

**CS 127/227A, Fall 2022**

# **Optimization Models in Engineering**

Gireeja Ranade, UC Berkeley  
Zhiyu An

## Contents

<b>1. Linear Algebra</b>	<b>3</b>
1.a. Least-Squares Problem Statement . . . . .	3
1.b. Norm . . . . .	4
1.c. Gram-Schmidt . . . . .	5
1.d. Symmetric Matrices . . . . .	6
1.e. Principal Component Analysis . . . . .	7
1.f. Singular Value Decomposition . . . . .	8
1.g. Low-Rank Approximation . . . . .	10

# 1. Linear Algebra

## 1.a. Least-Squares Problem Statement

### Definition 1.1 (Least Squares)

Assume matrix  $A$  and vectors  $\vec{x}$  and  $\vec{b}$ . The problem defined by

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|^2$$

is a Least Squares Problem (LSP).

### Example 1.2

Assume we have two dimensional data set  $\vec{x}$  and  $\vec{y}$  and we want to formalize a LSP to find a linear correlation between  $x$  and  $y$ . We first formalize the goal linear correlation as

$$y = mx + c$$

where we want to find the optimal values for  $m$  and  $c$  to minimize the squared loss across all data points. Summarizing the above equation for all data points gives us

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Where

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} m \\ c \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

And therefore

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|^2 = \min_{m,c} \sum_{i=1}^n (y_i - (mx_i + c))^2$$

**Theorem 1.3 (Ordinary Least Squares)**

Given the column space of the matrix  $A$ , for vector  $\vec{b}$  not in the said column space,  $A\vec{x} - \vec{b} = \vec{e}$  must be orthogonal to the columns of  $A$ . (Pythagora's theorem)

Therefore, the dot products of every column of  $A$  and  $\vec{e}$  must be zero, i.e.

$$\begin{aligned} A^T(A\vec{x} - \vec{b}) &= 0 \\ A^T A\vec{x} - A^T \vec{b} &= 0 \\ A^T A\vec{x} &= A^T \vec{b} \\ \vec{x} &= (A^T A)^{-1} A^T \vec{b} \end{aligned}$$

We conclude that the solution for Ordinary Least Squares (OLS) is

$$\vec{x}^* = \underset{\vec{x}}{\operatorname{argmin}} \|A\vec{x} - \vec{b}\|^2 = (A^T A)^{-1} A^T \vec{b}$$

**1.b. Norm****Definition 1.4 (Norm)**

A Norm is defined as

$$f := \mathbf{X} \rightarrow \mathbb{R}$$

For vector space  $\mathbf{X}$ .

The norm of  $x$  is denoted as  $\|x\|$ .

For any vector  $x$  and  $y$ , we have

- $\|x\| \geq 0$  and  $\|x\| = 0$  iff  $x = \vec{0}$
- $\|x + y\| \leq \|x\| + \|y\|$
- $\|\alpha x\| = |\alpha| * \|x\|$

**Definition 1.5 (l-p Norm)**

Generally, l-p norm is defined as

$$\|\vec{x}\|_p := \left( \sum |x_i|^p \right)^{\frac{1}{p}}; \quad 1 \leq p < \infty$$

Commonly used norms:

- $\|\vec{x}\|_1 := \sum |x_i|$
- $\|\vec{x}\|_2 := \sqrt{\sum |x_i|^2}$
- $\|\vec{x}\|_\infty := \max |x_i|$

**Theorem 1.6** (Cauchy-Schwartz Inequality)

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^\top \vec{y} = \|\vec{x}\|_2 \|\vec{y}\|_2 \cos \theta$$

Since  $-1 \leq \cos \theta \leq 1$ ,

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^\top \vec{y} \leq \|\vec{x}\|_2 \|\vec{y}\|_2$$

**Theorem 1.7** (Holder's Inequality)

For  $p, q \geq 1$  s.t.  $\frac{1}{p} + \frac{1}{q} = 1$ ,

$$|\vec{x}^\top \vec{y}| \leq \sum_{i=1}^n |x_i y_i| \leq \|\vec{x}\|_p \|\vec{y}\|_q$$

i.e., Cauchy-Schwartz is a narrowed case of Holder's Inequality.

**1.c. Gram-Schmidt****Theorem 1.8** (Gram-Schmidt/QR-decomposition)

Let  $X$  be a vector space with basis  $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\}$ , which is orthonormal. For any matrix  $A$ ,

$$A = QR$$

$$[\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n] = [\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n] \begin{bmatrix} \vec{r}_{11} & \vec{r}_{12} & \cdots & \vec{r}_{1n} \\ 0 & \vec{r}_{22} & \cdots & \vec{r}_{2n} \\ 0 & 0 & \ddots & \vec{r}_{3n} \\ 0 & 0 & 0 & \vec{r}_{nn} \end{bmatrix}$$

Where  $Q$  is orthonormal and  $R$  is upper-triangular.

**Theorem 1.9** (Fundamental Theorem of Linear Algebra)

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

$$\text{Null}(A) \oplus \text{Range}(A^\top) = \mathbb{R}^n$$

Where  $\oplus$  denotes "direct sum" and  $\text{Range}(A^\top)$  is the column space of  $A^\top$ . With the said equation we can also conclude that

$$\text{Range}(A) \oplus \text{Null}(A^\top) = \mathbb{R}^m$$

**Theorem 1.10** (orthogonal decomposition theorem)

$X$  a vector space and  $S$  a subspace of  $X$ . Then for any  $\vec{x}$  in  $X$ ,

$$\vec{x} = \vec{s} + \vec{r}, \quad \vec{s} \in S, \quad \vec{r} \in S^\perp$$

Such that

$$S^\perp = \{\vec{r} \mid \langle \vec{r}, \vec{s} \rangle = 0, \quad \forall \vec{s} \in S\}$$

Therefore,

$$X = S \oplus S^\perp$$

**Example 1.11** (Minimum Norm Problem)

We want to find

$$\min \|\vec{x}\|_2^2$$

subject to  $A\vec{x} = \vec{b}$ . From FTLA we know that

$$\vec{x} = \vec{y} + \vec{z} \quad s.t. \quad \vec{y} \in N(A); \quad \vec{z} \in R(A^\top).$$

And

$$A(\vec{y} + \vec{z}) = 0 + A\vec{z} = \vec{b}$$

Since  $\vec{y} \perp \vec{z}$ ,

$$\|\vec{x}\|_2^2 = \|\vec{y}\|_2^2 + \|\vec{z}\|_2^2$$

Consider  $\vec{z} = A^\top \vec{w}$ ,

$$A\vec{z} = \vec{b}$$

$$AA^\top \vec{w} = \vec{b}$$

$$\vec{w} = (AA^\top)^{-1} \vec{b}$$

Therefore

$$\vec{z} = \min \|\vec{x}\|_2^2 = A^\top (AA^\top)^{-1} \vec{b}$$

**1.d. Symmetric Matrices****Definition 1.12**

Matrix  $A$  is symmetric if  $A = A^\top$ , i.e.  $A_{ij} = A_{ji}$ .

Set  $\mathbb{S}^n$  means the set of symmetric matrices of dimension  $n$ .

**Theorem 1.13 (Spectral Theorem)**

If matrix  $A \in \mathbb{S}^n$ , then

- All eigenvalues of  $A$  are real numbers
- Eigenspaces are orthogonal
- $\dim(N(\lambda_i I - A)) = \mu_i$  where  $\mu_i$  is the algebraic multiplicity of  $\lambda_i$

This means that  $A$  is always diagonalizable. i.e.:

$$A = U\Lambda U^\top$$

where  $U$  orthonormal and  $\Lambda$  diagonal. Orthonormal (or, unitary) means that the columns of  $U$  are orthogonal and all columns are normalized, i.e.

$$U^{-1} = U^\top$$

**Theorem 1.14**

For a diagonalizable  $n \times n$  matrix  $A$  that has  $n$  linearly independent eigenvectors,  $A$  can be factorized as

$$A = U\Lambda U^\top$$

Where  $U$  orthonormal and  $\Lambda$  is a diagonal matrix consists of the eigenvalues of  $A$  such that

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_i \end{bmatrix}$$

Therefore it is also called an eigenvalue decomposition.

**1.e. Principal Component Analysis****Definition 1.15**

For  $A \in \mathbb{S}$ , its Rayleigh coefficient is defined as

$$R = \frac{\vec{x}^\top A \vec{x}}{\vec{x}^\top \vec{x}}$$

The Rayleigh coefficient can bound the eigenvalues of  $A$  such that,

$$\lambda_{\min}(A) \leq \frac{\vec{x}^\top A \vec{x}}{\vec{x}^\top \vec{x}} \leq \lambda_{\max}(A)$$

PCA is very similar to Singular Value Decomposition (SVD). SVD has more nice properties than PCA.

## 1.f. Singular Value Decomposition

### Theorem 1.16 (SVD)

Let  $A \in \mathbb{R}^{m \times n}$ , the SVD of  $A$  is given as

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

Where

$$U \in \mathbb{R}^{m \times m}, \Sigma \in \mathbb{R}^{m \times n}, V \in \mathbb{R}^{n \times n}$$

and  $\Sigma$  has real entries in its diagonal (the singular values) and zero's else where. If  $\text{Rank}(A) = r$ , we can rewrite  $A$  as

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T$$

### Remark 1.17 (geometric interpretation of SVD)

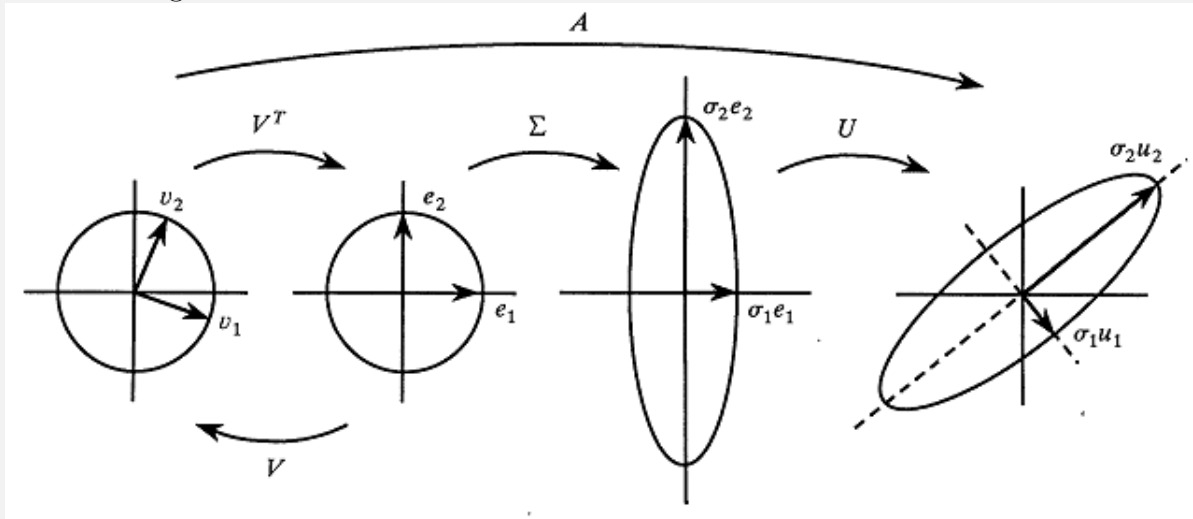
Consider linear transformation on vector  $\vec{x}$  given by matrix  $A$ , s.t.

$$A\vec{x} = U\Sigma V^T \vec{x}$$

SVD helps breaking the transformation into three smaller steps, i.e.

- orthonormal transformation (rotate/reflect) by  $V$ ,
- scaling by  $\Sigma$ ,
- orthonormal transformation by  $U$ .

The following illustration is an example of a 2D transformation  $A\vec{x}$ . It shows the decomposed linear transformation through the unit circles relative to the original unit circle at different stages of the transformation.





**Theorem 1.18** (Proof of SVD)

For  $A \in \mathbb{R}^{m \times n}$ , consider symmetric matrix  $A^\top A$  that has eigenvalues  $\lambda_1 \cdots \lambda_r > 0$  with corresponding eigenvectors  $v_1 \cdots v_r$  and  $\lambda_{r+1} \cdots \lambda_n = 0$ . Then we know that

$$A^\top A \vec{v}_i = \lambda_i \vec{v}_i$$

Let

$$V = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}$$

Define  $\sigma_i = \sqrt{\lambda_i}$ , let

$$A \vec{v}_i = \sigma_i \vec{u}_i \quad i \leq r$$

for some vector  $\vec{u}_i$ .

**Claim.**  $\vec{u}_i$  are orthonormal.

*Proof.*

$$\begin{aligned} \vec{u}_i^\top \vec{u}_j &= \frac{(A \vec{v}_i)^\top (A \vec{v}_j)}{\sigma_i \sigma_j} \\ &= \frac{1}{\sigma_i \sigma_j} \vec{v}_i^\top A^\top A \vec{v}_j & A^\top A \vec{v}_j &= \lambda_j \vec{v}_j \\ &= \frac{1}{\sigma_i \sigma_j} \vec{v}_i^\top \lambda_j \vec{v}_j \\ &= \frac{\lambda_j}{\sigma_i \sigma_j} \vec{v}_i^\top \vec{v}_j & \vec{v}_i \vec{v}_j &\text{ orthonormal} \\ &= \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \end{aligned}$$

Therefore  $\vec{u}_i$  are orthonormal. ■

Recall that  $A$  has rank  $r$ , we let

$$V_r = V = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_r \\ | & & | \end{bmatrix}$$

Hence

$$\begin{aligned} A V_r &= \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_r \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} = U_r \Sigma_r \\ A &= U \Sigma V^\top \end{aligned}$$

Since  $V$  orthonormal and  $V^{-1} = V^\top$

## 1.g. Low-Rank Approximation

### Definition 1.19 (matrix norms)

There are two ways to interpret a matrix, either as an operator or as a block of data. Frobenius norm consider the matrix as a block of data.

**Frobenius norm** of matrix  $A$  is defined as

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

Frobenius norm is invariant to orthonormal transformations, i.e. given  $U$  an orthonormal matrix,

$$\|UA\|_F = \|AU\|_F = \|A\|_F$$

**Spectral norm**, or  $l_2$  norm, interpret the matrix as an operator and is defined as

$$\|A\|_2 = \max_{\|\vec{x}\|_2=1} \|A\vec{x}\|_2 = \max_{\|\vec{x}\|_2=1} \sqrt{\vec{x}^\top A^\top A \vec{x}} = \sqrt{\lambda_{\max}(A^\top A)} = \sigma_{\max}(A^\top A)$$

Intuitively, the spectral norm of a matrix  $A$  is the largest scaling that  $A$  can do (recall the  $\Sigma$  matrix that is used to scale the unit circle in the three steps of transformation after SVD).

### Theorem 1.20 (Eckart-Young-Mirsky Theorem)

$A \in \mathbb{R}^{m \times n}$ . Do SVD gives us

$$A = U \Sigma V^\top = \sum_{i=1}^n \sigma_i \vec{u}_i \vec{v}_i^\top$$

Define

$$A_k = \sum_{i=1}^k \sigma_i \vec{u}_i \vec{v}_i^\top$$

We want to find the best  $k$ -rank (lower than  $r$ ) approximation of  $A$ , i.e.

$$\underset{B \in \mathbb{R}^{m \times n}, \text{Rank}(B)=k}{\text{argmin}} \|A - B\|_F$$

Surprisingly, Eckart-Young-Mirsky Theorem tells us that

$$\underset{B \in \mathbb{R}^{m \times n}, \text{Rank}(B)=k}{\text{argmin}} \|A - B\|_F = A_k$$

Moreover,

$$\underset{B \in \mathbb{R}^{m \times n}, \text{Rank}(B)=k}{\text{argmin}} \|A - B\|_2 = A_k$$

This theorem relates two completely different norms and is not obvious at all. It shows how fundamental SVD is, such that in any way of looking at a matrix, the decomposition shows up.

**Remark 1.21**

Eckart-Young-Mirsky Theorem can be used to **compress images**. For an image, the matrix that represents the pixels of the image can be reduced to a lower rank matrix, and hence a smaller set of data, while remains relatively high resolution. The  $A_k$  matrix **captures the key features of the image because it keeps  $k$  largest singular values and their corresponding vectors that contribute most to the dataset/transformation.**

**Definition 1.22** (trace)

The trace of a matrix is defined as

$$\text{trace} := \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$$

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

**Remark 1.23** (Orthonormal transformation invariance of Frobenius norm)

Proof that  $\|UA\|_F = \|AU\|_F = \|A\|_F$

*Proof.* Recall that  $\|A\|_F = \sqrt{\text{tr}(A^\top A)}$ . By definition, for any matrices A and B, we have  $\text{tr}(AB) = \text{tr}(BA)$ . Then,

$$\begin{aligned} \|AU\|_F &= \sqrt{\text{tr}((AU)^\top (AU))} \\ &= \sqrt{\text{tr}(U^\top A^\top AU)} \\ &= \sqrt{\text{tr}(UU^\top A^\top A)} \\ &= \sqrt{\text{tr}(A^\top A)} \\ &= \|A\|_F \end{aligned}$$

■

**Remark 1.24** (Frobenius norm is the sum of singular values)

$$\begin{aligned} \|A\|_F &= \|U\Sigma V^\top\|_F = \|\Sigma\|_F \\ &= \sqrt{\sum_{i=1}^n \sigma_i^2} \end{aligned}$$