Foundations of Robotics

Math Introduction

August 22, 2022

Topics

Part 1: Linear Algebra

Part 2: Probability

Content courtesy: Modern Robotics Linear Algebra Review: http://hades.mech.northwestern.edu/images/c/c8/AppendixE-linear-algebra-review-Dec20-2019.pdf

Preliminaries

From Wikipedia, a **matrix** is a rectangular array of table of numbers and symbols arranged in rows and columns.
In robotics, we will encounter matrices very frequently. It is important that you understand them and develop a good intuition.

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = [\boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \dots \ \boldsymbol{a}_n]$$

where $\boldsymbol{a}_1 \ \dots \ \boldsymbol{a}_n$ are column vectors.

Preliminaries

Matrix Addition:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} 3 & 1 & 6 \\ 5 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 1+3 & 2+1 & 3+6 \\ 4+5 & 5+2 & 6+4 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 9 \\ 9 & 7 & 10 \end{bmatrix}$$

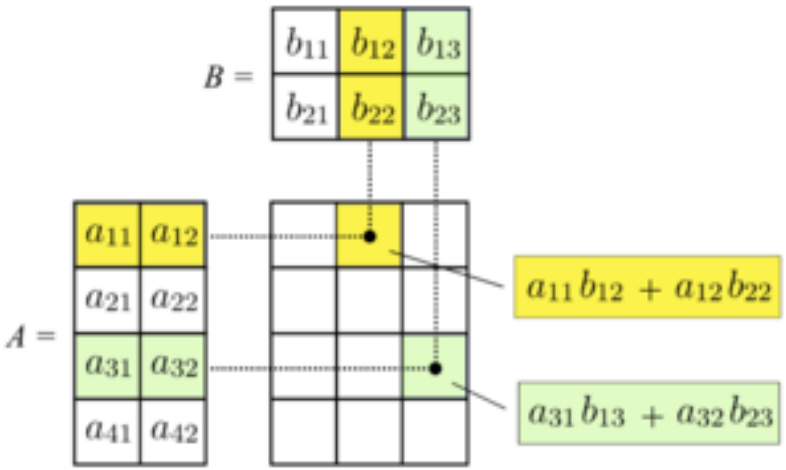
Scalar Multiplication:

$$2 \cdot \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}$$

Preliminaries

Matrix Multiplication:

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} * \begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 1*3+2*5 & 1*1+2*2 \\ 4*3+5*5 & 4*1+5*2 \end{bmatrix} = \begin{bmatrix} 13 & 5 \\ 37 & 14 \end{bmatrix}$$



Preliminaries

Transpose of matrix A is written as

$$A^T = \begin{bmatrix} a_{11} & \dots & a_{m1} \\ \vdots & \ddots & \\ a_{1n} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^T \\ \vdots \\ \boldsymbol{a}_n^T \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

Preliminaries

A very common equation we will encounter with matrices is

$$A\boldsymbol{x} = \boldsymbol{b}$$

, where $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$

Special Note

We only consider the case where all elements of matrix A are real. Results covered later may not generalize to matrices with imaginary elements.

Domain, Range, and Span

$A\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$

A different perspective...

We can view matrix $A \in \mathbb{R}^{m \times n}$ as a linear operator (a function) from space $\mathbb{R}^n$ to $\mathbb{R}^m$. In this context, $\mathbb{R}^n$ is called the **domain**, and $\mathbb{R}^m$ is called the **codomain**. The space of possible values $A\boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}^n$[domain] is the **range, image, or column space** of the linear mapping A.
$Range(A) := \{A\boldsymbol{x} | x \in \mathbb{R}^n\}$

More on Range

Range is also called a **linear span** of the **columns vectors** of A.

$$span(\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}) = \{k_1\boldsymbol{a}_1 + \cdots + k_n\boldsymbol{a}_n | k_1 \ldots k_n \in \mathbb{R}\}$$

Linearity (The superposition principle)

If we view A as a function $A : R^n \longrightarrow R^m$, it is linear if

$$\forall \alpha, \beta \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, A(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha A\boldsymbol{x} + \beta A\boldsymbol{y}$$

Rank and Null Space

For matrix $A \in \mathbb{R}^{m \times n}$

Definition

Null Space or **Kernel** of A is $Null(A) := \{A\boldsymbol{x} = 0 | \boldsymbol{x} \in \mathbb{R}^n\}$
Rank of A is the dimension of A's **range**.
Nullity of A is the dimension of A's null space.

Note, again

Note that above definitions are limited to real matrices and real vectors. Keep in mind that the actual definitions can be much broader. Typically, matrices and vectors can be defined over any field, and $\mathbb{R}$ is just one such field.

Rank

For matrix $A \in \mathbb{R}^{m \times n}$
$Rank(A) \leq min(m, n)$
The **rank** is the number of linearly independent columns of A.
If $Rank(A) = min(m, n)$, matrix A is **full rank**.
If A is not full rank, it is **singular**.
The **rank-nullity theorem**: rank(A) + nullity(A) = # of columns of A

Square Matrices

For square matrices, the number of rows equals the number of columns.
Both the domain and the codomain are $\mathbb{R}^n$
A **diagonal matrix** is a square matrix with all elements not on the diagonal equal to zero.
An **identity matrix** *I* is a diagonal matrix with all elements along the diagonal equal to one.

Determinant

The **determinant** is a scalar value, and it is a **function** of the entries of a square matrix. It characterizes some properties of the matrix and the linear mapping represented by the matrix.
The determinant of a matrix is nonzero if and only if the matrix is invertable.

$$det(A) = |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$det(A) = |A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Determinant

Example:

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1 * 4 - 3 * 2 = -2$$

$$\begin{vmatrix} 1 & 3 & 2 \\ 2 & 4 & 1 \\ 1 & 3 & 0 \end{vmatrix} = 1 * \begin{vmatrix} 4 & 1 \\ 3 & 0 \end{vmatrix} - 3 * \begin{vmatrix} 2 & 1 \\ 1 & 0 \end{vmatrix} + 2 * \begin{vmatrix} 2 & 4 \\ 1 & 3 \end{vmatrix} = 4$$

Square Matrices

If we consider matrix A as a mapping from R^n to R^n by stretching, squeezing, rotating, etc, and for some nonzero vectors v, the mapping may be particularly simple: Av is just a scaled version of v, $Av = \lambda v$.
For any λ and v satisfying $Av = \lambda v$, v is called an **eigenvector** of A and λ is the corresponding **eigenvalue**.

$$Av = \lambda v$$

$$Av - \lambda v = 0$$

$$Av - \lambda \mathbb{I} v = 0$$

, where $\mathbb{I}$ is the identity matrix

$$(A - \lambda \mathbb{I})v = 0$$

If v non-zero, the equation will have a solution only if $|A - \lambda \mathbb{I}| = 0$

Example

Eigenvalue Example:

$$A = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix}$$

$$|A - \lambda\mathbb{I}| = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1) = 0$$

When $\lambda = 1$, $(A - \lambda\mathbb{I})v = 0$, $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} v = 0$, $v = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. When $\lambda = 3$, $(A - \lambda\mathbb{I})v = 0$, $\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} v = 0$, $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

$$eigenvector : v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \lambda_1 = 1$$

$$eigenvector : v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \lambda_2 = 3$$

Inverse

For matrix $A \in \mathbb{R}^{m \times n}$,
the matrix inverse, A^{-1} is the **unique** matrix that satisfies
$AA^{-1} = A^{-1}A = I$.
Matrix inverse can be used to solve $Ax = b$, where $x = A^{-1}b$.

Pseudoinverse

If A is not full rank (it is singular), then the inverse does not exist, but we can still calculate the Moore-Penrose pseudoinverse of A, denoted $A^{\dagger}$. The peudoinverse has "inverse-like" properties and can be used to find solutions or approximate solutions to $Ax = b$, where $x = A^{\dagger}b$. The pseudoinverse $A^{\dagger}$ is equivalent to inverse A^{-1} when A is invertible.

$$AA^{\dagger}A = A$$

Pseudoinverse Properties

For $A^\dagger \in \mathbb{R}^{mxn}$ of any real matrix $A \in \mathbb{R}^{nxm}$:

$$AA^\dagger A = A$$

$$A^\dagger AA^\dagger = A^\dagger$$

$AA^\dagger$ is symmetric

$A^\dagger A$ is symmetric

Symmetric Matrices

A square matrix A is symmetric if it is equal to its transpose, $A = A^T$. A matrix A is skew symmetric if $A = -A^T$.

Basic Matrix Identities

$$(AB)C = A(BC)$$

$$(A^T)^{-1} = (A^{-1})^T$$

$$(Ax)^T = x^T A^T$$

$$(AB)^T = B^T A^T$$

$$(ABC...)^T = ...C^T B^T A^T$$

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

$$(ABC...)^{-1} = ...C^{-1} B^{-1} A^{-1}$$

Topics

Part 1: Linear Algebra

Part 2: Probability

Content Courtesy: Mathematical Statistics and Data Analysis, 3rd Edition

Random Variables

A **random variable** is a **function** from sample space Ω to real numbers. Example: A coin is tossed 3 times, and we can observe the sequence of heads and tails:

$$\Omega = \{hhh, hht, htt, hth, ttt, tth, thh, tht\}$$

The *total number of heads* is a random variable defined on Ω, so are the *total number of tail*s and the *number of heads minus the number of tails*. In general, we denote random variables with uppercase letters. We will use lowercase letters to denote the values that random variables can take on. For example, the possible values of random variable X are $x_1.x_2, ...$

Discrete and Continuous Random Variables

- A **discrete random variable** is a random variable that can take on a finite or at most a countably infinite number of values.
  Ex. Let X be a random variable that denotes the total number of tosses until a tail turns up. The possible value of X is 0, 1, 2, 3, ...
- A **continuous random variable** is a random variable that can take on a continuum of values rather than a finite or countably infinite number.
  Ex. A uniform random variable on the interval [0, 1].

Probability Mass Function and Probability Density Function

Discrete Random Variables:
There is a function p that determines the probabilities of the various values of X. If these possible values are denoted by $x_1, x_2, ...$, then $p(x_i) = P(X = x_i)$ and $\sum_i p(x_i) = 1$. This function p is called the **probability mass function**, or the **frequency function**.
Ex. Let random variable X denotes the total number of heads in 3 tosses. If the coin is fair,

$$p(0) = P(X = 0) = \frac{1}{8}$$

$$p(1) = P(X = 1) = \frac{3}{8}$$

$$p(2) = P(X = 2) = \frac{3}{8}$$

$$p(3) = P(X = 3) = \frac{1}{8}$$

Probability Mass Function and Probability Density Function

Continuous Random Variables:
the role of the frequency function is taken by the **probability density function (PDF)**, $f(x)$.

$$f(x) \geq 0$$

$$f \text{ is piece-wise continuous}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$P(a < X < b) = \int_{a}^{b} f(x)dx$$

Example: uniform random variable on the interval [0, 1]

$$f(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1 \\ 0, & x < 0 \text{ or } x > 1 \end{cases}$$

Probability Mass Function and Probability Density Function

Another common example, Normal Distribution
Ex. One-dimensional normal distribution with mean μ and variance σ^2.
$p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\}$
If x is a multi-dimensional vector, we have multivariate normal distribution
$p(x) = det(2\pi\Sigma^2)^{-\frac{1}{2}} exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$ μ is now the mean vector, and Σ is a symmetric matrix called the covariance matrix.

Cumulative distribution function(cdf)

In addition to frequency/density function, cumulative distribution function (cdf) can be also helpful. It is defined as

$$F(x) = P(X \leq x)$$

cdf is non-decreasing and satisfies

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to +\infty} F(x) = 1$$

The cdf of a random variable can be represented as an integral of its pdf, $f_X(x)$.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, f_X(x) = \frac{dF_X(x)}{dx}$$

Joint Probability

Discrete Random Variables:
Let X and Y be discrete random variables defined on the same sample space, and that they take on values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$ respectively. Their joint probability mass function $p(x, y)$ is

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$

Ex. Let X denote the number of heads on the first toss and Y the total number of heads.

	y			
x	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

Joint Probability

Let X and Y be continuous random variables with a joint cdf $F(x, y)$. Their joint density function is a piece-wise continuous function of two variables, $f(x, y)$.

Ex.

$$f(x, y) = \frac{12}{7}(x^2 + xy), 0 \leq x \leq 1, 0 \leq y \leq 1$$

$$P(X > Y) = \frac{12}{7} \int_0^1 \int_0^x (x^2 + xy) dy dx = \frac{9}{14}$$

Marginals

Let's look at this table again.

	y			
x	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

Finding random variable Y's density function from this table is easy!

$$p_Y(0) = P(Y = 0) = P(Y = 0, X = 0) + P(Y = 0, X = 1) = \frac{1}{8} + 0 = \frac{1}{8}$$

$$p_Y(1) = P(Y = 1) = P(Y = 1, X = 0) + P(Y = 1, X = 1) = \frac{2}{8} + \frac{1}{8} = \frac{3}{8}$$

p_Y is called the **marginal frequency function** of Y.

Marginals

For continuous random variables, the **marginal cdf** of X,(F_X) is

$$F_X(x) = P(X \leq x) = \lim_{y \to \infty} = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, y) dy du$$

The marginal density is

$$f_X(x) = F_X^{'}(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Ex.

$$f(x, y) = \frac{12}{7}(x^2 + xy), 0 \leq x \leq 1, 0 \leq y \leq 1$$

$$f_X(x) = \frac{12}{7} \int_0^1 (x^2 + xy) dy = \frac{12}{7}(x^2 + \frac{x}{2})$$

Conditional Probability

Discrete Case: Let X and Y be jointly distributed discrete random variables. The conditional probability of $X = x_i$ given that $Y = y_i$ is

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_i)}{P(Y = y_j)} = \frac{p(X = x_i, Y = y_j)}{p_Y(y_j)}$$

(Joint over marginal)
Continuous case: Let X and Y be continuous random variables, the conditional density of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Conditional Probability

Discrete Case Example:

	y			
x	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

$$p_{X|Y}(0|2) = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}$$

Independent Random Variables

We say two events, A and B, are independent if knowing that one had occurred provided us no information whether the other event had occurred.

$$P(A|B) = P(A), P(B|A) = P(A)$$

Now, if events A and B are independent,

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

$$P(A \cap B) = P(A)P(B)$$

Independent Random Variables

Ex. A card is drawn from a deck randomly. Random variable A denotes the event that the card is an ace. Random variable B denotes the event that the card is a spade.
A and B are independent because knowing that the card is an ace gives no information about its suit.
Mathematically,

$$P(A) = \frac{4}{52} = \frac{1}{13}, P(B) = \frac{1}{4}$$

$$P(A \cap B) = \frac{1}{52} = P(A)P(B)$$

Conditional Independence

Events A, B are conditionally independent given C if

$$P(A|B,C) = P(A|C)$$

or

$$P(A,B|C) = P(A|C)P(B|C)$$

Conditional Independence

Ex. A fair coin and a two-headed coin (P(H) = 1) are hidden in a box, one is chosen randomly. Consider the following events

- A = First coin toss results in an H.
- B = Second coin toss results in an H.
- C = Coin 1 (fair) has been selected.

$$P(B) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}, P(A) = \frac{1}{2} * \frac{1}{2} + \frac{1}{2} * 1 = \frac{3}{4}$$

$$P(A \cup B) = \frac{1}{2} * \frac{1}{4} + \frac{1}{2} * 1 = \frac{5}{8}$$

Clearly, A and B are not independent $P(A)P(B) \neq P(A \cap B)$
However, if C is given (the coin is fair), we know A and B are independent.
Thus, A and B are conditionally independent given C.
[example credit: `https://www.probabilitycourse.com/chapter1/1_4_4_conditional_independence.php`]

Bayes Rule

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

Bayes Rule

Ex. Cornell is doing COVID Testing. Let +/- denote the event that the test is positive/negative. Let P denote the event that the person actually has COVID. Let N denote the event that the person does not have COVID. Say,

$$P(+|P) = 0.88, P(-|P) = 0.12, P(-|N) = 0.86, P(+|N) = 0.14$$

Contextually, it means that of a person has COVID, the probability that this is detected by school's test is 0.88. If a person does not have COVID, the probability that the tests is negative is 0.86. Say the COVID cases at Cornell is actually rare, so P(N) = 0.99, and P(P) = 0.01.
Now, a subject tests positive. What is the probability that the test is incorrect and he is actually healthy?

$$P(N|+) = \frac{P(+|N)P(N)}{P(+|N)P(N) + P(+|P)P(P)} = \frac{(0.14)(0.99)}{(0.14)(0.99) + (0.88)(0.01)}$$

Expectation

The concept of the expected value can be interpreted as a weighted average. Intuitively, the possible values of a random variable are weighted by their probabilities.
For discrete random variable X with frequency/mass function $p(x)$:

$$E[X] = \sum_i x_i p(x_i)$$

For continuous random variable X with frequency/density function $f(x)$:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Expectation

Ex: Normal Distribution

$$E[X] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \mu$$

References

- en.wikipedia.org/wiki/Matrix_(mathematics)#Addition,_scalar_multiplication,_and_transposition
- en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors
- Modern Robotics Linear Algebra Review: http://hades.mech.northwestern.edu/images/c/c8/AppendixE-linear-algebra-review-Dec20-2019.pdf
- Mathematical Statistics and Data Analysis, 3rd Edition, ISBN-10: 9788131519547
- https://en.wikipedia.org/wiki/Conditional_independence