CS 127/227A, Fall 2022

Optimization Models in Engineering

Gireeja Ranade, UC Berkeley
Zhiyu An

Contents

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

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

Where

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

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.

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

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

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

1.b. Norm

Definition 1.4 (Norm)

A Norm is defined as

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

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)

$$< \vec{x}, \vec{y} >= \vec{x}^{\top}\vec{y} = \|\vec{x}\|_2\|\vec{y}\|_2 \cos\theta$$

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

$$< \vec{x}, \vec{y} >= \vec{x}^{\top}\vec{y} \leq \|\vec{x}\|_2\|\vec{y}\|_2$$

Theorem 1.7 (Holder's Inequality)

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

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

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

1.c. Gram-Schimdt

Theorem 1.8 (Gram-Schimdt/QR-decomposition)

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

$$A = QR$$

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

Where Q is orthonormal andR is upper-triangular.

Theorem 1.9 (Fundamental Theorem of Linear Algebra)

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

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

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

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

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

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

Therefore,

$$\mathbf{X} = S \bigoplus S^{\perp}$$

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

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

And

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

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

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

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

$$\begin{aligned} A\vec{z} &= \vec{b} \\ AA^{\intercal}\vec{w} &= \vec{b} \\ \vec{w} &= (AA^{\intercal})^{-1}\vec{b} \end{aligned}$$

Therefore

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

1.d. Symmetric Matrices

Definition 1.12
Matrix A is symmetric if $A = A^{\intercal}$, 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 $\in \mathbb{S}^{\ltimes}$, then

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

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

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

where U orthonormal and Λ 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*n matrix A that has n linearly independent eigenvectors, A can be factorized as

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

Where U orthonormal and Λ 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 $A \in \mathbb{R}^{m*n}$, the SVD of A is given as

$$A = U\Sigma V^{\top}$$

Where

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

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

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

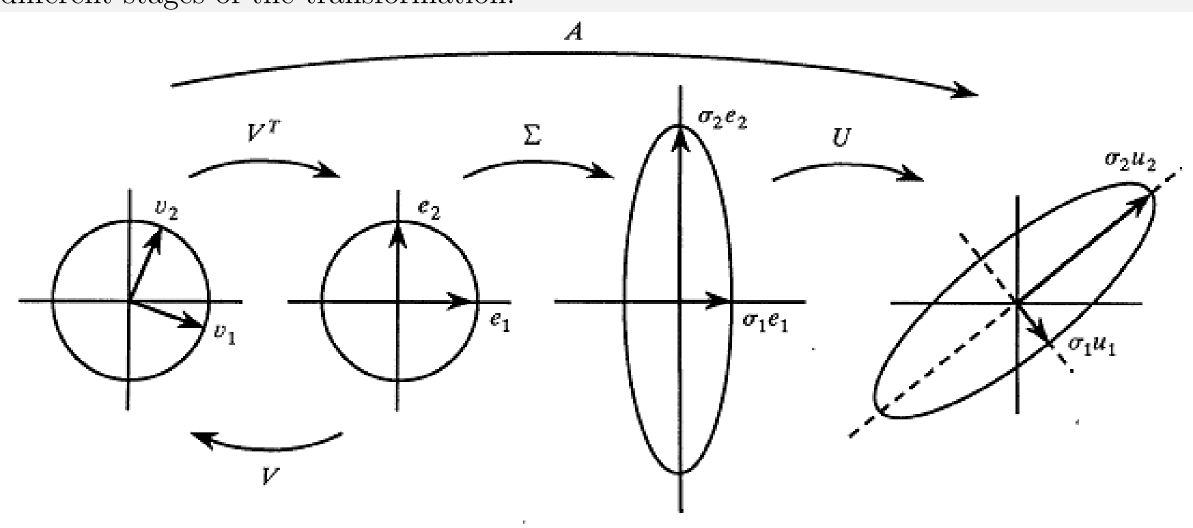
Remark 1.17
Consider linear transformation on vector $\vec{x}$ given by matrix A, s.t.

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

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

- orthonormal transformation (rotate/reflect) by V,
- scaling by Σ,
- 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 $A \in \mathbb{R}^{m*n}$, 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.

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

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

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

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