Linear Algebra & Convex Optimization – Lecture 6

Text : Introduction to Applied Linear Algebra, S. Boyd: Chapters 8, 11.

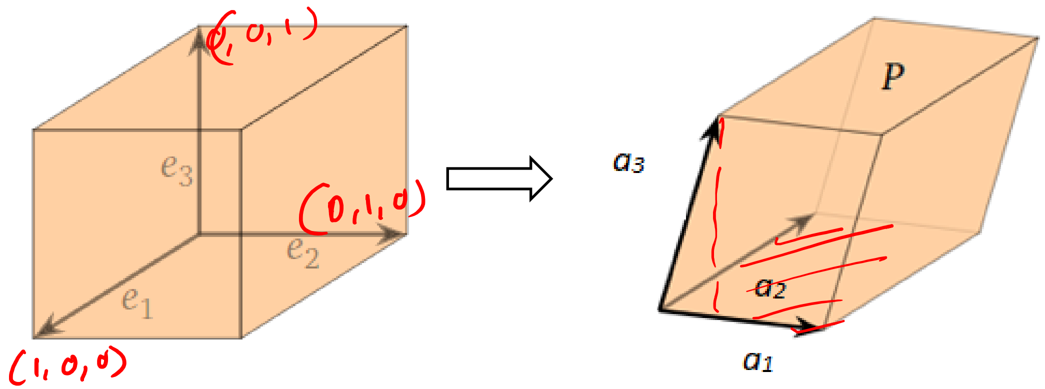
Reference: Linear Algebra, Gilbert Strang

Determinant & Singularity

Determinant is a scalar value associated with every $n \times n$ square matrix $A: \mathbb{R}^n \to \mathbb{R}^n$

$$A = [\ \mathbf{a}_1 \mid \mathbf{a}_2 \mid \cdots \mid \mathbf{a}_n\ ]$$

$$A\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} = \mathbf{a}_1, \quad A\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix} = \mathbf{a}_2, \quad \ldots, \quad A\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix} = \mathbf{a}_n$$

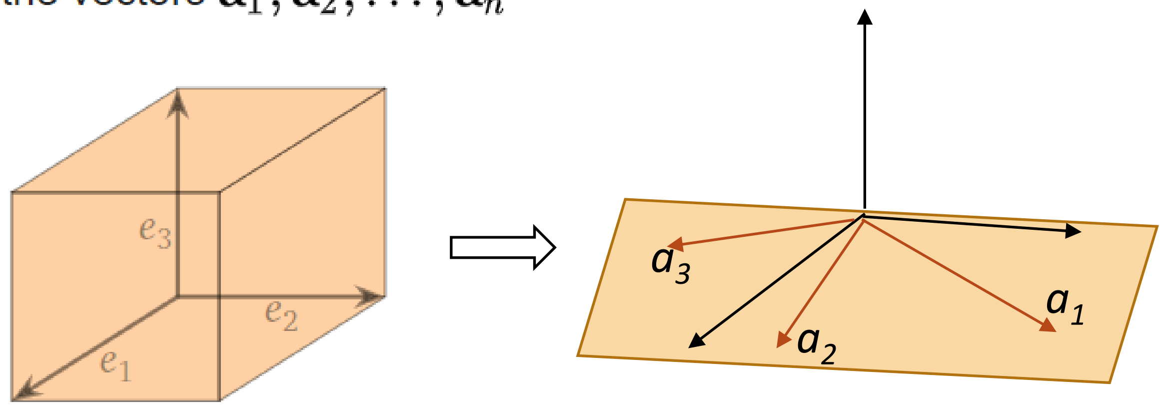


A maps the unit n-cube to the n-dimensional parallelotope defined by the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$

$$|\det(A)| = \mathrm{vol}(P).$$

Determinant is measure of linear transformation induced by A

If columns of A are linearly dependent, det(A) = 0



parallelotope is ‘squashed’ for Singular Matrices

System of Linear Equations

$$
\begin{aligned}
A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= b_1 \\
A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= b_2 \\
&\vdots \\
A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n &= b_m
\end{aligned}
$$

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & & \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

$$Ax = b$$

$m \times n$ Matrix A: *Coefficient Matrix*

Example 1:

$$x_1 + x_2 = 1$$
$$x_1 = -1$$
$$x_1 - x_2 = 0$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}$$

3×2

$$b = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

3×1

No Solution in this case

Example 2:

$$x_1 + x_2 = 1$$
$$x_2 + x_3 = 2$$

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

2×3

$$b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

2×1

Many Solutions !

$$x = (1, 0, 2)$$
$$x = (0, 1, 1)$$

Ax = b , A is square and invertible

SOLVING LINEAR EQUATIONS VIA QR FACTORIZATION

given an $n \times n$ invertible matrix A and an n-vector b.

1. *QR factorization.* Compute the QR factorization $A = QR$.
2. Compute $Q^T b$.
3. *Back substitution.* Solve the triangular equation $Rx = Q^T b$ using back substitution.

Back-substitution

$$
\begin{aligned}
R_{11}x_1 + R_{12}x_2 + \cdots + R_{1,n-1}x_{n-1} + R_{1n}x_n &= b_1 \\
&\vdots \\
R_{n-2,n-2}x_{n-2} + R_{n-2,n-1}x_{n-1} + R_{n-2,n}x_n &= b_{n-2} \\
R_{n-1,n-1}\underline{x_{n-1}} + R_{n-1,n}x_n &= b_{n-1} \\
R_{nn}\underline{x_n} &= b_n. \checkmark
\end{aligned}
$$

$$x_{n-1} = (b_{n-1} - R_{n-1,n}x_n)/R_{n-1,n-1}.$$

$Ax = b$

$QRx = b$

$Q^T QRx = Q^T b$

$Rx = Q^T b$

$$\begin{bmatrix} q_1^T b \\ q_2^T b \\ \vdots \\ q_n^T b \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

$x^2 = -1$

$x^3 = -x$

Overdetermined & Underdetermined Systems

$$Ax = b$$

		Coefficient Matrix, A	Solution
Overdetermined Systems	More Equations than unknowns	m > n, Tall	No Solution when $b \notin C(A)$
Underdetermined Systems	Less Equations than unknowns	m < n, Wide	Many Solutions

Ax = b with no solutions

$$\begin{array}{ccc} A & x & b \\ \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = & \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \end{array}$$

Usually $b \notin C(A)$

Then no Solutions !

What is the best we can do ?

Find a vector p in Column space of A that is the closest to vector b

Solve for $A\hat{x} = p$

Projections

Projection of a vector b to a Sub-space S = "closest' point p in the sub-space to the vector b

$$e = (b - p) \perp S \qquad p = c_1 a$$

$$a^T(b - c_1 a) = 0 \implies c_1 = \frac{a^T b}{a^T a}$$

$$p = a\frac{a^T b}{a^T a}$$

$$= \frac{aa^T}{a^T a} b = P\,b$$

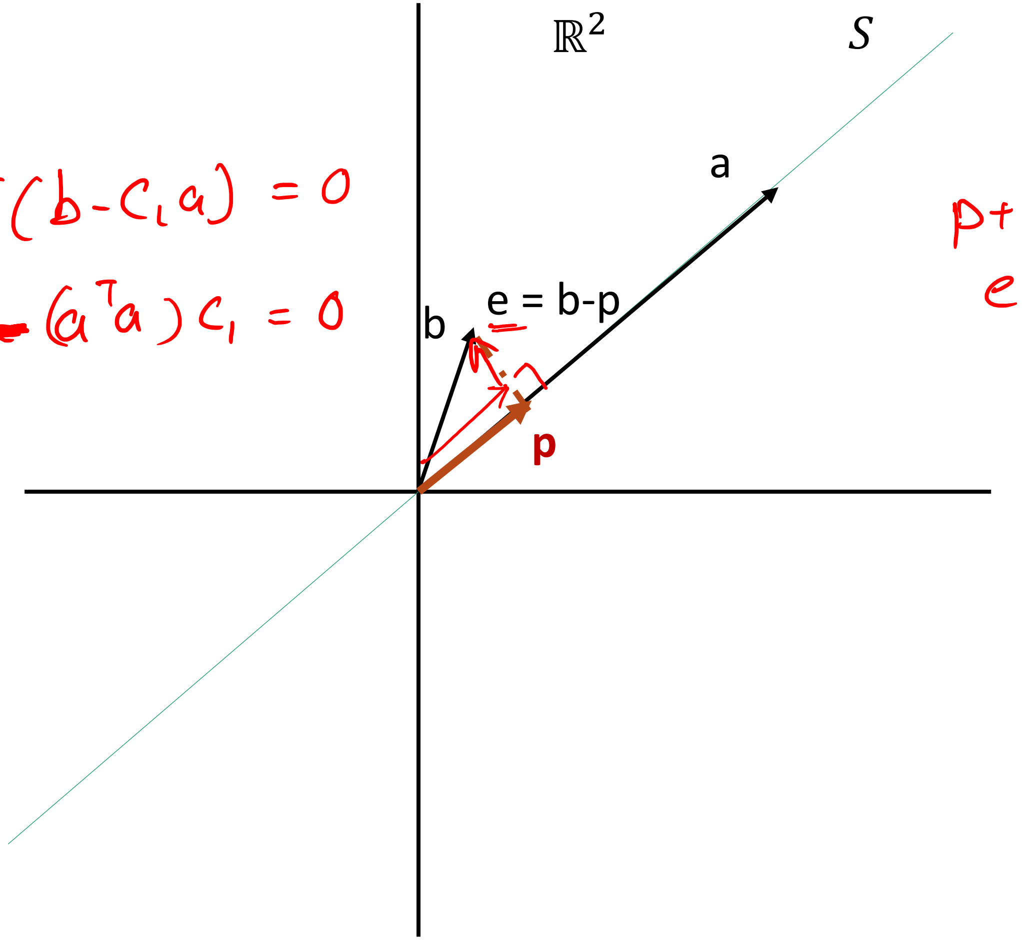
P is the projection matrix.

$$P(Pb) = P^2 b = p$$

$$P^2 = \frac{aa^T}{a^T a}\cdot\frac{aa^T}{a^T a} = \frac{aa^T}{a^T a} = P$$

$e = b - p$
$= b - c_1 a$
$a^T e = 0 \quad a^T(b - c_1 a) = 0$
$a^T b - (a^T a) c_1 = 0$

$Pp = p$



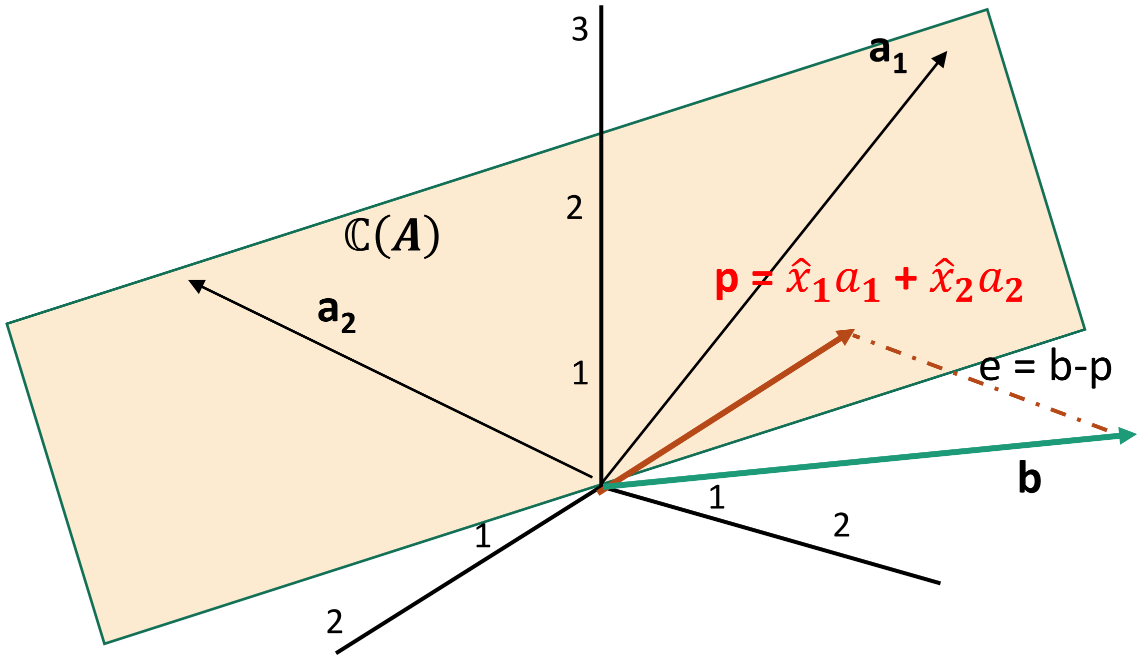
$p + e = b$
$e = b - p$

Projection Method for best solution

Problem: Best Solution for A**x**=b when $b \notin C(A)$

$$\underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix} \qquad \mathbf{A}x = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$Ax = b$



If $b \in C(A)$, $Pb = b$

$$p = \underline{\hat{x}_1} a_1 + \underline{\hat{x}_2} a_2 = A\hat{x}$$

$$\mathrm{e} = b - p = (b - A\hat{x}) \perp \text{plane } C(A)$$

$\leftarrow e \rightarrow$

$$\begin{aligned} a_1^T(b - A\hat{x}) &= 0 \\ a_2^T(b - A\hat{x}) &= 0 \end{aligned} \Rightarrow \begin{bmatrix} \leftarrow a_1^T \rightarrow \\ \leftarrow a_2^T \rightarrow \end{bmatrix} \begin{bmatrix} (b - A\hat{x}) \end{bmatrix} \Rightarrow A^T(b - A\hat{x}) = 0$$

$A^Tb - A^TA\hat{x} = 0$

$A^TA\hat{x} = A^Tb$

$$A^TA\hat{x} = A^Tb$$

$$\hat{x} = \underbrace{(A^TA)^{-1}}A^Tb$$

$$p = A\hat{x} = \underbrace{A\underbrace{(A^TA)^{-1}A^Tb}_{\hat{x}}}_{P}$$

$$= \underline{Pb}$$

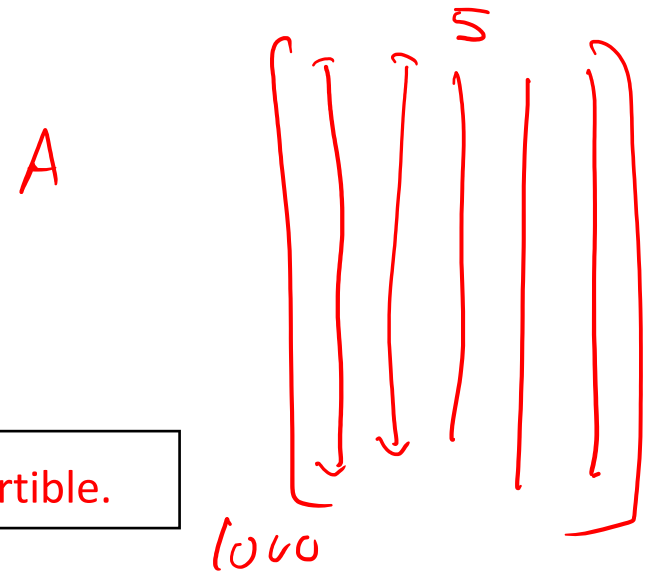
$\hat{x} = (A^TA)^{-1} A^T b$

The best approximate solution for A**x**=b when $b \notin C(A)$ is

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$$

Ax= b : Gram Matrix $A^T A$

- If A has linearly independent columns, A is tall or square
- A has left Inverse



If columns of Matrix A are linearly independent, the Gram Matrix $A^T A$ is invertible.

suppose that the columns of A are linearly independent. $\Longrightarrow$ $Ax = 0$ is true only if $x = 0$

Let x be an n-vector which satisfies $(A^T A)x = 0$.

$x^T A^T A x = 0$

$(Ax)^T (Ax) = 0$

$0 = x^T 0 = x^T (A^T A x) = x^T A^T A x = \|Ax\|^2$ $\Longrightarrow$ $Ax = 0$ $\Longrightarrow$ $x = 0$

$\|Ax\| = 0$

Hence, if tall Matrix A has linearly independent columns, the Square Matrix , $A^T A$ has independent columns OR $A^T A$ is invertible

Pseudo Inverse

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

- Matrix A is tall or square
- If columns of Matrix A are linearly independent, A has left Inverse.
- If columns of Matrix A are linearly independent, the Gram Matrix $A^T A$ is invertible

$(A^TA)^{-1}(A^T A) = I$

Matrix $(A^T A)^{-1} A^T$ is the left inverse of Matrix A

$(A^TA)^{-1} = \underbrace{A^{-1}(A^T)^{-1}}$

$$((A^T A)^{-1} A^T)\, A = (A^T A)^{-1} (A^T A) = I$$

Moore –Penrose Psuedo Inverse: $A^\dagger = (A^T A)^{-1} A^T = A^{-1}$

When A is square, the pseudo-inverse $A^\dagger$ reduces to the ordinary inverse:

$$A^\dagger = (A^T A)^{-1} A^T = A^{-1} A^{-T} A^T = A^{-1} I = A^{-1}$$

$(A^T)^{-1}$ does not exist when the matrix is not square

Poll: If A has independent columns, A can have many left inverses. It also means A can have many Moore-Penrose Pseudo Inverses? TRUE / FALSE ?

Least Squares

Problem Setting:

We have $Ax = b$

Matrix A is tall with linearly independent columns, $b \notin C(A)$

Objective: Seek an x for which the residual, $r = Ax - b$ is minimal.

$$\text{minimize} \quad \|Ax - b\|^2 = \|r\|^2 = r_1^2 + \cdots + r_m^2$$

If $\|A\hat{x} - b\|^2 < \|Ax - b\|^2$ is true for all x , then $\hat{x}$ is a *solution*

solution $\hat{x}$ of $Ax = b$ need not satisfy the equations $A\hat{x} = b$;

Least Squares: Column/Row Interpretations

Column Interpretation:

$$\|Ax - b\|^2 = \|(x_1 a_1 + \cdots + x_n a_n) - b\|^2$$

If $\hat{x}$ is a solution of the least squares problem, then the vector

$$A\hat{x} = \hat{x}_1 a_1 + \cdots + \hat{x}_n a_n$$

is closest to the vector b, among all linear combinations of the vectors $a_1, \ldots, a_n$.

Row Interpretation:

$$\begin{bmatrix} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \vdots \\ \tilde{a}_m^T \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

Suppose the rows of A are the n-row-vectors $\tilde{a}_1^T, \ldots, \tilde{a}_m^T$,

$$r_i = \tilde{a}_i^T x - b_i, \quad i = 1, \ldots, m.$$

Minimize $$\|Ax - b\|^2 = (\tilde{a}_1^T x - b_1)^2 + \cdots + (\tilde{a}_m^T x - b_m)^2$$

Poll: Which interpretation fits in a classifier training setting.
A) Row B) Column

The least squares objective is to find the $\hat{x}$ that minimizes the sum of residuals

Least Squares : Loss Function Optimization

$f(x) : \mathbb{R}^n \to \mathbb{R}^m$

$\mathbb{R}^{1000} \to \mathbb{R}^1$

What is n & m
if my system has
1000 weight parameters

$$f(x) = \|Ax - b\|^2 = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j - b_i\right)^2.$$

$(Ax-b)^T(Ax-b)$

For $\hat{x}$ to become optimal point $\frac{\partial f}{\partial x_i}(\hat{x}) = 0, \quad i = 1, \ldots, n,$

Once we solve the equation for $\hat{x}$, we get the same solution,

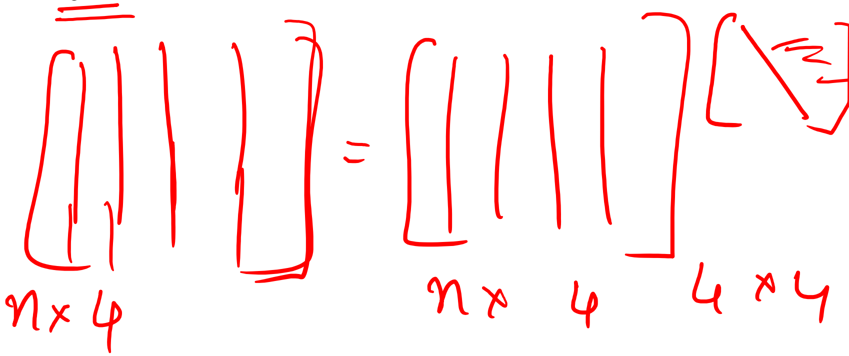
$$\hat{x} = (A^TA)^{-1}A^Tb$$

(Details of the differentiation will be re-visited after introducing vector calculus)

Pseudo Inverse via QR Factorization

$(A^T A)^{-1} A^T$

- If columns of Matrix A are linearly independent, A has left Inverse.
- If columns of Matrix A are linearly independent ,QR Factorization of Matrix A, $A = QR$ exists.



$$A^T A = (QR)^T (QR)$$

$$= R^T Q^T Q R = R^T R$$

$$A^\dagger = (A^T A)^{-1} A^T$$

$$= (R^T R)^{-1} (QR)^T \quad = R^{-1} R^{-T} R^T Q^T = R^{-1} Q^T.$$

Do we have to compute pseudo inverse to get the solution $\hat{x}$?

$$A\hat{x} = QR\hat{x} = b$$

$$\hat{x} = R^{-1} Q^T b \qquad \Longrightarrow \qquad R\hat{x} = Q^T b$$

If columns of A are linearly independent, the 'approximate solution' for $Ax = b$ for can be found by solving $R\hat{x} = Q^T b$