Lecture 15, Feb 4, 2026

Probability Concepts

- Bayes' rule underpins all of state estimation: $p(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})}$
  - $p(\boldsymbol{x})$ is the *prior density*
  - $p(\boldsymbol{x}|\boldsymbol{y})$ is the *posterior density*
  - $p(\boldsymbol{y}|\boldsymbol{x})$ is a *generative model/measurement likelihood*, e.g. a sensor model
  - $p(\boldsymbol{y})$ is the *marginal likelihood*, which is a normalization
- We can define *moments* for probability functions
  - The zeroth moment is always 1 by the law of total probability
  - The first (raw) moment is the mean: $\boldsymbol{\mu} = E[\boldsymbol{x}]$
  - The second (central) moment is the covariance: $\boldsymbol{\Sigma} = E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T]$
  - Third and fourth (central) moments are called *skewness* and *kurtosis*
- We denote $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ when $\boldsymbol{x}$ is Gaussian distributed; when $\boldsymbol{\mu} = \boldsymbol{0}, \boldsymbol{\Sigma} = \boldsymbol{1}$ we have a *standard Gaussian*
  - Gaussians are completely determined by their first and second moments
- A Gaussian can be factored for Bayesian inference:
  - $p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}yx & \boldsymbol{\Sigma}_{yy}\end{bmatrix}\right) = p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y})$
  - $p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})$
  - $p(\boldsymbol{y}) = \mathcal{N}(\mu_y, \boldsymbol{\Sigma}_{yy})$
- Statistical independence $p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y})$ implies uncorrelated $E[\boldsymbol{x}\boldsymbol{y}^T] = E[\boldsymbol{x}]E[\boldsymbol{y}]$, but not vice versa in general
  - For Gaussians, these are equivalent, as uncorrelated means $\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{yx}^T = 0$ and we just have $p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$ independent of $\boldsymbol{y}$
- The direct product of two Gaussian PDFs is also a Gaussian, where $\boldsymbol{\Sigma}^{-1} = \displaystyle\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1}, \boldsymbol{\mu} = \boldsymbol{\Sigma} \sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k$
- Given $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$ and a nonlinear map $\boldsymbol{g}(\boldsymbol{x})$ such that $p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{R})$, we can linearize the nonlinear map using its Jacobian, and get an approximate posterior distribution as $p(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{g}(\boldsymbol{\mu}_x), \boldsymbol{R} + \boldsymbol{G}\boldsymbol{\Sigma}_{xx}\boldsymbol{G}^T)$ where $\boldsymbol{G} = \left.\dfrac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{\mu}_x}$
- The *Sherman-Morrison-Woodbury* (SMW) identities (aka *matrix inversion lemma*) is often used to break up a matrix inversion in probabilistic estimation contexts
- To quantify the uncertainty, we can look at the *negative entropy* (aka *Shannon information*): $H(\boldsymbol{x}) = -E[\ln p(\boldsymbol{x})] = -\displaystyle\int_a^b p(\boldsymbol{x}) \ln p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$
  - For a Gaussian, $H(\boldsymbol{x}) = \dfrac{1}{2}\ln((2\pi)^N \det \boldsymbol{\Sigma}) + \dfrac{1}{2}E[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})] = \dfrac{1}{2}\ln((2\pi e)^N \det \boldsymbol{\Sigma})$
    * The second term is a *Mahalanobis distance*, similar to Euclidean distance but weighted by the inverse covariance to account for uncertainty
    * We can rewrite it using trace, and using the linearity property we can derive that the second term is always N
- Geometrically, we can interpret $\sqrt{\det \boldsymbol{\Sigma}}$ as the volume of the *uncertainty ellipsoid* formed by the PDF
  - The ellipsoids are essentially level sets of the Gaussian PDF