Linear Algebra review

Avinash Prabhu*

IIIT Hyderabad

avinash.prabhu@students.iiit.ac.in

May 17, 2021

* These slides are heavily derived from Mithun Nallana's Linear Algebra slides from previous RRC summer schools.

Overview

Why?

Why?

- Almost everything is represented in the form of matrices and vectors in robotics.
- Eg- Transformations between viewpoints, calibration, input to standard optimization libraries, etc.

Goals:

- Review important material before diving-in.
- Collect all notations at one place.
- Able to read books and seminal papers and actively participate in projects.

Spend some time after class to connect the dots.

Introduction

Key questions:

- Where am I?
- Where am I going?
- How do I get there?

Robot has:

- Sensors
- Computational resources
- Actuators

We need to model Environment, Sensors, Actuators.

Vectors and Matrices

$$
\begin{aligned}
3u + 4v &= -6 \\
9u - 5v &= 3
\end{aligned}
$$

Compactly represented as

$$Ax = b$$

$$A = \begin{bmatrix} 3 & 4 \\ 9 & -5 \end{bmatrix}, x = \begin{bmatrix} u \\ v \end{bmatrix}, b = \begin{bmatrix} -6 \\ 3 \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

Vector space and Operations

Vector Space: We define vector space as a set of all elements that are closed under vector addition and scalar multiplication operations.

Operations on vectors:

- Vector addition
- Scalar multiplication
- Element wise operation
- Dot product $(x \cdot y)$
- Cross product $(x \times y)$

Operations

Norm.
Norm is a function that takes a vector and returns non-negative number. It satisfies four properties:

- Positivity: $||x|| >= \mathbf{0}$
- Definiteness: $||x|| = \mathbf{0}$ if and only if $x = 0$
- Absolutely homogeneous: $||\alpha x|| = |\alpha|||x||$
- Triangle inequality: $||x + y|| \leq ||x|| + ||y||$

Operations

- L2 or Euclidean norm: $||x||_2 = \sqrt{x_1^2 + x_2^2 + \ldots x_n^2}$
- L1 or Manhattan norm: $||x||_1 = \sum_{i=1}^{n} |x_i|$
- Infinity norm: $||x||_\infty = \max_i |x_i|$
- p-norm: $||x||_p = (|x_1|^p + |x_2|^p + \ldots |x_n|^p)^{\frac{1}{p}}$

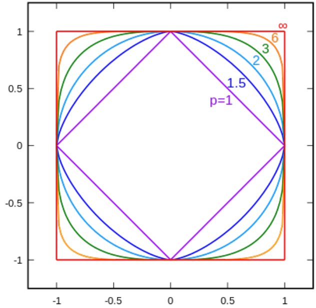


Figure: $||x||_p = 1$, for the 2-D case.

Operations

Inner Product.

Inner product is a function that takes two vectors from vector space and returns a real number. Function needs to satisfy five properties:

- Positivity: $\langle x, x \rangle \geq 0$
- Definiteness: $\langle x, x \rangle = 0$ if and only if $x = 0$
- Additivity: $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- Homogeneity: $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$
- Symmetry: $\langle x, y \rangle = \langle y, x \rangle$

Operations

Given two vectors $x, y \in \mathbb{R}^n$, we define dot product as

$$x \cdot y = x^T y = \begin{bmatrix} x_1 & x_2 & \dots x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i .$$

Inner product and norm can be related to each other as follows.

$$||x|| = \sqrt{\langle x, x \rangle}$$

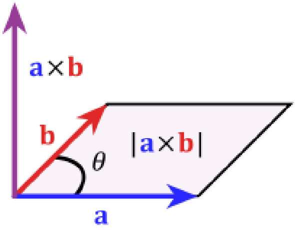
Orthogonal vectors are vectors that satisfy $\langle x, y \rangle = 0$
Using these we can establish several known facts:

- Pythagorean Theorem.
- Cauchy-Schwarz inequality.
- Triangle inequality.

Operations

We deal with $\mathbb{R}^3$ more often. So, we can define cross product or exterior product as follows:
Given two vectors x and y, $x \times y$ is a vector z that is perpendicular to both x and y with a direction given by right-hand rule and magnitude equals to area of parallelogram that vectors span.

$$x \times y = ||x||||y||sin(\theta)\mathbf{n}$$



Four equivalent forms of cross product:

$$x \times y = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

$$x \times y = (x_2 y_3 - x_3 y_2)\mathbf{i} + (x_1 y_3 - x_3 y_1)\mathbf{j} + (x_1 y_2 - x_2 y_1)\mathbf{k}$$

$$x \times y = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$x \times y = [x]_{\mathbf{x}} y$$

Operations

Given vectors $v_1, v_2, \ldots v_n$ are said to be linearly independent if no vector can be written as a linear combination of other vectors.
Other way, this set is said to be linearly dependant if one or more vectors can be written as linear combination of other remaining vectors.

$$v_1 = \sum_{i=2}^{n} \alpha_i v_i$$

We can form a vector space V from a given set of vectors by taking linear combinations.

Span : All linear combinations of elements.

Basis : Linear Independence + Span

Dimension : cardinality of basis.

Operations

Normalized vector: $||x||_2 = 1$

Orthogonal vectors: $x^T y = 0$

Orthonormal basis

$$\vec{e_1}, \vec{e_2}, \ldots \vec{e_3}$$

Commonly used orthonormal basis

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \cdots \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Matrix-vector Operations

Given $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$, their product $y = Ax \in \mathbb{R}^m$
As we have seen earlier,

$$Ax = b$$

$$A = \begin{bmatrix} 3 & 4 \\ 9 & -5 \end{bmatrix}, x = \begin{bmatrix} u \\ v \end{bmatrix}, b = \begin{bmatrix} -6 \\ 3 \end{bmatrix}$$

We can think of Ax in two ways:

$$y = Ax = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

Matrix-vector Operations

$$\begin{bmatrix} | & | & | & & | \\ a_1 & a_2 & a_3 & \dots & a_n \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} | \\ a_1 \\ | \end{bmatrix} x_1 + \begin{bmatrix} | \\ a_2 \\ | \end{bmatrix} x_2 + \dots \begin{bmatrix} | \\ a_n \\ | \end{bmatrix} x_n$$

First method writes it as vector consisting of dot products
Second method writes it as a linear combination of column vectors of A.

Matrix-Matrix Operations

Now, we can generalize a bit further to matrix multiplications. Given two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$, their product $C \in \mathbb{R}^{m \times p}$ can be written as follows:

$$C = AB$$

$$\begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} \begin{bmatrix} | & | & | & & | \\ b_1 & b_2 & b_3 & \dots & b_p \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \dots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \dots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \dots & a_m^T b_p \end{bmatrix}$$

Matrix Properties and Operations

Matrix multiplication is:

- not commutative.
- associative $(AB)C = A(BC)$
- distributive $A(B + C) = AB + AC$

Matrix Properties and Operations

Transpose of a matrix

$$B = A^T$$
$$B_{ij} = A_{ji}$$

Properties:

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

Matrix Properties and Operations

Trace of a matrix

$$tr(A) = \sum_{i=1}^{n} A_{ii}$$

Properties:

- $tr(A) = tr(A^T)$
- $tr(A + B) = tr(A) + tr(B)$
- $tr(AB) = tr(BA)$

Matrix Properties and Operations

Inverse of a square matrix

$$AA^{-1} = A^{-1}A = I$$

Properties:

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

However, inverse may not exist for all square matrices.

Matrix Properties and Operations

Row rank and Column rank of a matrix:
Size of largest subset of rows/columns of A that constitute a linearly independent set.
Row rank = Column rank
Properties:

- $rank(A) = rank(A^T)$
- $rank(A + B) \leq rank(A) + rank(B)$
- $rank(AB) \leq min(rank(A), rank(B))$

Inverse doesn't exist when A is not full rank.

Matrix Properties and Operations

Range and Nullspace of a matrix.
Span of a set of vectors $\{\ v_1, v_2, v_3, \ldots v_n\ \}$ is the set of of all possible linear combinations of elements in the set.
Range is the span of columns of A.

$$col(A) = \{v \in \mathbb{R}^{\mathrm{m}} : v = Ax, x \in \mathbb{R}^{\mathrm{n}}\}$$

Nullspace is the set of vectors when multiplied by A give a zero vector.

$$N(A) = \{v \in \mathbb{R}^{\mathrm{n}} : Av = 0\}$$

Row space and null space are orthogonal compliments of one another.

$$col(A^T) \cap N(A) = \{\mathbf{0}\}$$

Matrix Properties and Operations

Orthonormal matrix: Columns are orthogonal to one another and each column is normalized.

$$U^T U = I = UU^T$$

So, inverse of a orthonormal matrix is it's transpose.

Other important matrices

- Identity matrix, I_n

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

- Symmetric matrix, $A = A^T$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{12} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{mm} \end{bmatrix}$$

Other important matrices

- Anti-Symmetric matrix, $A = -A^T$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ -a_{12} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{1m} & -a_{2m} & \dots & a_{mm} \end{bmatrix}$$

- Triangular matrix

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ 0 & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{mm} \end{bmatrix}$$

Other important matrices

- Diagonal matrix: $diag(a_1, a_2, a_3, \ldots a_m)$

$$\begin{bmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_m \end{bmatrix}$$

- Zero matrix

$$\begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}$$

Others

- Frobenius norm

$$||A||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$$
$$= \sqrt{tr(A^TA)}$$

- Kronecker product

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix}$$

Quadratic forms

- Positive definite

$$\text{For all non-zero vectors } x \in \mathbb{R}^{\mathrm{n}}, x^T A x > 0,$$

then A is said to positive definite. It is denoted by A $\succ 0$

- Positive semi-definite

$$\text{For all non-zero vectors } x \in \mathbb{R}^{\mathrm{n}}, x^T A x \geq 0,$$

then A is said to positive definite. It is denoted by A $\succcurlyeq 0$

Eigenvalues and Eigenvectors

Eigenvector: A non zero $x \in \mathrm{I\!R}^{\mathrm{n}}$ is called called a eigenvector if it satisfies:

$$Ax = \lambda x$$

λ is called eigenvalue of matrix A.

If x is a eigenvector of A, αx is also a eigen vector of A. So, we use normalized eigenvector.

Properties:

- $tr(A) = \sum_{i=1}^{n} \lambda_i$
- $det(A) = \prod_{i=1}^{n} \lambda_i$

Eigen Decomposition

$$
\begin{aligned}
Ax_1 &= \lambda_1 x_1 \\
Ax_2 &= \lambda_2 x_2 \\
&= \vdots \\
Ax_n &= \lambda_n x_n
\end{aligned}
$$

This can be compactly written as

$$
A \begin{bmatrix} | & | & | & & | \\ x_1 & x_2 & x_3 & \dots & x_n \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | \\ x_1 & x_2 & x_3 & \dots & x_n \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}
$$

$$
AV = V\Sigma
$$

Eigen Decomposition

If V is invertible,

$$A = V\Sigma V^{-1}$$

Why carry out such a decomposition? Here is a simple yet important application. Suppose we want to calculate A^P. We know we can write A^P as the following:

$$\overbrace{V\Sigma V^{-1} V\Sigma V^{-1} ... V\Sigma V^{-1}}^{\text{P times}}$$

$$=$$

$$V\Sigma^P V^{-1}$$

Since, Σ is a diagonal matrix, Σ^P would just be a single matrix with its diagonal elements raised to P, thus significantly reducing the computation required.

SVD (Singular Value Decomposition)

However, Eigenvalue decomposition is applicable to only square matrices which are diagonalizable. Thus, SVD.

$$M = U\Sigma V^T$$

If M is a $m \times n$ matrix then U, Σ, V are $m \times m, m \times n$ and $n \times n$ respectively.

U, V are orthonormal matrices.

Σ is a matrix with positive real entries.

Columns of U, V are eigen vectors of MM^T and M^TM respectively.

LU Decomposition

A is a square matrix,

$$A = LU$$

$$A = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nn} \end{bmatrix}$$

Where L is the lower triangular matrix and U is upper triangular matrix.

Can be computed using Gauss-Jordan elimination.

Used a lot in solving a system of equations.

Other variants: LDU decomposition.

QR decomposition

A is a real square matrix,

$$A = QR$$

Where Q is an orthogonal matrix ($QQ^T = I$) and R is upper triangle matrix.

Can be computed using Gram-Schmidt process (a method to find orthogonal basis for a given set of vectors)

Used in the linear least squares problem

Other variants: RQ decomposition.

Cholesky decomposition

A is a positive definite square matrix.

$$A = LL^T$$

Where L is a lower triangular matrix

Can be computed using Cholesky algorithm.

Used in variants of Kalman filters.

Other variants: LDL decomposition.

References

- *Introduction to linear algebra* by Gilbert Stang.
- *Linear Algebra done right* by Sheldon Axler
- MITOCW 18.06 course
- Notes by Brendan Fortuner
- Notes by Hal Daume
- Zico Kolter review notes
- Zico kolter review videos
- Aaditya Ramdas review videos
- *Wikipedia*

The End