ECE276A: Sensing & Estimation in Robotics
Lecture 7: Probabilistic SLAM and Bayesian Filtering

Nikolay Atanasov

natanasov@ucsd.edu

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

Outline

Probability Theory Review

Probabilistic Formulation of SLAM

Bayesian Filtering

Measurable Space

- **Experiment**: repeatable procedure with a well-defined set of outcomes
- **Sample space**: set Ω of possible experiment outcomes
  - Example: $\Omega = \{HH, HT, TH, TT\}$ or $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$
- **Event**: subset A of the sample space Ω
  - Example: $A = \{HH\}$, $B = \{HT, TH\}$, $A, B \subseteq \Omega$
- σ-**algebra**: set $\mathcal{F}$ of subsets of Ω closed under complementation and countable union
- **Borel** σ-**algebra**: the smallest σ-algebra $\mathcal{B}$ containing all open sets from a topological space Ω (needed because there is no translation invariant way to assign a finite measure to all subsets of $[0, 1)$)
- **Measurable space**: tuple $(\Omega, \mathcal{F})$, where Ω is a sample space and $\mathcal{F}$ is a σ-algebra

Probability Space

- **Measure on** $(\Omega, \mathcal{F})$: function $\mu : \mathcal{F} \to \mathbb{R}$ satisfying:
  - **non-negativity**: $\mu(A) \geq 0$ for all $A \in \mathcal{F}$ and $\mu(\emptyset) = 0$
  - **countable additivity**: $\mu\left(\cup_i A_i\right) = \sum_i \mu(A_i)$ for countable number of sets $A_i \in \mathcal{F}$ that are pairwise disjoint, i.e., $A_i \cap A_j = \emptyset$
- Properties of measure μ on $(\Omega, \mathcal{F})$:
  - **subadditivity**: $\mu\left(\cup_i A_i\right) \leq \sum_i \mu(A_i)$ for countable number of sets $A_i \in \mathcal{F}$
  - $\max\{\mu(A), \mu(B)\} \leq \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \leq \mu(A) + \mu(B)$
- **Probability measure**: measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ that satisfies $\mathbb{P}(\Omega) = 1$
- **Probability space**: tuple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a sample space, $\mathcal{F}$ is a σ-algebra, and $\mathbb{P}$ is a probability measure

Conditional and Total Probability

- **Conditional probability**: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$
- **Bayes rule**: assume $\mathbb{P}(B) > 0$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

- **Total probability law**: if $\{A_1, \ldots, A_n\}$ is a partition of Ω, i.e., $\Omega = \bigcup_i A_i$ and $A_i \cap A_j = \emptyset$, $\forall i \neq j$, then:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i)$$

- **Corollary**: if $\{A_1, \ldots, A_n\}$ is a partition of Ω, then:

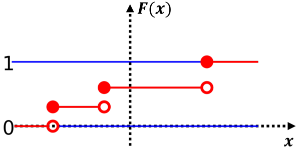
$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}$$

- **Independent events**: $\mathbb{P}\left(\bigcap_i A_i\right) = \prod_i \mathbb{P}(A_i)$
  - observing one event does not give any information about another
  - disjoint events are **not** independent: observing one tells us that the other will not occur

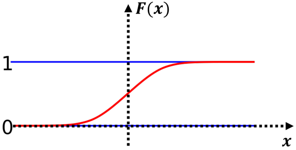
Random Variable

- **Random variable**: function $X : \Omega \to \mathbb{R}^n$ from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^n, \mathcal{B})$ such that, for every $B \in \mathcal{B}$, the set $A = \{\omega \in \Omega \mid X(\omega) \in B\}$ is contained in $\mathcal{F}$

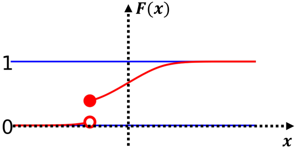
- **Cumulative distribution function** (CDF) of random variable X: function $F(\mathbf{x}) := \mathbb{P}(X \leq \mathbf{x})$ with the following properties:
  - **non-decreasing**: $\mathbf{x} \leq \mathbf{y}$ (elementwise) $\Rightarrow F(\mathbf{x}) \leq F(\mathbf{y})$
  - **right-continuous**: $\lim_{\mathbf{x} \downarrow \mathbf{y}} F(\mathbf{x}) = F(\mathbf{y})$ for all $\mathbf{y} \in \mathbb{R}^n$
  - $\lim_{x_1, \ldots, x_n \to \infty} F(\mathbf{x}) = 1$ and $\lim_{x_i \to -\infty} F(\mathbf{x}) = 0$ for all i



(a) Discrete CDF

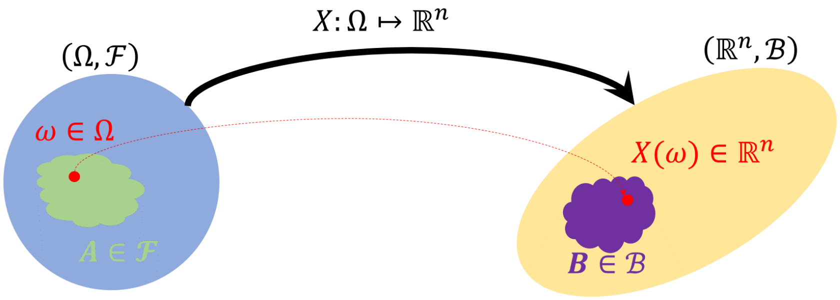


(b) Continuous CDF



(c) Mixed CDF

Random Variable



$\mathbb{P}: \mathcal{F} \mapsto \mathbb{R}$

$$\mathbb{P}(X \in B) = \mathbb{P}(A = \{\omega \in \Omega \mid X(\omega) \in B\})$$

"Volume of the preimage of B *under* X*"*

$$F_X(b) = \mathbb{P}(X \leq b) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in (-\infty, b_1] \times \cdots \times (-\infty, b_n]\})$$

$$= \int_{-\infty}^{b_n} \cdots \int_{-\infty}^{b_1} p_X(x_1, \dots, x_n) dx_1 \dots dx_n$$

CDF Examples

- $X \sim \mathcal{U}([a, b])$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

- $X \sim \mathcal{U}(\{a, b\})$

$$F(x) = \begin{cases} 0 & x < a \\ 1/2 & a \leq x < b \\ 1 & x \geq b \end{cases}$$

- $X \sim Exp(\lambda)$ with $\lambda > 0$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right) dy$$

Probability Density Function

- **Probability density function** (pdf) of a continuous random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B})$: function $p : \mathbb{R}^n \mapsto [0, 1]$ such that:
  - $p(\mathbf{x}) \geq 0$
  - $\int p(\mathbf{x}) d\mathbf{x} = 1$

- Intuition: the pdf $p(\mathbf{x})$ of X behaves like a derivative of the CDF $F(\mathbf{x})$:
  - $F(\mathbf{x}) = \mathbb{P}(X \leq \mathbf{x}) = \int_{-\infty}^{\mathbf{x}} p(\mathbf{y}) d\mathbf{y}$
  - $\mathbb{P}(\mathbf{a} < X \leq \mathbf{b}) = F(\mathbf{b}) - F(\mathbf{a}) = \int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{y}) d\mathbf{y}$
  - $\mathbb{P}(X = \mathbf{x}) = \lim_{\epsilon \to 0} \int_{\mathbf{x}}^{\mathbf{x} + \epsilon \delta \mathbf{x}} p(\mathbf{y}) d\mathbf{y} = 0$

Probability Mass Function

- **Integer set**: $\mathbb{Z} := \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
- **Probability mass function** (pmf) of a discrete random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{Z}, 2^{\mathbb{Z}})$: function $m : \mathbb{Z} \mapsto [0, 1]$ such that:
  - $m[i] \geq 0$
  - $\sum_{i \in \mathbb{Z}} m[i] = 1$
- Properties of the pmf m of X:
  - $F(i) = \mathbb{P}(X \leq i) = \sum_{j \leq i} m[j]$
  - $\mathbb{P}(a < X \leq b) = F(b) - F(a) = \sum_{a < j \leq b} m[j]$
  - $\mathbb{P}(X = i) = m[i] \in [0, 1]$
- **Dirac delta function**:

$$\delta(x) := \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases} \qquad \int_{-\infty}^{\infty} f(x)\delta(x)dx = f(0) \qquad \int_{-\infty}^{\infty} \delta(x)dx = 1$$

- A pdf can be defined for a discrete random variable $X \in \mathbb{Z}$ with pmf m using the Dirac delta function:

$$p(x) = \sum_{i \in \mathbb{Z}} m[i]\delta(x - i)$$

pdf and pmf Examples

- $X \sim \mathcal{U}([a, b])$

$$p(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$$

- $X \sim \mathcal{U}(\{a, b\})$

$$m[i] = \begin{cases} \frac{1}{2} & i \in \{a, b\} \\ 0 & \text{else} \end{cases}$$

- $X \sim Exp(\lambda)$ with $\lambda > 0$

$$p(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

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

Expectation and Variance

- Consider a random variable X with pdf p and a (measurable) function g
- The **expectation** of $g(X)$ is:

$$\mathbb{E}\left[g(X)\right] = \int g(x)p(x)dx$$

- The **variance** of $g(X)$ is:

$$\begin{aligned} Var[g(X)] &= \mathbb{E}\left[\left(g(X) - \mathbb{E}[g(X)]\right)\left(g(X) - \mathbb{E}[g(X)]\right)^\top\right] \\ &= \mathbb{E}\left[g(X)g(X)^\top\right] - \mathbb{E}[g(X)]\mathbb{E}[g(X)]^\top \end{aligned}$$

- The variance of a sum of random variables is:

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + \sum_{i=1}^{n}\sum_{j\neq i} Cov[X_i, X_j]$$

$$Cov[X_i, X_j] = \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])^\top\right] = \mathbb{E}\left[X_i X_j^\top\right] - \mathbb{E}[X_i]\mathbb{E}[X_j]^\top$$

Expectation and Variance Examples

- $X \sim \mathcal{U}([a, b])$

$$\mathbb{E}[X] = \int yp(y)dy = \frac{1}{b-a}\int_a^b ydy = \frac{b^2 - a^2}{2(b-a)} = \frac{1}{2}(a+b)$$

$$Var[X] = \int y^2 p(y)dy - \mathbb{E}[X]^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{1}{4}(a+b)^2 = \frac{1}{12}(b-a)^2$$

- $X \sim \mathcal{U}(\{a, b\})$

$$\mathbb{E}[X] = \sum_{i \in \{a,b\}} i\ m[i] = \frac{1}{2}(a+b)$$

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{2}(a^2 + b^2) - \frac{1}{4}(a+b)^2 = \frac{1}{4}(b-a)^2$$

Expectation and Variance Examples

- $X \sim Exp(\lambda)$ with $\lambda > 0$

$$\mathbb{E}[X] = \int_0^\infty y\lambda e^{-\lambda y} dy \overset{z=\lambda y,\ dz=\lambda dy}{=\!=\!=} \frac{1}{\lambda}\int_0^\infty ze^{-z} dz$$

$$\underset{du=dz,\ v=-e^{-z}}{\overset{u=z,\ dv=e^{-z}dz}{=\!=\!=}} \frac{1}{\lambda}\left(\left(-ze^{-z}\right)\Big|_0^\infty + \int_0^\infty e^{-z} dz\right) = \frac{1}{\lambda}(0+1) = \frac{1}{\lambda}$$

$$Var[X] = \int_0^\infty y^2\lambda e^{-\lambda y} dy - \frac{1}{\lambda^2} \overset{z=\lambda y,\ dz=\lambda dy}{=\!=\!=} \frac{1}{\lambda^2}\left(\int_0^\infty z^2 e^{-z} dz - 1\right)$$

$$\underset{du=2zdz,\ v=-e^{-z}}{\overset{u=z^2,\ dv=e^{-z}dz}{=\!=\!=}} \frac{1}{\lambda^2}\left(\left(-z^2e^{-z}\right)\Big|_0^\infty + 2\int_0^\infty e^{-z} dz - 1\right) = \frac{1}{\lambda^2}$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

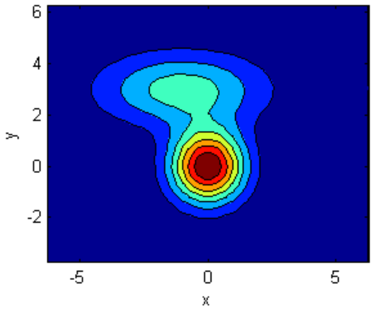
$$\mathbb{E}[X-\mu] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty \frac{(y-\mu)}{\sigma}\exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right) dy$$

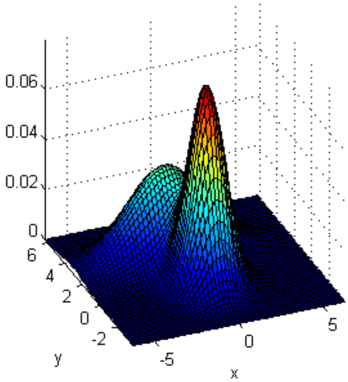
$$\underset{dz=\frac{(y-\mu)}{\sigma}dy}{\overset{z=\frac{(y-\mu)^2}{2\sigma}}{=\!=\!=}} \frac{1}{\sqrt{2\pi}}\left(\int_\infty^{\mu^2/2\sigma} e^{-z/\sigma} dz + \int_{\mu^2/2\sigma}^\infty e^{-z/\sigma} dz\right) = 0$$

Gaussian Distribution

- **Gaussian random vector** $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$
  - parameters: **mean** $\boldsymbol{\mu} \in \mathbb{R}^n$, **covariance** $\Sigma \in \mathbb{S}^n_{\succ 0}$ (symmetric positive definite $n \times n$ matrix)
  - pdf: $\phi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) := \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$
  - expectation: $\mathbb{E}[X] = \int \mathbf{x}\phi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) d\mathbf{x} = \boldsymbol{\mu}$
  - variance: $Var[X] = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\right] = \Sigma$
- **Gaussian mixture** $X \sim \mathcal{NM}(\{\alpha_k\}, \{\boldsymbol{\mu}_k\}, \{\Sigma_k\})$
  - parameters: **weights** $\alpha_k \geq 0$, $\sum_k \alpha_k = 1$,
    means $\boldsymbol{\mu}_k \in \mathbb{R}^n$, **covariances** $\Sigma_k \in \mathbb{S}^n_{\succeq 0}$
  - pdf: $p(\mathbf{x}) := \sum_k \alpha_k \phi(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k)$
  - expectation: $\mathbb{E}[X] = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x} = \sum_k \alpha_k \boldsymbol{\mu}_k =: \bar{\boldsymbol{\mu}}$
  - variance: $Var[X] = \mathbb{E}\left[XX^\top\right] - \mathbb{E}[X]\mathbb{E}[X]^\top = \sum_k \alpha_k \left(\Sigma_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^\top\right) - \bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}^\top$

pdf of a Mixture of Two 2-D Gaussians





Independent Random Variables

- The random variables $\{X_i\}_{i=1}^n$ with joint CDF $F(x_1, \ldots, x_n)$ and marginal CDFs $\{F_i(x_i)\}_{i=1}^n$ are **jointly independent** iff:

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i), \qquad \text{for all } x_1, \ldots, x_n \in \mathbb{R}.$$

- The random variables $\{X_i\}_{i=1}^n$ with joint pdf/pmf $p(x_1, \ldots, x_n)$ and marginal pdfs/pmfs $\{p_i(x_i)\}_{i=1}^n$ are **jointly independent** iff:

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_i(x_i), \qquad \text{for all } x_1, \ldots, x_n \in \mathbb{R}.$$

- Let X and Y be random variables and suppose $\mathbb{E}[X]$, $\mathbb{E}[Y]$, and $\mathbb{E}[XY]$ exist. Then, X and Y are **uncorrelated** iff $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ or equivalently $Cov[X, Y] = 0$.

- Independence implies uncorrelatedness

Conditional and Total Probability

- **Total probability**: If two random variables X, Y have a joint pdf $p(x, y)$, the marginal pdf $p(x)$ of X is:

$$p(x) = \int p(x, y) dy$$

- **Conditional probability**: If two random variables X, Y have a joint pdf $p(x, y)$, the pdf $p(x|y)$ of X conditioned on $Y = y$ and the pdf $p(y|x)$ of Y conditioned on $X = x$ satisfy

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x)$$

- **Bayes rule**: The pdf $p(x|y)$ of X conditioned on $Y = y$ can be expressed in terms of the pdf $p(y|x)$ of Y conditioned on $X = x$ and the marginal pdf $p(x)$ of X:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y \mid x')p(x')dx'}$$

Joint and Marginal Distribution Example

- Suppose $V = (X, Y)$ is a continuous random vector with density $p_V(x, y) = 8xy$ for $0 < y < x$ and $0 < x < 1$
- Let $g(x, y) = 2x + y$
  - Determine $\mathbb{E}[g(V)]$
  - Evaluate $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ by finding the marginal densities of X and Y and then evaluating the appropriate univariate integrals
  - Determine $Var[g(V)]$

Joint and Marginal Distribution Example

$$
\begin{aligned}
\mathbb{E}[2X+Y] &= \int_0^1 \int_0^x (2x+y) 8xy \; dydx = \frac{32}{15} \\
p_X(x) &= \int_0^x 8xy \; dy = 4x^3 \text{ for } 0 \le x \le 1 \\
\mathbb{E}[X] &= \int_0^1 x p_X(x) dx = \int_0^1 4x^4 dx = \frac{4}{5} \\
p_Y(y) &= \int_y^1 8xy \; dx = 4y - 4y^3 \text{ for } 0 \le y \le 1 \\
\mathbb{E}[Y] &= \int_0^1 y p_Y(y) dy = \int_0^1 4y^2 - 4y^4 dy = \frac{8}{15} \\
Var[g(V)] &= \mathbb{E}\left[(g(V) - \mathbb{E}[g(V)])^2\right] = \mathbb{E}\left[\left(2X + Y - \frac{32}{15}\right)^2\right] \\
&= \int_0^1 \int_0^x \left(2x + y - \frac{32}{15}\right)^2 8xy \; dydx = \frac{17}{75}
\end{aligned}
$$

Conditional Probability Example

- Suppose that $V = (X, Y)$ is a discrete random vector with probability mass function:

$$p_V(x,y) = \begin{cases} 0.10 & \text{if } (x,y) = (0,0) \\ 0.20 & \text{if } (x,y) = (0,1) \\ 0.30 & \text{if } (x,y) = (1,0) \\ 0.15 & \text{if } (x,y) = (1,1) \\ 0.25 & \text{if } (x,y) = (2,2) \\ 0 & \text{elsewhere} \end{cases}$$

  - What is the conditional probability that V is $(0,0)$ given that V is $(0,0)$ or $(1,1)$?
  - What is the conditional probability that X is 1 or 2 given that Y is 0 or 1?
  - What is the probability that X is 1 or 2?
  - What is the probability mass function of $X \mid Y = 0$?
  - What is the expected value of $X \mid Y = 0$?

Conditional Probability Example

$$\mathbb{P}\left(V \in \{(0,0)\} \mid V \in \{(0,0),(1,1)\}\right) = \frac{\mathbb{P}\left(V \in \{(0,0)\} \cap \{(0,0),(1,1)\}\right)}{\mathbb{P}\left(V \in \{(0,0),(1,1)\}\right)}$$

$$= \frac{0.10}{0.25} = 0.4$$

$$\mathbb{P}\left(X \in \{1,2\} \mid Y \in \{0,1\}\right) = \mathbb{P}\left(V \in \{1,2\} \times \mathbb{R} \mid V \in \mathbb{R} \times \{0,1\}\right)$$

$$= \frac{\mathbb{P}\left(V \in \{(1,0),(1,1)\}\right)}{\mathbb{P}\left(V \in \{(0,0),(0,1),(1,0),(1,1)\}\right)} = \frac{0.45}{0.75} = 0.6$$

$$\mathbb{P}\left(X \in \{1,2\}\right) = \mathbb{P}\left(V \in \{1,2\} \times \mathbb{R}\right) = 0.7$$

$$p_{X|Y=0}(x) = \frac{p_V(x,0)}{\sum_{x' \in \{0,1\}} p_V(x',0)} = \frac{1}{0.4} p_V(x,0) = \begin{cases} 0.25 & \text{if } x = 0 \\ 0.75 & \text{if } x = 1 \end{cases}$$

$$\mathbb{E}\left[X \mid Y = 0\right] = \sum_{x \in \{0,1\}} x p_{X|Y=0}(x) = p_{X|Y=0}(1) = 0.75$$

Change of Density

- **Convolution**: Let X and Y be independent random variables with pdfs p and q, respectively. Then, the pdf of $Z = X + Y$ is given by the convolution of p and q:

$$[p * q](z) = \int p(z - y)q(y)dy = \int p(x)q(z - x)dx$$

- **Change of Density**: Let $Y = f(X)$ be random variables related by an invertible function f such that $dy = \left|\det\left(\frac{df}{dx}(x)\right)\right| dx$. The pdf of $p_y(y)$ of Y and the pdf $p_x(x)$ of X are related by change of variables:

$$\begin{aligned}\mathbb{P}(Y \in A) = \mathbb{P}(X \in f^{-1}(A)) &= \int_{f^{-1}(A)} p_x(x)dx \\ &= \int_A \underbrace{\frac{1}{\left|\det\left(\frac{df}{dx}(f^{-1}(y))\right)\right|} p_x(f^{-1}(y))}_{p_y(y)} dy\end{aligned}$$

Change of Density Example

- Let $X \sim \mathcal{N}(0, \sigma^2)$ and $Y = f(X) = \exp(X)$
- Note that $f(x)$ is invertible $f^{-1}(y) = \log(y)$
- The infinitesimal integration volumes for y and x are related by:

$$dy = \left| \det \left(\frac{df}{dx}(x) \right) \right| dx = \exp(x) dx$$

- Using change of density with $A = [0, \infty)$ and $f^{-1}(A) = (-\infty, \infty)$:

$$\mathbb{P}(Y \in [0, \infty)) = \int_{-\infty}^{\infty} \phi(x; 0, \sigma^2) dx = \int_0^{\infty} \frac{1}{\exp(\log(y))} \phi(\log(y); 0, \sigma^2) dy$$

$$= \int_0^{\infty} \underbrace{\frac{1}{y} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{\log^2(y)}{\sigma^2} \right)}_{p(y)} dy$$

Change of Density Example

- Let $V := (X, Y)$ be a random vector with pdf:

$$p_V(x,y) := \begin{cases} 2y - x & x < y < 2x \text{ and } 1 < x < 2 \\ 0 & \text{else} \end{cases}$$

- Let $T := (M, N) = g(V) := \left(\frac{2X-Y}{3}, \ \frac{X+Y}{3}\right)$ be a function of V
- Note that $X = M + N$ and $Y = 2N - M$ and, hence, the pdf of V is non-zero for $0 < m < n/2$ and $1 < m + n < 2$. Also:

$$\det\left(\frac{dg}{dv}\right) = \det\begin{bmatrix} 2/3 & -1/3 \\ 1/3 & 1/3 \end{bmatrix} = \frac{1}{3}$$

- The pdf T is:

$$p_T(m,n) = \begin{cases} \frac{1}{\left|\det\left(\frac{dg}{dv}(m+n, 2n-m)\right)\right|} p_V(m+n, 2n-m), & 0 < m < n/2 \text{ and } \\ & 1 < m + n < 2, \\ 0, & \text{else.} \end{cases}$$

Outline

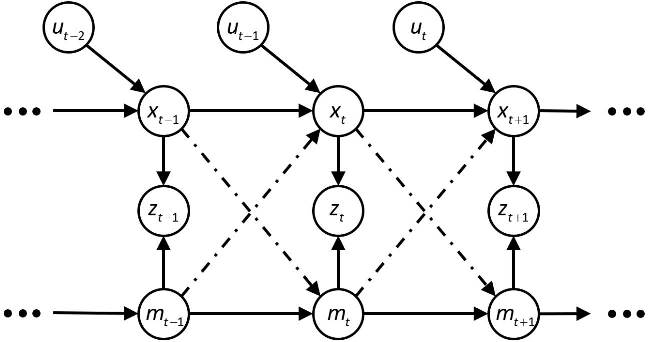
Probability Theory Review

Probabilistic Formulation of SLAM

Bayesian Filtering

Structure of Robotics Problems

- **Time**: t (discrete or continuous)
- **Robot state**: $\mathbf{x}_t$ (e.g., position, orientation, velocity)
- **Control input**: $\mathbf{u}_t$ (e.g., force, torque)
- **Observation**: $\mathbf{z}_t$ (e.g., image, laser scan, inertial measurements)
- **Map state**: $\mathbf{m}_t$ (e.g., occupancy map)



Markov Assumptions

- The control inputs $\mathbf{u}_{0:t}$ and observations $\mathbf{z}_{0:t}$ are known (observable)
- The robot states $\mathbf{x}_{0:t}$ and map $\mathbf{m}_{0:t}$ are unknown (partially observable)
- **Overloaded notation**: we consider the joint robot and map state $(\mathbf{x}_t, \mathbf{m}_t)$ as a single random variable $\mathbf{x}_t$
- **Markov assumptions**
  - The state $\mathbf{x}_{t+1}$ only depends on the previous input $\mathbf{u}_t$ and state $\mathbf{x}_t$, i.e., $\mathbf{x}_{t+1}$ given $\mathbf{u}_t$, $\mathbf{x}_t$ is independent of the history $\mathbf{x}_{0:t-1}$, $\mathbf{z}_{0:t-1}$, $\mathbf{u}_{0:t-1}$
  - The observation $\mathbf{z}_t$ only depends on the state $\mathbf{x}_t$
- **Motion model**: function f or equivalently probability density function p_f that describes the state $\mathbf{x}_{t+1}$ resulting from applying input $\mathbf{u}_t$ at state $\mathbf{x}_t$:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) \sim p_f(\cdot \mid \mathbf{x}_t, \mathbf{u}_t) \qquad \mathbf{w}_t = \text{motion noise}$$

- **Observation model**: function h or equivalently probability density function p_h that describes the observation $\mathbf{z}_t$ depending on $\mathbf{x}_t$

$$\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{v}_t) \sim p_h(\cdot \mid \mathbf{x}_t) \qquad \mathbf{v}_t = \text{observation noise}$$

Joint Distribution Factorization

- The Markov assumptions induce a factorization of the joint probability density function of the states $\mathbf{x}_{0:T}$, observations $\mathbf{z}_{0:T}$, and inputs $\mathbf{u}_{0:T-1}$:

$$
\begin{aligned}
&p(\mathbf{x}_{0:T}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}) \\
&\overset{\text{Conditional}}{\underset{\text{probability}}{=\!=\!=}} p(\mathbf{z}_T | \mathbf{x}_{0:T}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-1}) p(\mathbf{x}_{0:T}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-1}) \\
&\overset{\text{Markov}}{\underset{\text{assumption}}{=\!=\!=}} \underbrace{p_h(\mathbf{z}_T | \mathbf{x}_T)}_{\text{observation model}} \; p(\mathbf{x}_{0:T}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-1}) \\
&\overset{\text{Conditional}}{\underset{\text{probability}}{=\!=\!=}} p_h(\mathbf{z}_T | \mathbf{x}_T) p(\mathbf{x}_T | \mathbf{x}_{0:T-1}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-1}) p(\mathbf{x}_{0:T-1}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-1}) \\
&\overset{\text{Markov}}{\underset{\text{assumption}}{=\!=\!=}} p_h(\mathbf{z}_T | \mathbf{x}_T) \underbrace{p_f(\mathbf{x}_T | \mathbf{x}_{T-1}, \mathbf{u}_{T-1})}_{\text{motion model}} \underbrace{p(\mathbf{u}_{T-1} | \mathbf{x}_{T-1})}_{\text{control policy}} p(\mathbf{x}_{0:T-1}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-2}) \\
&= \cdots \\
&= \underbrace{p(\mathbf{x}_0)}_{\text{prior}} \prod_{t=0}^{T} \underbrace{p_h(\mathbf{z}_t \mid \mathbf{x}_t)}_{\text{observation model}} \prod_{t=0}^{T-1} \underbrace{p_f(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t)}_{\text{motion model}} \prod_{t=0}^{T-1} \underbrace{p(\mathbf{u}_t \mid \mathbf{x}_t)}_{\text{control policy}}
\end{aligned}
$$

Probabilistic Parameter Estimation

- Consider data D generated by probabilistic model $p(D|\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$

- **Maximum Likelihood Estimation (MLE)**: maximize the likelihood of the data D given the parameters $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}_* \in \arg\max_{\boldsymbol{\theta}} p(D|\boldsymbol{\theta})$$

- **Maximum A Posteriori (MAP)**: maximize the likelihood of the parameters $\boldsymbol{\theta}$ given the data D:

$$\boldsymbol{\theta}_* \in \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|D) = \arg\max_{\boldsymbol{\theta}} p(D|\boldsymbol{\theta})p(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} p(D, \boldsymbol{\theta})$$

MAP Formulation of SLAM

- SLAM as a MAP problem:
  - data: $D = \{\mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}\}$
  - parameters: $\boldsymbol{\theta} = \mathbf{x}_{0:T}$
  - joint pdf: $p(D, \boldsymbol{\theta}) = p(\mathbf{x}_0) \prod_{t=0}^{T} p_h(\mathbf{z}_t \mid \mathbf{x}_t) \prod_{t=0}^{T-1} p_f(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t) \prod_{t=0}^{T-1} p(\mathbf{u}_t \mid \mathbf{x}_t)$
- Factor graph optimization (usually $p(\mathbf{u}_t|\mathbf{x}_t)$ is not considered):

$$\min_{\mathbf{x}_{0:T}} \; -\log p(\mathbf{x}_0) - \sum_{t=0}^{T} \log p_h(\mathbf{z}_t|\mathbf{x}_t) - \sum_{t=0}^{T-1} \log p_f(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t)$$

  - Start with initial guess $\hat{\mathbf{x}}_{0:T}$, e.g., from odometry and feature triangulation
  - Linearize motion model $f(\mathbf{x}, \mathbf{u}, \mathbf{w})$ and observation model $h(\mathbf{x}, \mathbf{v})$
  - Solve the linearized problem to obtain a descent direction $\tilde{\mathbf{x}}_{0:T}$
  - Update the guess $\hat{\mathbf{x}}'_{0:T} = \hat{\mathbf{x}}_{0:T} + \alpha \, \tilde{\mathbf{x}}_{0:T}$
  - Perform descent by re-linearizing around $\hat{\mathbf{x}}'_{0:T}$ and obtaining a new descent direction $\tilde{\mathbf{x}}'_{0:T}$

Motion Model Linearization

- Motion model linearization around state $\hat{\mathbf{x}}_t$ and noise 0:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) \approx f(\hat{\mathbf{x}}_t, \mathbf{u}_t, 0) + F_t(\mathbf{x}_t - \hat{\mathbf{x}}_t) + Q_t \mathbf{w}_t$$

- Motion model Jacobians:

$$F_t = \frac{df}{d\mathbf{x}}(\hat{\mathbf{x}}_t, \mathbf{u}_t, 0) \qquad\qquad Q_t = \frac{df}{d\mathbf{w}}(\hat{\mathbf{x}}_t, \mathbf{u}_t, 0)$$

- Let $\tilde{\mathbf{x}}_t := \mathbf{x}_t - \hat{\mathbf{x}}_t$ and $\boldsymbol{\eta}_{t+1} := \hat{\mathbf{x}}_{t+1} - f(\hat{\mathbf{x}}_t, \mathbf{u}_t, 0)$:

$$\begin{aligned}\tilde{\mathbf{x}}_{t+1} + \hat{\mathbf{x}}_{t+1} &\approx f(\hat{\mathbf{x}}_t, \mathbf{u}_t, 0) + F_t \tilde{\mathbf{x}}_t + Q_t \mathbf{w}_t \\ \tilde{\mathbf{x}}_{t+1} + \boldsymbol{\eta}_{t+1} &\approx F_t \tilde{\mathbf{x}}_t + Q_t \mathbf{w}_t\end{aligned}$$

- Motion model pdf with $\mathbf{w}_t \sim \mathcal{N}(0, W)$ and $W_t := Q_t W Q_t^\top$:

$$p_f(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{u}_t) \approx$$

$$\frac{1}{\sqrt{(2\pi)^{d_x} \det(W_t)}} \exp\left(-\frac{1}{2}\left(\tilde{\mathbf{x}}_{t+1} + \boldsymbol{\eta}_{t+1} - F_t \tilde{\mathbf{x}}_t\right)^\top W_t^{-1} \left(\tilde{\mathbf{x}}_{t+1} + \boldsymbol{\eta}_{t+1} - F_t \tilde{\mathbf{x}}_t\right)\right)$$

$$\log p_f(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{u}_t) \approx$$

$$-\frac{1}{2}\log((2\pi)^{d_x} \det(W_t)) - \frac{1}{2}\left(\tilde{\mathbf{x}}_{t+1} + \boldsymbol{\eta}_{t+1} - F_t \tilde{\mathbf{x}}_t\right)^\top W_t^{-1} \left(\tilde{\mathbf{x}}_{t+1} + \boldsymbol{\eta}_{t+1} - F_t \tilde{\mathbf{x}}_t\right)$$

Observation Model Linearization

- Observation model linearization around state $\hat{\mathbf{x}}_t$ and noise 0:

$$\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{v}_t) \approx h(\hat{\mathbf{x}}_t, 0) + H_t(\mathbf{x}_t - \hat{\mathbf{x}}_t) + R_t \mathbf{v}_t$$

- Observation model Jacobians:

$$H_t = \frac{dh}{d\mathbf{x}}(\hat{\mathbf{x}}_t, 0) \qquad R_t = \frac{dh}{d\mathbf{v}}(\hat{\mathbf{x}}_t, 0)$$

- Let $\tilde{\mathbf{x}}_t := \mathbf{x}_t - \hat{\mathbf{x}}_t$ and $\tilde{\mathbf{z}}_t := \mathbf{z}_t - h(\hat{\mathbf{x}}_t, 0)$:

$$\tilde{\mathbf{z}}_t = H_t \tilde{\mathbf{x}}_t + R_t \mathbf{v}_t$$

- Observation model pdf with $\mathbf{v}_t \sim \mathcal{N}(0, V)$ and $V_t := R_t V R_t^\top$:

$$p_h(\mathbf{z}_t | \mathbf{x}_t) \approx \frac{1}{\sqrt{(2\pi)^{d_z} \det(V_t)}} \exp\left(-\frac{1}{2}\left(\tilde{\mathbf{z}}_t - H_t \tilde{\mathbf{x}}_t\right)^\top V_t^{-1} \left(\tilde{\mathbf{z}}_t - H_t \tilde{\mathbf{x}}_t\right)\right)$$

$$\log p_h(\mathbf{z}_t | \mathbf{x}_t) \approx -\frac{1}{2}\log((2\pi)^{d_z} \det(V_t)) - \frac{1}{2}\left(\tilde{\mathbf{z}}_t - H_t \tilde{\mathbf{x}}_t\right)^\top V_t^{-1} \left(\tilde{\mathbf{z}}_t - H_t \tilde{\mathbf{x}}_t\right)$$

Descent Direction from Linearized MAP Problem

- Linearized MAP problem is a least-squares problem:

$$\min_{\tilde{\mathbf{x}}_{0:T}} \left\{ \|\Sigma_0^{-1/2}\tilde{\mathbf{x}}_0\|_2^2 + \sum_{t=0}^{T} \| V_t^{-1/2} \left(\tilde{\mathbf{z}}_t - H_t\tilde{\mathbf{x}}_t\right) \|_2^2 + \sum_{t=0}^{T-1} \| W_t^{-1/2} \left(\tilde{\mathbf{x}}_{t+1} + \boldsymbol{\eta}_{t+1} - F_t\tilde{\mathbf{x}}_t\right) \|_2^2 \right\}$$

- Using that $\left\| \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} - \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \right\|_2^2 = \|\mathbf{x}_1 - \mathbf{y}_1\|_2^2 + \|\mathbf{x}_2 - \mathbf{y}_2\|_2^2$ for $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{R}^{d_1}$, $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{R}^{d_2}$, rewrite the least-squares cost in matrix notation:

$$\begin{aligned}
&\|\Sigma_0^{-1/2}\tilde{\mathbf{x}}_0\|_2^2 + \sum_{t=0}^{T} \| V_t^{-1/2} \left(\tilde{\mathbf{z}}_t - H_t\tilde{\mathbf{x}}_t\right) \|_2^2 + \sum_{t=0}^{T-1} \| W_t^{-1/2} \left(\tilde{\mathbf{x}}_{t+1} + \boldsymbol{\eta}_{t+1} - F_t\tilde{\mathbf{x}}_t\right) \|_2^2 \\
&= \|\Sigma_0^{-1/2}\tilde{\mathbf{x}}_0\|^2 + \left\| \begin{bmatrix} V_0^{-1/2}\left(\tilde{\mathbf{z}}_0 - H_0\tilde{\mathbf{x}}_0\right) \\ \vdots \\ V_T^{-1/2}\left(\tilde{\mathbf{z}}_T - H_T\tilde{\mathbf{x}}_T\right) \end{bmatrix} \right\|_2^2 + \left\| \begin{bmatrix} W_0^{-1/2}\left(\boldsymbol{\eta}_1 + \tilde{\mathbf{x}}_1 - F_0\tilde{\mathbf{x}}_0\right) \\ \vdots \\ W_{T-1}^{-1/2}\left(\boldsymbol{\eta}_T + \tilde{\mathbf{x}}_T - F_{T-1}\tilde{\mathbf{x}}_{T-1}\right) \end{bmatrix} \right\|_2^2
\end{aligned}$$

Descent Direction from Linearized MAP Problem

- Using that $\left\| \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} - \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \right\|_2^2 = \|\mathbf{x}_1 - \mathbf{y}_1\|_2^2 + \|\mathbf{x}_2 - \mathbf{y}_2\|_2^2$ for $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{R}^{d_1}$, $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{R}^{d_2}$, rewrite the least-squares cost in matrix notation:

$$\|\Sigma_0^{-1/2}\tilde{\mathbf{x}}_0\|_2^2 + \sum_{t=0}^{T} \| V_t^{-1/2} \left(\tilde{\mathbf{z}}_t - H_t \tilde{\mathbf{x}}_t\right) \|_2^2 + \sum_{t=0}^{T-1} \| W_t^{-1/2} \left(\tilde{\mathbf{x}}_{t+1} + \boldsymbol{\eta}_{t+1} - F_t \tilde{\mathbf{x}}_t\right) \|_2^2$$

$$= \|\Sigma_0^{-1/2}\tilde{\mathbf{x}}_0\|^2 + \left\| \begin{bmatrix} V_0^{-1/2} H_0 & & \\ & \ddots & \\ & & V_T^{-1/2} H_T \end{bmatrix} \begin{pmatrix} \tilde{\mathbf{x}}_0 \\ \vdots \\ \tilde{\mathbf{x}}_T \end{pmatrix} - \begin{bmatrix} V_0^{-1/2} \tilde{\mathbf{z}}_0 \\ \vdots \\ V_T^{-1/2} \tilde{\mathbf{z}}_T \end{bmatrix} \right\|_2^2$$

$$+ \left\| \begin{bmatrix} W_0^{-1/2} F_0 & -W_0^{-1/2} & & \\ & W_1^{-1/2} F_1 & \ddots & \\ & & \ddots & -W_{T-1}^{-1/2} \\ & & & W_{T-1}^{-1/2} F_{T-1} \end{bmatrix} \begin{pmatrix} \tilde{\mathbf{x}}_0 \\ \vdots \\ \tilde{\mathbf{x}}_T \end{pmatrix} - \begin{bmatrix} W_0^{-1/2} \eta_1 \\ \vdots \\ W_{T-1}^{-1/2} \eta_T \end{bmatrix} \right\|_2^2$$

Descent Direction from Linearized MAP Problem

$$
\left\| \underbrace{\begin{bmatrix}
\Sigma_{0|0}^{-1/2} & & & \\
V_0^{-1/2} H_0 & & & \\
& \ddots & & \\
& & \ddots & \\
& & & V_T^{-1/2} H_T \\
W_0^{-1/2} F_0 & -W_0^{-1/2} & & \\
& W_1^{-1/2} F_1 & \ddots & \\
& & \ddots & -W_{T-1}^{-1/2} \\
& & & W_{T-1}^{-1/2} F_{T-1}
\end{bmatrix}}_{J}
\begin{pmatrix} \tilde{\mathbf{x}}_0 \\ \vdots \\ \tilde{\mathbf{x}}_T \end{pmatrix}
-
\underbrace{\begin{bmatrix}
0 \\
V_0^{-1/2} \tilde{\mathbf{z}}_0 \\
\vdots \\
\vdots \\
V_T^{-1/2} \tilde{\mathbf{z}}_T \\
W_0^{-1/2} \boldsymbol{\eta}_1 \\
W_1^{-1/2} \boldsymbol{\eta}_2 \\
\vdots \\
W_{T-1}^{-1/2} \boldsymbol{\eta}_T
\end{bmatrix}}_{\mathbf{b}}
\right\|_2^2
$$

$$
= \| J \tilde{\mathbf{x}}_{0:T} - \mathbf{b} \|_2^2
$$

Descent Direction from Linearized MAP Problem

- Obtain a descent direction $\tilde{\mathbf{x}}_{0:T}$ from the linearized MAP problem:

$$\min_{\tilde{\mathbf{x}}_{0:T}} \|J\tilde{\mathbf{x}}_{0:T} - \mathbf{b}\|_2^2$$

- Setting the gradient to zero leads to the **normal equations**:

$$J^\top J\tilde{\mathbf{x}}_{0:T} = J^\top \mathbf{b}$$

- The Jacobian matrix J is **sparse**

- $J^\top J$ is the **info matrix** of the Gaussian distribution of $\tilde{\mathbf{x}}_{0:T} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}$

- The normal equations can be solved via:
  - Cholesky factorization of $J^\top J$
  - QR factorization of J
  - QR factorization is a more efficient and robust way to solve the normal equations because it avoids computing $J^\top J$, which is slow and squares the condition number of J

Descent Direction from Linearized MAP Problem

- Number of variables: n
- Number of measurement constraints: m
- QR factorization: $J = Q \begin{bmatrix} R \\ 0 \end{bmatrix} \in \mathbb{R}^{m \times n}$
- $R \in \mathbb{R}^{n \times n}$ is the **upper-triangular square root information matrix**

$$R^\top R = J^\top J$$

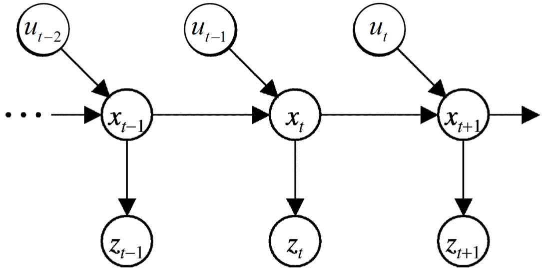
- $Q \in \mathbb{R}^{m \times m}$ is an orthogonal matrix: $Q^\top Q = I$
- Descent direction via QR factorization:

$$\begin{aligned} \|J\tilde{\mathbf{x}}_{0:T} - \mathbf{b}\|_2^2 &= \left\| Q \begin{bmatrix} R \\ 0 \end{bmatrix} \tilde{\mathbf{x}}_{0:T} - \mathbf{b} \right\|_2^2 = \left\| Q^\top Q \begin{bmatrix} R \\ 0 \end{bmatrix} \tilde{\mathbf{x}}_{0:T} - Q^\top \mathbf{b} \right\|_2^2 \\ &= \left\| \begin{bmatrix} R \\ 0 \end{bmatrix} \tilde{\mathbf{x}}_{0:T} - \begin{bmatrix} \mathbf{b}_1' \\ \mathbf{b}_2' \end{bmatrix} \right\|_2^2 = \|R\tilde{\mathbf{x}}_{0:T} - \mathbf{b}_1'\|_2^2 + \underbrace{\|\mathbf{b}_2'\|_2^2}_{\text{residual}} \end{aligned}$$

- Since R is upper-triangular, back-substitution can be used to compute $\tilde{\mathbf{x}}_{0:T}$

Outline

Markov Assumptions



- **Motion model**: given $\mathbf{x}_t$, $\mathbf{u}_t$, the state $\mathbf{x}_{t+1}$ is independent of the history $\mathbf{x}_{0:t-1}$, $\mathbf{z}_{0:t-1}$, $\mathbf{u}_{0:t-1}$:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) \sim p_f(\cdot \mid \mathbf{x}_t, \mathbf{u}_t)$$

- **Observation model**: given $\mathbf{x}_t$, the observation $\mathbf{z}_t$ is independent of the history $\mathbf{x}_{0:t-1}$, $\mathbf{z}_{0:t-1}$, $\mathbf{u}_{0:t-1}$:

$$\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{v}_t) \sim p_h(\cdot \mid \mathbf{x}_t)$$

Bayes Filter

- **Bayes filter**: a probabilistic inference technique for estimating the state $\mathbf{x}_t$ of a dynamical system by combining evidence from control inputs $\mathbf{u}_t$ and observations $\mathbf{z}_t$ using the **Markov assumptions**, **conditional probability**, **total probability**, and **Bayes rule**

- The Bayes filter keeps track of:
  - **Predicted pdf**: $p_{t+1|t}(\mathbf{x}_{t+1}) := p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t})$
  - **Updated pdf**: $p_{t+1|t+1}(\mathbf{x}_{t+1}) := p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t+1}, \mathbf{u}_{0:t})$

- Special cases of the Bayes filter:
  - Particle filter
  - Kalman filter
  - Forward algorithm for Hidden Markov Models

Bayes Filter Prediction and Update Steps

- Starting with a prior pdf $p_{t|t}(\mathbf{x}_t)$, the Bayes filter uses a prediction step to obtain a predicted pdf $p_{t+1|t}(\mathbf{x}_{t+1})$ by incorporating information about the motion model p_f and input $\mathbf{u}_t$ and an update step to obtain an updated pdf $p_{t+1|t+1}(\mathbf{x}_{t+1})$ by incorporating information about the observation model p_h and observation $\mathbf{z}_{t+1}$

- **Prediction step**: given a prior pdf $p_{t|t}$ of $\mathbf{x}_t$ and control input $\mathbf{u}_t$, use the motion model p_f to compute the predicted pdf $p_{t+1|t}$ of $\mathbf{x}_{t+1}$:

$$\boxed{p_{t+1|t}(\mathbf{x}) = \int p_f(\mathbf{x} \mid \mathbf{s}, \mathbf{u}_t) p_{t|t}(\mathbf{s}) d\mathbf{s}}$$

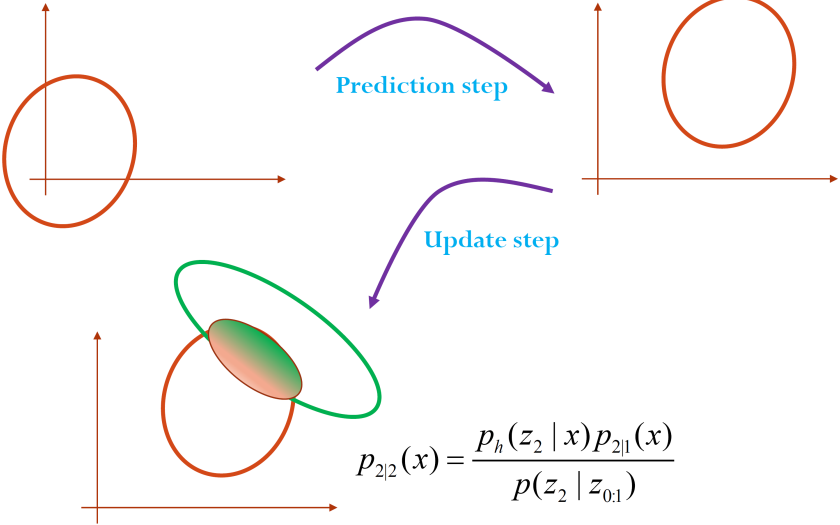
- **Update step**: given a predicted pdf $p_{t+1|t}$ of $\mathbf{x}_{t+1}$ and measurement $\mathbf{z}_{t+1}$, use the observation model p_h to obtain the updated pdf $p_{t+1|t+1}$ of $\mathbf{x}_{t+1}$:

$$\boxed{p_{t+1|t+1}(\mathbf{x}) = \frac{p_h(\mathbf{z}_{t+1} \mid \mathbf{x}) p_{t+1|t}(\mathbf{x})}{\int p_h(\mathbf{z}_{t+1} \mid \mathbf{s}) p_{t+1|t}(\mathbf{s}) d\mathbf{s}}}$$

Bayes Filter Illustration

$$p_{1|1}(x) := p(x_1 \mid z_{0:1}, u_0)$$

$$p_{2|1}(x) = \int p_f\,(x \mid s, u_1)\, p_{1|1}(s) ds$$



$$p_{2|2}(x) = \frac{p_h(z_2 \mid x) p_{2|1}(x)}{p(z_2 \mid z_{0:1})}$$

Bayes Filter Derivation

$$
\begin{aligned}
p_{t+1|t+1}(\mathbf{x}_{t+1}) &= p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t+1}, \mathbf{u}_{0:t}) \\
&\underset{\text{rule}}{\overset{\text{Bayes}}{=\!=\!=}} \frac{1}{\eta_{t+1}} p(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}, \mathbf{z}_{0:t}, \mathbf{u}_{0:t}) p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}) \\
&\underset{\text{assumption}}{\overset{\text{Markov}}{=\!=\!=}} \frac{1}{\eta_{t+1}} p_h(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}) \\
&\underset{\text{probability}}{\overset{\text{Total}}{=\!=\!=}} \frac{1}{\eta_{t+1}} p_h(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) \int p(\mathbf{x}_{t+1}, \mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}) d\mathbf{x}_t \\
&\underset{\text{probability}}{\overset{\text{Conditional}}{=\!=\!=}} \frac{1}{\eta_{t+1}} p_h(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) \int p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}, \mathbf{x}_t) p(\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}) d\mathbf{x}_t \\
&\underset{\text{assumption}}{\overset{\text{Markov}}{=\!=\!=}} \frac{1}{\eta_{t+1}} p_h(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) \int p_f(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t) p(\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}) d\mathbf{x}_t \\
&= \frac{1}{\eta_{t+1}} p_h(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) \int p_f(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t) p_{t|t}(\mathbf{x}_t) d\mathbf{x}_t
\end{aligned}
$$

- **Normalization constant**: $\eta_{t+1} = p(\mathbf{z}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t})$

Bayes Filter

- **Motion model**: $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) \sim p_f(\cdot \mid \mathbf{x}_t, \mathbf{u}_t)$
- **Observation model**: $\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{v}_t) \sim p_h(\cdot \mid \mathbf{x}_t)$
- **Bayes filter**: recursive computation of $p(\mathbf{x}_T | \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1})$ that tracks:
  - **Updated pdf**: $p_{t|t}(\mathbf{x}_t) := p(\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1})$
  - **Predicted pdf**: $p_{t+1|t}(\mathbf{x}_{t+1}) := p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t})$

$$p_{t+1|t+1}(\mathbf{x}_{t+1}) = \underbrace{\overbrace{\frac{1}{p(\mathbf{z}_{t+1}|\mathbf{z}_{0:t}, \mathbf{u}_{0:t})}}^{\frac{1}{\eta_{t+1}}} p_h(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) \overbrace{\int p_f(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t) p_{t|t}(\mathbf{x}_t) d\mathbf{x}_t}^{\textbf{Predict: } p_{t+1|t}(\mathbf{x}_{t+1})}}_{\textbf{Update}}$$

Bayes Smoother

- **Bayes smoother**: recursive computation of $p(\mathbf{x}_t|\mathbf{z}_{0:T}, \mathbf{u}_{0:T-1})$ for all $t \in \{0, \ldots, T\}$ instead of only the most recent state $\mathbf{x}_T$
  - **Smoothed pdf**: $p_{t|T}(\mathbf{x}_t) := p(\mathbf{x}_t \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1})$ for $t \in \{0, \ldots, T\}$

- **Forward pass**: compute $p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t+1}, \mathbf{u}_{0:t})$ and $p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t})$ for $t = 0, \ldots, T$ via the Bayes filter

- **Backward pass**: for $t = T-1, \ldots, 0$ compute:

$$
\begin{aligned}
p(\mathbf{x}_t \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}) &\xlongequal[\text{Probability}]{\text{Total}} \int p(\mathbf{x}_t \mid \mathbf{x}_{t+1}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}) p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}) d\mathbf{x}_{t+1} \\
&\xlongequal[\text{Assumption}]{\text{Markov}} \int p(\mathbf{x}_t \mid \mathbf{x}_{t+1}, \mathbf{z}_{0:t}, \mathbf{u}_{0:t}) p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}) d\mathbf{x}_{t+1} \\
&\xlongequal[\text{Rule}]{\text{Bayes}} \underbrace{p(\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1})}_{\text{forward pass}} \int \left[\frac{\overbrace{p_f(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t)}^{\text{motion model}} p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1})}{\underbrace{p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t})}_{\text{forward pass}}} \right] d\mathbf{x}_{t+1}
\end{aligned}
$$