# Linear Algebra Crash Course

Sungjun Eom

February 11, 2024

# Contents

# 1. Vector Spaces and Subspaces

# 2. Linear Independence

If there are only zero $\alpha_i$ for all i which satisfy the equation below, we say a set of vectors $\{x_1, \dots, x_n\}$ is linearly independent.

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0 \tag{1}$$

# 3. Basis, Representation and Orthonormalization

## 3.1 Basis

Basis is a set of linearly independent vectors $\{x_1, \dots, x_n\}$ which is used to form an n-dimensional space. $\iff$ Basis $\{x_1, \dots, x_n\}$ spans n-dimensional space.

## 3.2 Representation

In this subsection, we restrict vectors to arrays. Representation of vectors and representation of matrices are introduced differently in this article.

### 3.2.1 Vector

$x$ is a vector whose basis is $I$. $Q$ is a new basis on which we want to represent the vector. $\bar{x}$ is a representation for the vector $x$ with respect to the basis $Q$.[1]

$$x = Q\bar{x} \tag{2}$$

We assume the basis $Q$ is formed by stacking linearly independent column vectors $q_1, q_2, \dots, q_n$.

$$Q := \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$$

### 3.2.2 Matrix

A representation of matrix $A$ is $\bar{A}$. $\bar{A}_i$ means the $i^{th}$ column of the basis $Q$.

$$Aq_i = Q\bar{A}_i \tag{3}$$

Why does the representation of a matrix look like this? What we really want to do is actually this: to represent $y = Ax$ in another form $\bar{y} = \bar{A}\bar{x}$. So we should start from the equation $y = Ax$.

$$\begin{aligned} y &= Ax \\ Q\bar{y} &= AQ\bar{x} \end{aligned}$$

Suppose $Q$ is an $n \times n$ matrix. The matrix $Q$ consists of vectors from the basis so the column vectors of the matrix are linearly independent. Then the matrix $Q$ is invertible.

$$\begin{aligned} \bar{y} &= Q^{-1}AQ\bar{x} \\ \bar{y} &= \bar{A}\bar{x} \end{aligned}$$

[1] Actually we cannot say Q is a basis. Q is just a matrix made up of vectors of the basis.

We get the equation:

$$\bar{A} = Q^{-1} A Q \tag{4}$$

Next we go forward to get the equation 3.

$$\begin{aligned} AQ &= Q\bar{A} \\ A\begin{bmatrix} q_1 & \dots & q_n \end{bmatrix} &= Q\begin{bmatrix} \bar{A}_1 & \dots & \bar{A}_n \end{bmatrix} \\ \begin{bmatrix} Aq_1 & \dots & Aq_n \end{bmatrix} &= \begin{bmatrix} Q\bar{A}_1 & \dots & Q\bar{A}_n \end{bmatrix} \end{aligned}$$

If we choose $i^{th}$ column then we get the equation 3 $Aq_i = Q\bar{A}_i$.

## 3.3 Orthonormalization

### 3.3.1 Orthonormal

Two vectors $v_1$ and $v_2$ are **orthogonal** if and only if

$$v_1^T v_2 = 0$$

If all orthogonal vectors in a set have magnitude one then we say them **orthonormal**.

$$||v_i|| = 1 \quad or \quad v_i^T v_i = 1$$

### 3.3.2 Gram-Schmidt process

Suppose we have vectors $e_1, e_2, \dots, e_n$ which we want to orthonormalize. The following steps are called Gram-Schmidt processs and are used to make orthonormal vectors $q_1, q_2, \dots, q_n$.

$$\begin{aligned} v_1 &:= e_1 & q_1 &:= \frac{v_1}{||v_1||} \\ v_2 &:= e_2 - q_1^T e_2 & q_2 &:= \frac{v_2}{||v_2||} \\ v_3 &:= e_3 - q_1^T e_3 - q_2^T e_3 & q_2 &:= \frac{v_3}{||v_3||} \\ &\vdots \\ v_n &:= e_n - \sum_{i=1}^{n-1} q_i^T e_n & q_n &:= \frac{v_n}{||v_n||} \end{aligned}$$

This process incremently removes the effect of former vectors from the current vector and normalizes the current vector.

# 4. Linear Algebraic Equation

A linear algebraic equation is:

$$\mathbf{y}_{n\times 1} = \mathbf{A}_{n\times m}\mathbf{x}_{m\times 1} \tag{5}$$

## 4.1 Projection

Let $\{z_1, \dots, z_n\}$ be an orthonormal basis for $\mathbb{R}^n$, $\{z_1, \dots, z_r\}$ be basis for $V$, and $\{z_{r+1}, \dots, z_n\}$ be basis for $V^\perp$. Next, we form matrices:

$$Z = \begin{bmatrix} z_1 & \dots & z_n \end{bmatrix}_{n\times n}, \quad Z_1 = \begin{bmatrix} z_1 & \dots & z_r \end{bmatrix}_{n\times r}, \quad Z_2 = \begin{bmatrix} z_{r+1} & \dots & z_n \end{bmatrix}_{n\times n-r}$$

$C(Z)$ means a space spanned by column vectors of Z. Then the equations below hold.

$$\mathbb{R}^n = C(Z) = span(\{z_1, \dots, z_n\}), \quad V = C(Z_1) = span(\{z_1, \dots, z_r\}), \quad V^\perp = C(Z_2) = span(\{z_{r+1}, \dots, z_n\})$$

Just to be clear, we write relations between them.

$$V \subset \mathbb{R}^n, \quad V^\perp \subset \mathbb{R}^n, \quad V \perp V^\perp$$

**Definition of projection:** We say a linear mapping $\pi_V : \mathbb{R}^n \to V$ is a projection onto $V$ when the equation below holds.

$$\forall \mathbf{v} \in V, \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad < \mathbf{v}, \pi_V(\mathbf{x}) - \mathbf{x} >= 0$$

**Definition of projection matrix:** $P$ is a projection matrix when the following equation holds.

$$\pi_V(\mathbf{x}) = P\mathbf{x}$$

$Z_1 Z_1^T$ is a projection matrix onto $V$. Likewise $Z_2 Z_2^T$ is a projection matrix onto $V^\perp$. We verify this:
For $\forall \mathbf{x} = Z\alpha \in \mathbb{R}^n$,

$$\begin{aligned}
Z_1 Z_1^T \mathbf{x} &= Z_1 Z_1^T Z\alpha \\
&= Z_1 Z_1^T \begin{bmatrix} Z_1 & Z_2 \end{bmatrix} \alpha \\
&= Z_1 Z_1^T (Z_1 \alpha_1 + Z_2 \alpha_2) \\
&= Z_1 Z_1^T Z_1 \alpha_1 + Z_1 Z_1^T Z_2 \alpha_2 \\
&= Z_1 \alpha_1
\end{aligned}$$

We used the fact that $Z_1^T Z_1 = I$. Also, notice that $I - Z_1 Z_1^T$ is a projection matrix onto $V^\perp$ and $I - Z_2 Z_2^T$ is a projection matrix onto $V$.

In the paragraph above, we only consider the case of orthogonal $Z$s. The next question which arises will be how we form a projection matrix when the matrices don't have orthonormal column vectors. Although a matrix doesn't have orhtonormal columns, we can make them be orthonormal. For example, Gram-Schmidt process. So there is always a matrix $A_{r \times r}$ which converts vectors of a matrix $X_{n \times r}$ to be orthonormal.

$$Z_1 = XA$$

Use what we already know:

$$\begin{aligned}
Z_1 Z_1^T &= XAA^T X^T \\
&\& \\
I_r = Z_1^T Z_1 &= A^T X^T XA \\
\iff (A^T)^{-1} A^{-1} &= X^T X \\
AA^T &= (X^T X)^{-1}
\end{aligned}$$

Finally, we get the equation below.

$$\therefore Z_1 Z_1^T = X(X^T X)^{-1} X^T$$

This is a projection matix onto $V$. Likewise, $I - X(X^T X)^{-1} X^T$ is a projection matrix onto $V^\perp$.
We conclude that the projection onto V is:

$$\pi_V(x) = X(X^T X)^{-1} X^T x$$

And its projection matrix is:

$$P_\pi = X(X^T X)^{-1} X^T$$

## 4.2 Least Square Method

Consider a linear model $y = X\beta + \epsilon$. We want to minimize the square sum of errors $\epsilon^T \epsilon = \epsilon_1^2 + \cdots + \epsilon_n^2$.

$$\begin{aligned}\min_\beta e^T e &= (y - X\beta)^T (y - X\beta) \\ &= < y - X\beta, y - X\beta > \\ &= < y - X\beta, y - X\beta >\end{aligned}$$