

Linear Algebra review

Mithun Nallana

IIIT Hyderabad

mithun.babu@research.iiit.ac.in

May 11, 2019

Overview

- 1 Goals and Introduction
- 2 Vector and Matrices
- 3 Operations
- 4 Matrix-vector Operations
- 5 Matrix-Matrix Operations
- 6 Matrix Properties and Operations
- 7 Other Important matrices
- 8 Matrix Decompositions
- 9 References

Why?

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.

"Before a man studies Zen, to him mountains are mountains and waters are waters; after he gets an insight into the truth of Zen through the instruction of a good master, mountains to him are not mountains and waters are not waters; but after this when he really attains to the abode of rest, mountains are once more mountains and waters are waters." - D.T.Suzuki

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

$$3u + 4v = -6$$

$$9u - 5v = 3$$

Compactly represented as

$$Ax = b$$

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

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

Vectors and Matrices

We use x to represent column vector and x^T to represent row vector.
So, A can be viewed in two ways:

$$A = \begin{bmatrix} | & | & | & \dots & | \\ a_1 & a_2 & a_3 & \dots & a_n \\ | & | & | & & | \end{bmatrix}$$

$$A = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} \quad A = \begin{bmatrix} - & a^{1T} & - \\ - & a^{2T} & - \\ & \vdots & \\ - & a^{mT} & - \end{bmatrix}$$

Vector space and Operations

Now, we can define vector space as a set of all elements that are closed under vector addition and scalar multiplication operations.

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

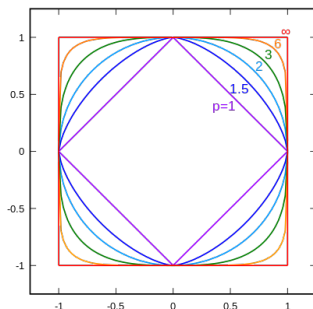
Length.

To define such properties using norm. Norm is a function that takes a vector and returns non-negative number. It satisfies four properties:

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

Operations

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



Length and Angle.

To define such properties using 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

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

Every inner product induces a norm called as induced norm.

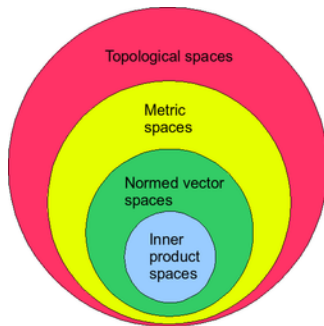
$$||x|| = \sqrt{\langle x, x \rangle}$$

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.

We can use any other norm in a inner product space to measure lengths.
More general setting:

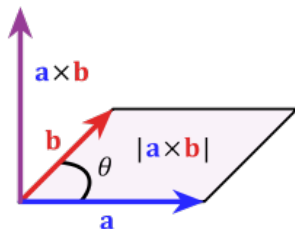


Operations

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:

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

$$\mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2)\mathbf{i} + (x_1y_3 - x_3y_1)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k}$$

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

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

Given vectors v_1, v_2, \dots, 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 vector: $x^T y = 0$

Ortho-normal basis

$$\vec{e}_1, \vec{e}_2, \dots, \vec{e}_3$$

Commonly used basis

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

Linear transformation

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

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

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 looked in three (actually four) ways:

$$C = AB$$

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

Matrix-Matrix Operations

$$\begin{bmatrix} | & | & | & & | \\ a_1 & a_2 & a_3 & \dots & a_n \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} - & b_1^T & - \\ - & b_2^T & - \\ & \vdots & \\ - & b_n^T & - \end{bmatrix} = \sum_{i=1}^n a_i b_i^T$$

$$A \begin{bmatrix} | & | & | & & | \\ b_1 & b_2 & b_3 & \dots & b_p \\ | & | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & | & & | \\ Ab_1 & Ab_2 & Ab_3 & \dots & Ab_p \\ | & | & | & & | \end{bmatrix}$$

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

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

Properties:

- $\text{tr}(A) = \text{tr}(A^T)$
- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
- $\text{tr}(AB) = \text{tr}(BA)$

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:

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

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 $\{v_1, v_2, v_3, \dots, v_n\}$ is the set of all possible linear combinations of elements in the set.

Range is the span of columns of A .

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

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

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

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

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

Matrix Properties and Operations

Determinant of a matrix.

$$\det(A) = \sum_{\sigma \in S_n} (\text{sgn}(\sigma)) \prod_{i=1}^n a_{i, \sigma_i}$$

Intuitively, Let S be the set of all linear combination of row vectors such that the coefficients of combination lie between 0 and 1, Then absolute value of determinant is the volume of the set S .

Properties:

- $\det(A) = \det(A^T)$
- $\det(A + B) \geq \det(A) + \det(B)$
- $\det(AB) = \det(A)\det(B)$

Matrix Properties and Operations

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

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

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

Other important matrices

- Identity matrix, I_n

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

- 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: $\text{diag}(a_1, a_2, a_3, \dots, a_m)$

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

- Zero matrix

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

- Frobenius norm

$$\begin{aligned}\|A\|_F &= \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} \\ &= \sqrt{\text{tr}(A^T A)}\end{aligned}$$

- 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

We also deal with 2^{nd} order functions such as conics, least squares. Such functions can be compactly represented in quadratic form:

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \\ &= \sum_{i=1}^n x_i \left(\sum_{j=1}^n a_{ij} x_j \right) \\ &= \sum_{i=1}^n x_i (Ax)_i \\ &= x^T Ax \end{aligned}$$

A is a symmetric matrix.

Properties:

- Positive definite

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

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

- Positive semi-definite

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

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

Eigenvalues and Eigenvectors

Eigenvector: A non zero $x \in \mathbb{R}^n$ 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

$$Ax_1 = \lambda x_1$$

$$Ax_2 = \lambda x_2$$

$$= \vdots$$

$$Ax_n = \lambda x_n$$

This can be compactly written as

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

$$AV = V\Sigma$$

Eigen Decomposition

If V is invertible,

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

$$A = V\Sigma V^T$$

Along with fact that, every matrix can be written as a unique product of an orthogonal matrix and a symmetric matrix.

$$M = QS$$

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

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

SVD for a matrix is:

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

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

U, V are orthonormal matrices and D is a matrix with positive real entries.

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

Columns of U, V are called left and right singular vectors of M .

Columns of U, V are eigen vectors of MM^T and $M^T M$ 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 ?

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
- Lagrange Multipliers notes by Yan Bin jia
- Zico Kolter review notes
- Zico kolter review videos
- Aaditya Ramdas review videos
- *Wikipedia*

The End