ROB 501 - Mathematics for Robotics
Recitation #1

Meet Gandhi (Courtesy: Jana Pavlasek)

Sept 01, 2021

1 Notation

1. Sets:
   $\mathbb{N}$: Natural numbers: $\{1, 2, 3, \ldots\}$
   $\mathbb{Z}$: Integers, such as $-2, -1, 0, 1, 2, \ldots$
   $\mathbb{Q}$: Rational numbers: fractions of integers, such as $\frac{1}{3}, \frac{-351}{7}, \frac{2}{5}, \ldots$
   $\mathbb{R}$: Real numbers
   $\mathbb{C}$: Complex numbers: $a + ib$, where $i = \sqrt{-1}$
   $\in$: is an element of
   $\notin$: is not an element of
   $\subset$ or $\not\subseteq$: is a subset of
   $\supset$ or $\not\supseteq$: contains
   $\{x : f(x) > 0\}$ or $\{x | f(x) > 0\}$: the set of x or a collection of all the x that satisfies $f(x) > 0$
   $\cup$: the union of two sets
   $\cap$: the intersection of two sets

   Ex: $0 \in \mathbb{Z}$, $\pi \notin \mathbb{Q}$, $\mathbb{Q} \subset \mathbb{R}$, $\mathbb{C} \supset \mathbb{R}$, $2.5 \notin \{x \in \mathbb{R} : x^2 > 10\}$.

2. Logic quantifiers:
   $\forall$: for all, for each, for every, for any
   $\exists$: there exists, there is some, there is at least one
   $\Rightarrow$:implies
   $\Leftrightarrow$: if and only if, iff, is equivalent to
   $\sim$ or $\neg$: negation
   $\vee$: or
   $\wedge$: and

   Ex: $p \implies q$, $p \Leftrightarrow q$, $p \vee q$, $p \wedge q$, $\sim q$, $\sim (p \wedge \sim q)$, $(p \implies q) \Leftrightarrow \big(\sim (p \wedge \sim q)\big)$

3. Others:
   $A \in \mathbb{R}^{m \times n}$: A is an m-by-n matrix of real numbers
   $A \in \mathbb{C}^{m \times n}$: A is an m-by-n matrix of complex numbers
   $[A]_{ij}$: the entry on the i-th row, j-th column of matrix A
   $f : D_1 \to D_2$: a function/mapping/transformation that maps set D_1 to D_2, D_1 is the domain of f and D_2 is the range of f. The range of f is defined as the set of output values generated by all elements in the domain D_1.

Ex:

- A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous at* x_0 if
$$\forall \epsilon > 0, \exists \delta > 0 : \forall x \in \{x : |x - x_0| < \delta\} \Rightarrow |f(x) - f(x_0)| < \epsilon.$$
- A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous on* $\mathbb{R}$ if
$$\forall \epsilon > 0, \forall x \in \mathbb{R}, \exists \delta > 0 : \forall y \in \{z : |z - x| < \delta\} \Rightarrow |f(y) - f(x)| < \epsilon.$$

2 Matrix

1. Operations:

(a) Product

i. Defintion: For matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times l}$, the product $AB \in \mathbb{R}^{n\times l}$ and
$$[AB]_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj}, \quad 1 \le i \le n, \quad 1 \le j \le l.$$

ii. Interpretation

A matrix pre-multiplied by a row vector is a linear combination of each row of this matrix.
$$\begin{bmatrix} 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 41 & 53 & 62 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 4 \end{bmatrix}.$$

A matrix post-multiplied by a column vector is a linear combination of each column of this matrix.
$$\begin{bmatrix} 0 & 5 & 39 \\ 1 & 0 & 23 \\ 0 & 6 & 11 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 6 \end{bmatrix}, \quad \begin{bmatrix} 0 & 5 \\ 1 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 6 \end{bmatrix}.$$

(b) Inverse:

Definition: B is the inverse of A if $AB = BA = I$.

i. $A^{-1} = \dfrac{\mathrm{adj}(A)}{\det(A)}$

ii. Elementary row operation $[\ A \mid I\ ] \to [\ I \mid A^{-1}\ ]$

Ex:

i. $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} = \dfrac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

$$\left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right] \xrightarrow[r_2 - \frac{1}{2}r_1]{\frac{1}{2}r_1} \left[\begin{array}{cc|cc} 1 & 1/2 & 1/2 & 0 \\ 0 & 3/2 & -1/2 & 1 \end{array}\right] \xrightarrow[r_1 - \frac{1}{3}r_2]{\frac{2}{3}r_2} \left[\begin{array}{cc|cc} 1 & 0 & 2/3 & -1/3 \\ 0 & 1 & -1/3 & 2/3 \end{array}\right]$$

ii. $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1/5 & 2/5 & 0 \\ -2/5 & -4/5 & 1 \end{bmatrix}, \quad AB = \begin{bmatrix} 1/5 & 2/5 & 0 \\ 2/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}, BA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$

iii. $A = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1/5 & -2/5 \\ 2/5 & -4/5 \\ 0 & 1 \end{bmatrix}$, $AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $BA = \begin{bmatrix} 1/5 & 2/5 & 0 \\ 2/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Note: A non-square matrix DOES NOT have an inverse.

(c) Transpose: $B = A^\top$ if $b_{ij} = a_{ji}, \forall i, j$.

(d) Properties: For $A, B \in \mathbb{R}^{n\times n}$,

i. $(AB)^\top = B^\top A^\top$

ii. $(A+B)^\top = A^\top + B^\top$

iii. $(AB)^{-1} = B^{-1}A^{-1}$

iv. $(A^\top)^{-1} = (A^{-1})^\top$

2. Determinant

(a) Definition:

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}, \forall i$$

$$M_{ij} : \text{Minor} \qquad (-1)^{i+j} M_{ij} : \text{cofactor}$$

Ex:

$$\det \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 1 & 3 \end{bmatrix} = 1 \cdot \det \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} + 2 \cdot \det \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} = 1 \cdot 3 + 2 \cdot (-2) = -1.$$

(b) Properties:

i. $\det(kA) = k^n \det(A)$

ii. $\det(A) = \prod_{i=1}^{n} \lambda_i$

iii. $\det(A^\top) = \det(A)$

iv. $\det(A^{-1}) = \dfrac{1}{\det A}$

v. $\det(AB) = \det(A)\det(B)$

3. Trace

(a) Definition:

$$\text{trace}(A) = \sum_{i=1}^{n} a_{ii}$$

(b) Properties

i. $\text{trace}(A) = \sum_{i=1}^{n} \lambda_i$

ii. $\text{trace}(A+B) = \text{trace}(A) + \text{trace}(B)$

iii. $\text{trace}(AB) = \text{trace}(BA)$

iv. $\text{trace}(ABC) = \text{trace}(BCA) = \text{trace}(CAB)$

Ex: $A = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

$$\begin{aligned}
A + B &= \begin{bmatrix} 3 & 5 \\ 4 & 2 \end{bmatrix} \\
\text{trace}(A) &= 2 + 1 = 3 \\
\text{trace}(B) &= 1 + 1 = 2 \\
\text{trace}(A + B) &= 3 + 2 = 5 \\
AB &= \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 7 \\ 4 & 5 \end{bmatrix} \\
BA &= \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 5 \\ 6 & 7 \end{bmatrix} \\
\text{trace}(AB) &= 8 + 5 = 13 \\
\text{trace}(BA) &= 6 + 7 = 13
\end{aligned}$$

3 Optimization: Minimization

1. Without constraint: Derivatives/Gradient

   Ex: $f(x, y) = x^2 + y^2 + 3xy + x - y$.

   Soln:

   $$\text{Necessary conditions } \Rightarrow \begin{cases} \dfrac{\partial}{\partial x} f(x, y) & = 2x + 3y + 1 = 0 \\ \dfrac{\partial}{\partial y} f(x, y) & = 2y + 3x - 1 = 0 \end{cases} \Rightarrow \begin{cases} x = 1 \\ y = -1 \end{cases}$$

   Using the second partial derivative test, $\det(\nabla^2 f(x, y)) = \det\left(\begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}\right) = 4 - 9 = -5 \ngeq 0 \Rightarrow (1, -1)$ is not minimal. Since this is the only point satisfying the necessary conditions for optimality, we know this function has no minimizer. In fact, this function is unbounded from below.

2. With equality constraints: Lagrange Multipliers

   Ex: $f(x, y) = 2x - y$ s.t. $x^2 + \frac{1}{4}y^2 = 2$.

   Soln: Let $g(x,\, y,\, \lambda) = 2x - y + \lambda\,(x^2 + \frac{1}{4}y^2 - 2)$.

   $$\text{Necessary conditions } \Rightarrow \begin{cases} \dfrac{\partial}{\partial x} g(x,\, y,\, \lambda) & = 2 + 2\lambda x = 0 \\ \dfrac{\partial}{\partial y} g(x,\, y,\, \lambda) & = -1 + \dfrac{1}{2}\lambda y = 0 \\ \dfrac{\partial}{\partial \lambda} g(x,\, y,\, \lambda) & = x^2 + \dfrac{1}{4}y^2 - 2 = 0 \end{cases} \Rightarrow \begin{cases} x = 1 \\ y = -2 \\ \lambda = -1 \end{cases} \text{ or } \begin{cases} x = -1 \\ y = 2 \\ \lambda = 1 \end{cases}$$

4 Probability

1. Event Probability

   Definition: The *sample space* is the set Ω of all possible outcomes.

   Definition: An *event* is a subset of the sample space Ω.

   Definition: *Probability* is a function $P(A)$ mapping event $A \subseteq \Omega$ to a number $\in \mathbb{R}$ which satisfies:

   (a) $P(A) \geq 0$
   (b) $P(\Omega) = 1$
   (c) For mutually exclusive events A_i where $A_i \cap A_j = \emptyset$ for $i \neq j$,

   $$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

   Conditional probability: The probability of event A given that event B occurred.

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

   Theorem of total probability: Suppose $\{A_1, \ldots, A_n\}$ is a partition of the sample space Ω, i.e., $\forall i, j : A_i \cap A_j = \emptyset$, $\bigcup_{i=1}^{n} A_i = \Omega$, then the probability of event B is

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

   Bayes rule:

   $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

2. Random Variables

   Cumulative distribution function (CDF): For continuous random variables,

   $$F_X(x) := P\big(X \leq x\big)$$

   With the following properties:

   (a) $\lim_{x \to \infty} F_X(x) = 1$ and $\lim_{x \to -\infty} F_X(x) = 0$
   (b) $F_X(x)$ is non-decreasing
   (c) $F_X(x)$ is continuous from the right: $F_X(X_0^+) = F_X(X_0)$
   (d) $P(\{X_1 < x \leq X_2\}) = F(X_2) - F_X(X_1)$
   (e) $P(\{x = X_1\}) = F_X(X_1) - F_X(X_1^-)$
   (f) $P(\{X_1 \leq x \leq X_2\}) = F(X_2) - F_X(X_1^-)$

   Probability density function (PDF): For continuous random variables,

   $$f_X(x) := \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)$$

   With the following properties:

(a) $f(x) \geq 0$

(b) $\int_{-\infty}^{x} f(\lambda)d\lambda = F_X(x)$

(c) $\int_{-\infty}^{\infty} f(\lambda)d\lambda = 1$

(d) $P(\{a \leq x \leq b\}) = \int_a^b f(\lambda)d\lambda$

Marginal PDF:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad f_{X_1}(x_1) = \int_{\mathbb{R}} f_X(x)\mathrm{d}x_2$$

Conditional PDF:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad f_{X_1|X_2}(x_1|X_2 = a) = \frac{f_{X_1X_2}(x_1, a)}{f_{X_2}(a)}$$

3. Normal distributions

- Scalar case: $X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\}$$

- Vector case: $X \sim N(\mu, \Sigma)$

$$f_X(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\Big\{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\Big\}$$

- **Standard Normal Distribution**
  Special case of normal distribution with $\mu = 0$ and $\sigma = 1$
  Ex: Figure shows a standard normal distribution, $X \sim N(0, 1)$. Overlay the plot for $Y = 4X$ and $Z = X + 5$

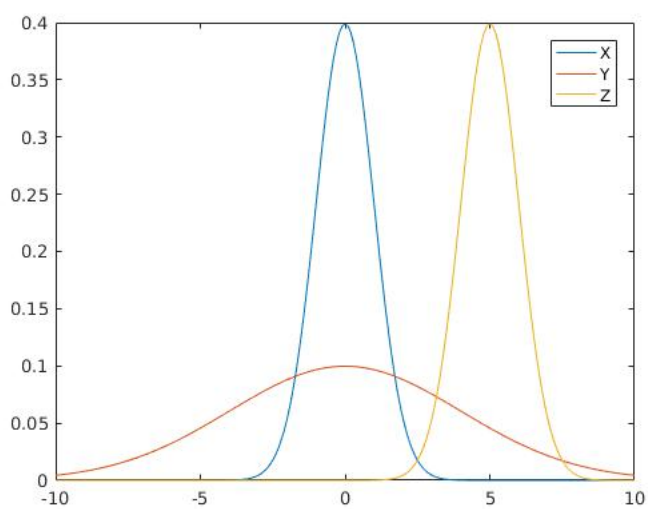

Figure 1: Standard Normal distribution

We will be covering more probability later into the course