}^{-} \boldsymbol{H}_n^{\mathrm{T}} \left(\boldsymbol{H}_n \boldsymbol{\Sigma}_{\boldsymbol{x}_n}^{-} \boldsymbol{H}_n^{\mathrm{T}} + \boldsymbol{\Sigma}_{\boldsymbol{v}_n} \right)^{-1}$$

- Mean and covariance update

$$\boldsymbol{\mu}_{\boldsymbol{x}_n} = \boldsymbol{\mu}_{\boldsymbol{x}_n}^{-} + \boldsymbol{K}_n \left(\boldsymbol{y}_n - \boldsymbol{H}_n \boldsymbol{\mu}_{\boldsymbol{x}_n}^{-} \right)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{x}_n} = \boldsymbol{\Sigma}_{\boldsymbol{x}_n}^{-} - \boldsymbol{K}_n \boldsymbol{H}_n \boldsymbol{\Sigma}_{\boldsymbol{x}_n}^{-}$$

Kalman Filter Properties (i)

- Kalman filter provides a procedure for calculating the entire posterior PDF $f(\boldsymbol{x}_n|\boldsymbol{y}_{1:n})$ of linear-Gaussian sequential Bayesian estimation problems

- The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}_n}$ does not depend on the measurements $\boldsymbol{y}_{1:n}$ and can thus be calculated offline

- Since for Gaussian distributions the mean is equal to the maximum (or mode), $\boldsymbol{\mu}_{\boldsymbol{x}_n}$ is the optimum MMSE estimate and MAP estimate, i.e.,

$$\hat{\boldsymbol{x}}_n^{\mathrm{MMSE}} = \hat{\boldsymbol{x}}_n^{\mathrm{MAP}} = \boldsymbol{\mu}_{\boldsymbol{x}_n}$$

and $\boldsymbol{\Sigma}_{\boldsymbol{x}_n}$ is the error covariance matrix of the estimate

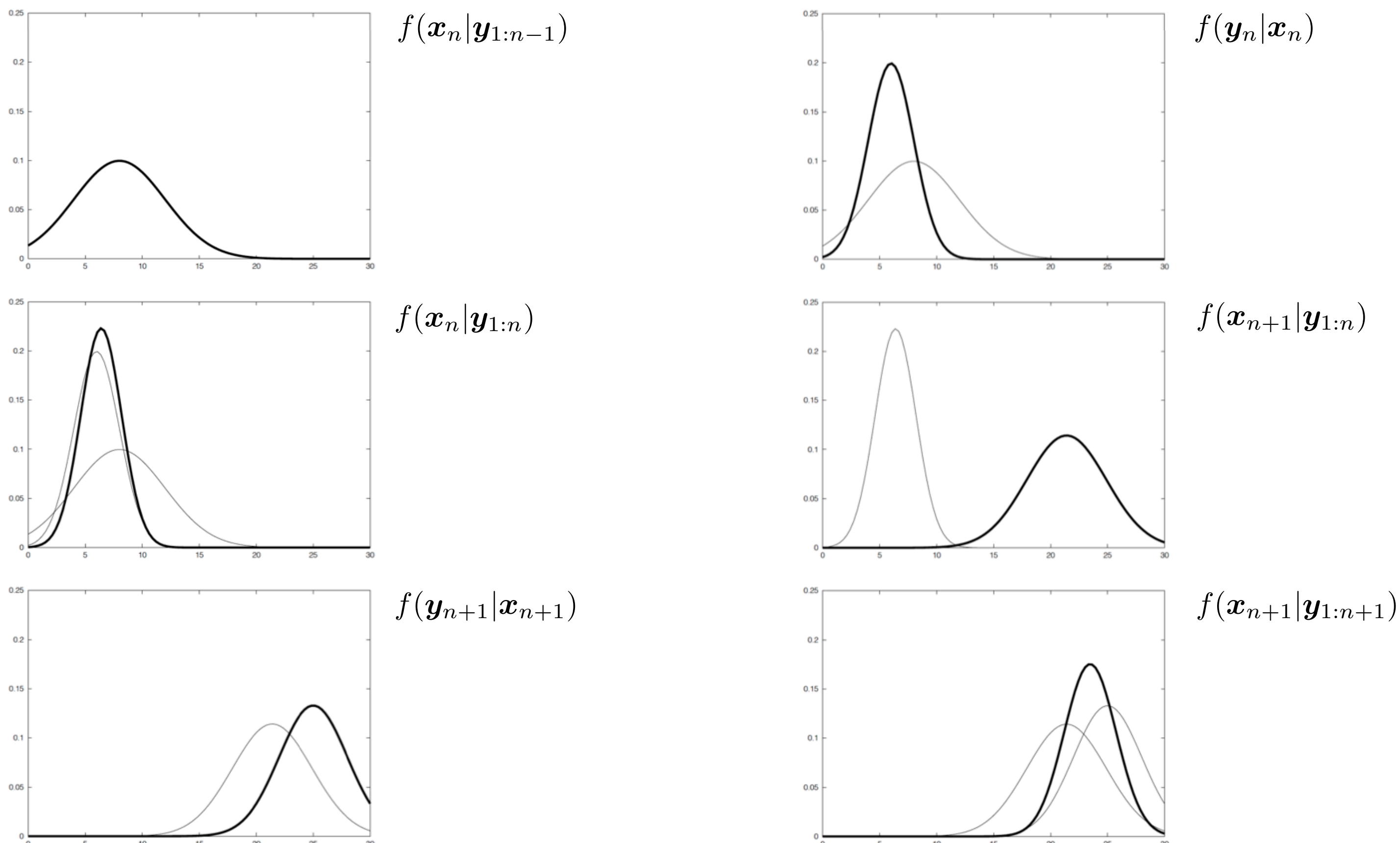
Kalman Filter Properties (ii)

- In case $f(\boldsymbol{u}_{n'})$ and $f(\boldsymbol{v}_{n'}), n' = 0, 1, \ldots, n$ or $f(\boldsymbol{x}_0)$ are not Gaussian distributions, it can be shown that the Kalman filter is the best linear MMSE estimator for $\boldsymbol{x}_n$

- The complexity of the Kalman filter is quite moderate, for I the dimension of the state space and L the dimension of the measurement space
  - it scales as $\mathcal{O}(I^2)$ due matrix multiplication in $\boldsymbol{\Sigma}_{\boldsymbol{x}_n} = (\boldsymbol{I} - \boldsymbol{K}_n \boldsymbol{H}_n)\boldsymbol{\Sigma}_{\boldsymbol{x}_n}^{-}$
  - it scales as $\mathcal{O}(L^{2.4})$ due matrix inversion in $\boldsymbol{K}_n = \boldsymbol{\Sigma}_{\boldsymbol{x}_n}^{-} \boldsymbol{H}_n^{\mathrm{T}} \left(\boldsymbol{\Sigma}_{\boldsymbol{z}_n}\right)^{-1}$

J. Anderson and B. Moore, *Optimal Filtering,* Prentice-Hall, 1979.

Kalman Filter Illustration



S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, MIT Press, 2006.