School of Computing
National University of Singapore
CS CS5240 Theoretical Foundations of Multimedia

More Linear Algebra

Singular Value Decomposition (SVD)

"The highpoint of linear algebra" – Gilbert Strang
Any $m \times n$ matrix $\mathbf{A}$ can be decomposed into:

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

$\mathbf{U}$	: $m \times m$	: columns are left singular vectors
$\mathbf{\Sigma}$	: $m \times n$	: diagonal : singular values
$\mathbf{V}$	: $n \times n$	: columns are right singular vectors

e.g. for $m > n$

$$\mathbf{A} = \begin{bmatrix} \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m \\ \vdots & & \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{bmatrix} \begin{bmatrix} \dots & \mathbf{v}_1^\top & \dots \\ & \vdots & \\ \dots & \mathbf{v}_r^\top & \dots \\ \dots & \mathbf{v}_{r+1}^\top & \dots \\ & \vdots & \\ \dots & \mathbf{v}_n^\top & \dots \end{bmatrix}$$

$\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > 0$, $r = \text{rank}(\mathbf{A})$
Economy version $\mathbf{A} = \underbrace{\mathbf{U}_r}_{m\times r}\underbrace{\mathbf{\Sigma}_r}_{r\times r}\underbrace{\mathbf{V}_r^\top}_{r\times n}$
$\mathbf{U}$,$\mathbf{V}$ orthogonal : $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_{m\times m}$, $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_{n\times n}$
Column Space: look at $\mathbf{Ax}$

$$\begin{aligned} \mathbf{Ax} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top\mathbf{x}, \quad \text{and let } \mathbf{y} = \mathbf{V}^\top\mathbf{x} \\ &= \begin{bmatrix} \vdots & & \vdots & \vdots & & \vdots \\ \sigma_1\mathbf{u}_1 & \dots & \sigma_r\mathbf{u}_r & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \end{bmatrix} \mathbf{y} \end{aligned}$$

so $\text{Col}(\mathbf{A}) = \text{Col}(\mathbf{U}_r)$. In fact, $\mathbf{u}_1,\dots,\mathbf{u}_r$ form an orthonormal basis for $\text{Col}(\mathbf{A})$.

Nullspace: look at

$$\begin{aligned} &\mathbf{Ax} = \mathbf{0} \\ &\Rightarrow \mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^\top\mathbf{x} = 0 \end{aligned}$$

pre-multiply by $\mathbf{U}_r^\top$: $\quad \mathbf{\Sigma}_r \mathbf{V}_r^\top \mathbf{x} = 0$
pre-multiply by $\mathbf{\Sigma}_r^{-1}$: $\quad \mathbf{V}_r^\top \mathbf{x} = 0$
i.e. want $\mathbf{x}$ to be orthogonal to $\mathbf{v}_1, \ldots, \mathbf{v}_r$
That's precisely $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$, since $\mathbf{V}$ is orthogonal!
Thus, $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ form an orthonormal basis for Null($\mathbf{A}$).

Consider

$$\begin{aligned}\mathbf{A}^\top \mathbf{A} &= \left(\mathbf{U\Sigma V}^\top\right)^\top \left(\mathbf{U\Sigma V}^\top\right) \\ &= \mathbf{V\Sigma}^\top \mathbf{U}^\top \mathbf{U\Sigma V}^\top \\ &= \mathbf{V\Sigma}^2 \mathbf{V}^\top\end{aligned}$$

But this is the eigen-decomposition of $\mathbf{A}^\top\mathbf{A}$! So $\mathbf{V}$ is the eigenvector matrix of $\mathbf{A}^\top\mathbf{A}$
$\mathbf{\Sigma}^2$ is the eigenvalue matrix of $\mathbf{A}^\top\mathbf{A}$ i.e. singular values are positive square roots of eigenvalues.

Similary, consider

$$\begin{aligned}\mathbf{A}\mathbf{A}^\top &= \mathbf{U\Sigma V}^\top \mathbf{V\Sigma}^\top \mathbf{U}^\top \\ &= \mathbf{U\Sigma}^2 \mathbf{U}^\top\end{aligned}$$

So $\mathbf{U}$ is the eigenvector matrix for $\mathbf{A}\mathbf{A}^\top$ with same eigenvalues.
In general, for $m \times n$ $\mathbf{A}$:

$$\begin{aligned}\mathbf{Ax} &= \mathbf{U\Sigma V}^\top \mathbf{x} \\ &= (\text{rotate in } \mathbb{R}^m)\,(\text{ scale })\,(\text{rotate in } \mathbb{R}^n)\,\mathbf{x}\end{aligned}$$

Low-rank approximation

SVD provides the best lower-rank approximation to $\mathbf{A}$, i.e. rank k approx. $\mathbf{A}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^\top$.
The idea is to use only the first k singular values/vectors, so that $\mathbf{A}_k \approx \mathbf{A}$.

Use SVD for compression:

Instead of storing $\mathbf{A}$	: mn numbers
store $\mathbf{u}_1, \ldots, \mathbf{u}_k$	: mk numbers
+ $\sigma_1, \ldots, \sigma_k$	: k numbers
+ $\mathbf{v}_1, \ldots, \mathbf{v}_k$	: nk numbers
=	$(m+n+1)k$ numbers

Use SVD to filter noise

Typically, small singular values are caused by noise.
using rank k approx ($k < r$), removes noise.

Linear Equations Revisited: $\mathbf{Ax} = \mathbf{b}$

Key: solution only when $\mathbf{b} \in \text{Col}(\mathbf{A})$

Case 1. $\mathbf{A}$ $\quad n \times n$ and invertible. Then unique solution : $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$
rank($\mathbf{A}$) $= n$, Col($\mathbf{A}$) $= \mathbb{R}^n$

Case 2. $\mathbf{A}$ $\quad n \times n$ and singular. $\text{rank}(\mathbf{A}) = r < n$, nullity $= n - r$
Two possibilities :

(a) $\mathbf{b} \in \text{Col}(\mathbf{A})$: many solutions.

(b) $\mathbf{b} \notin \text{Col}(\mathbf{A})$: no exact solution, closest solution.

(a) $\mathbf{b} \in \text{Col}(\mathbf{A})$: SVD gives particular solution $\mathbf{x}_p$ such that $\mathbf{A}\mathbf{x}_p = \mathbf{b}$
But we can add any vector from Nullspace, $\mathbf{x}_n$, since

$$\begin{aligned} \mathbf{A}(\mathbf{x}_p + \mathbf{x}_n) &= \mathbf{A}\mathbf{x}_p + \mathbf{A}\mathbf{x}_n \\ &= \mathbf{b} + \mathbf{0} \end{aligned}$$

$\therefore$ Infinitely many solutions!
What is the SVD solution? Invert only in rank r subspace
$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ $\qquad$ (all $n \times n$)

$$\text{where } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

$$\text{Let } \mathbf{A}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^\dagger\mathbf{U}^\top, \text{ where } \boldsymbol{\Sigma}^\dagger = \begin{bmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

Then $\mathbf{x}_p = \mathbf{A}^\dagger\mathbf{b}$. $\mathbf{A}^\dagger$: pseudoinverse. See Figure 1.

(b) $\mathbf{b} \notin \text{Col}(\mathbf{A})$: No exact solution, but can find $\mathbf{b}' \in \text{Col}(\mathbf{A})$ closest to $\mathbf{b}$
Solution $\mathbf{x}' = \mathbf{A}^\dagger\mathbf{b} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top\mathbf{b}$

Case 3. $\mathbf{A}$ $\quad m \times n$ with $m < n$ "underconstrained" fewer equations than unknowns.
$r = \text{rank}(\mathbf{A}) \leq \min(m,n)$, i.e. $r < n$, so Nullspace is not trivial. $\text{Col}(\mathbf{A}) \subseteq \mathbb{R}^m$
Situation similar to the previous case, either $\mathbf{b} \in \text{Col}(\mathbf{A})$ or $\mathbf{b} \notin \text{col}(\mathbf{A})$
In practice, usually $r = m$, so that $\mathbf{b} \in \text{Col}(\mathbf{A})$, i.e. many solutions

Case 4. $\mathbf{A}$ $\quad m \times n$ with $m > n$ "overconstrained", more equations that unknowns. rank, r, is at most, n. Therefore, $\text{Col}(\mathbf{A}) \subset \mathbb{R}^m$
Again, depends on whether $\mathbf{b} \in \text{col}(\mathbf{A})$, so we can only find "closest" or "least squares" solution. $\mathbf{x}' = \mathbf{A}^\dagger\mathbf{b}$

Pseudoinverse

$\mathbf{A}^\dagger$ solves $\mathbf{A}\mathbf{x} = \mathbf{b}$ in least squares sense, i.e $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ is minimum.

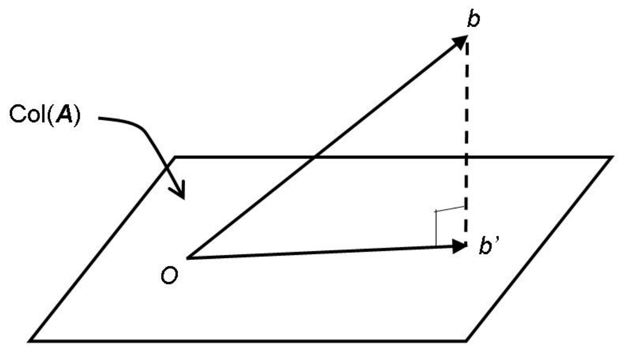

Figure 1: A singular matrix $\mathbf{A}$ has $\mathrm{Col}(\mathbf{A}) \subset \mathbb{R}^n$. This is represented by a plane in the diagram. If $\mathbf{b}$ lies outside of $\mathrm{Col}(\mathbf{A})$, then the best one can do is to obtain $\mathbf{b}'$, which is the vector in $\mathrm{Col}(\mathbf{A})$ that is closest to $\mathbf{b}$. This is what the pseudoinverse computes: $\mathbf{b}' = \mathbf{A}\mathbf{x}'$, where $\mathbf{x}' = \mathbf{A}^\dagger \mathbf{b}$.

$$\begin{aligned}\mathbf{A}^\dagger &= \mathbf{V}\boldsymbol{\Sigma}^\dagger\mathbf{U}^\top \quad \text{(using SVD)} \\ &= \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top \quad \text{but this requires } \mathrm{rank}(\mathbf{A}) = n\end{aligned}$$

Note: $\mathbf{A}^\dagger\mathbf{A} = \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\mathbf{A} = \mathbf{I}$, but $\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}\left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top \neq \mathbf{I}$ in general. Thus, pseudoinverse is only a left inverse, not a right inverse.
If $\mathbf{A}$ invertible, then pseudoinverse = true inverse:

$$\begin{aligned}\mathbf{A}^\dagger &= \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top \\ &= \mathbf{A}^{-1}\mathbf{A}^{-\top}\mathbf{A}^\top = \mathbf{A}^{-1}\end{aligned}$$

In Matlab, always use $A \setminus b$ to solve $Ax = b$. "\" will compute $\mathbf{A}^{-1}$ or $\mathbf{A}^\dagger$ accordingly.

Matrix Inversion Formulas

Excerpt from: *Statistical Signal Processing*, by Louis L.Scharf, Addison Wesley, 1991.

1. Lemma 1 (Inverse of a Partitioned Matrix)
   Let $\mathbf{R}$ denote the partitioned matrix

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]$$

   The inverse of $\mathbf{R}$ is

$$\mathbf{R}^{-1} = \left[\begin{array}{c|c} \mathbf{E}^{-1} & \mathbf{F}\mathbf{H}^{-1} \\ \hline \mathbf{H}^{-1}\mathbf{G} & \mathbf{H}^{-1} \end{array}\right]$$

$$
\begin{aligned}
\mathbf{E} &= \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \\
\mathbf{A}\mathbf{F} &= -\mathbf{B} \\
\mathbf{G}\mathbf{A} &= -\mathbf{C} \\
\mathbf{H} &= \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}
\end{aligned}
$$

All indicated inverses are assumed to exist. The matrix $\mathbf{E}$ is called Schur complement of $\mathbf{A}$, and the matrix $\mathbf{H}$ is called the Schur complement of $\mathbf{D}$.

2. Lemma 2 (Matrix Inversion Lemma)
   Let $\mathbf{E}$ denote the Schur complement of $\mathbf{A}$:

$$\mathbf{E} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$$

Then the inverse of $\mathbf{E}$ is

$$
\begin{aligned}
\mathbf{E}^{-1} &= \mathbf{A}^{-1} + \mathbf{F}\mathbf{H}^{-1}\mathbf{G} \\
\mathbf{A}\mathbf{F} &= -\mathbf{B} \\
\mathbf{G}\mathbf{A} &= -\mathbf{C} \\
\mathbf{H} &= \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}
\end{aligned}
$$

Lemmas 1 and 2 combine to form the following representation for the inverse of a partitioned matrix.

Theorem (Partitioned Matrix Inverse)

The inverse of the partitioned matrix

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]$$

is the matrix

$$\mathbf{R}^{-1} = \left[\begin{array}{c|c} \mathbf{A}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] + \left[\begin{array}{c} \mathbf{F} \\ \hline \mathbf{I} \end{array}\right] \left[\mathbf{H}^{-1}\right] \left[\begin{array}{c|c} \mathbf{G} & \mathbf{I} \end{array}\right]$$

$$
\begin{aligned}
\mathbf{A}\mathbf{F} &= -\mathbf{B} \\
\mathbf{G}\mathbf{A} &= -\mathbf{C} \\
\mathbf{H} &= \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}
\end{aligned}
$$

Corollary: Woodbury's Identity

The inverse of the matrix

$$\mathbf{R} = \mathbf{R}_0 + \gamma^2 \mathbf{u}\mathbf{u}^\top$$

is the matrix

$$\mathbf{R}^{-1} = \mathbf{R}_0^{-1} - \frac{\gamma^2}{1 + \gamma^2 \mathbf{u}^\top \mathbf{R}_0^{-1} \mathbf{u}} \mathbf{R}_0^{-1} \mathbf{u}\mathbf{u}^\top \mathbf{R}_0^{-1}$$

Projections

Often we want to project $\mathbf{x}$ onto some subspace, i.e. find $\mathbf{y}$ in subspace, "closest" to $\mathbf{x}$. Geometrically, this occurs when $\mathbf{x} - \mathbf{y}$ is orthogonal to subspace. Often the subspace of interest is $\mathrm{Col}(\mathbf{A})$. Recall that in the SVD of $\mathbf{A}$, $\mathbf{U}_r$ form an orthogonal basis for $\mathrm{Col}(\mathbf{A})$.

The projection matrix $\mathbf{P}_A$ that projects any vector onto $\mathrm{Col}(\mathbf{A})$ is :

$$\begin{aligned} \mathbf{P}_A &= \mathbf{U}_r \mathbf{U}_r^\top && \text{(SVD)} \\ &= \mathbf{A} \left(\mathbf{A}^\top \mathbf{A} \right)^{-1} \mathbf{A}^\top \end{aligned}$$

e.g. To project onto a line (vector) $\mathbf{u}$, $\mathbf{P}_u = \dfrac{\mathbf{u}\mathbf{u}^\top}{\mathbf{u}^\top \mathbf{u}}$.

In general, a projection matrix $\mathbf{P}$ is one that satisfies:

1. $\mathbf{P}^\top = \mathbf{P}$ symmetric
2. $\mathbf{P}^2 = \mathbf{P}$ idempotent

What are the eigenvalues of $\mathbf{P}$?

Derivatives

		Differentiate scalar	vector	matrix
w.r.t	scalar	scalar	vector	matrix
	vector	vector	matrix	
	matrix	matrix		

scalar—scalar: e.g. $\dfrac{d}{dx} x^2 = 2x$
vector—scalar: e.g.

$$\begin{aligned} \mathbf{y} &= [\cos\theta \quad \sin^2\theta]^\top \\ \frac{d\mathbf{y}}{d\theta} &= [-\sin\theta \quad 2\sin\theta\cos\theta]^\top \end{aligned}$$

matrix—scalar: e.g.

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} x^2 & x \\ 1 & \frac{1}{x} \end{bmatrix} \\ \frac{d\mathbf{A}}{dx} &= \begin{bmatrix} 2x & 1 \\ 0 & -\frac{1}{x^2} \end{bmatrix} \end{aligned}$$

scalar—vector: $f(\mathbf{x})$ scalar function of vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\frac{df}{d\mathbf{x}} = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

vector—vector: $\mathbf{y}(\mathbf{x}) \quad m \times 1$ vector function of vector $\mathbf{x} \in \mathbb{R}^n$
Then,

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

$$\mathbf{y} : m \times 1$$
$$\mathbf{x} : n \times 1$$
$$\frac{d\mathbf{y}}{d\mathbf{x}} : n \times m \text{ matrix}$$

scalar—matrix: $f(\mathbf{A})$ scalar function of $m \times n \quad \mathbf{A}$
Then,

$$\frac{df}{d\mathbf{A}} = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1n}} \\ \frac{\partial f}{\partial a_{21}} & \cdots & \cdots & \frac{\partial f}{\partial a_{2n}} \\ \vdots & & & \vdots \\ \frac{\partial f}{\partial a_{m1}} & \cdots & \cdots & \frac{\partial f}{\partial a_{mn}} \end{bmatrix} \quad m \times n \text{ matrix}$$

Commonly used derivatives

1. $\dfrac{d}{d\mathbf{x}}(\mathbf{A}\mathbf{x}) = \mathbf{A}^\top$
2. $\dfrac{d\mathbf{x}}{d\mathbf{x}} = \mathbf{I}$
3. $\dfrac{d\mathbf{y}^\top\mathbf{x}}{d\mathbf{x}} = \dfrac{d\mathbf{x}^\top\mathbf{y}}{d\mathbf{x}} = \mathbf{y}$
4. $\dfrac{d}{d\mathbf{x}}\left(\mathbf{x}^\top\mathbf{A}\mathbf{x}\right) = \begin{cases} \left(\mathbf{A} + \mathbf{A}^\top\right)\mathbf{x} & \text{if } \mathbf{A} \text{ square} \\ 2\mathbf{A}\mathbf{x} & \text{if } \mathbf{A} \text{ symmetric} \end{cases}$
5. $\dfrac{d}{d\mathbf{x}}\left(\mathbf{u}^\top(\mathbf{x}) \quad \mathbf{v}(\mathbf{x})\right) = \left[\dfrac{d\mathbf{u}^\top}{d\mathbf{x}}\right]\mathbf{v} + \left[\dfrac{d\mathbf{v}^\top}{d\mathbf{x}}\right]\mathbf{u}$ "product rule"
6. $\dfrac{d \quad \mathrm{tr}(\mathbf{A})}{d\mathbf{A}} = \mathbf{I}$

7. $\frac{d}{d\mathbf{A}}\det(\mathbf{A}) = \det(\mathbf{A})\mathbf{A}^{-\top}$

Example: to find pseudoinverse. Let $\mathbf{e} = \mathbf{Ax} - \mathbf{b}$. We want $\mathbf{x}$ such that $\|\mathbf{e}\|_2$ smallest., i.e. $\|\mathbf{e}\|_2^2$ smallest

$$
\begin{aligned}
\text{Let } y &= \|\mathbf{e}\|_2^2 \\
&= \mathbf{e}^\top \mathbf{e} \\
&= (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b}) \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} - 2\mathbf{b}^\top \mathbf{Ax} + \mathbf{b}^\top \mathbf{b} \\
\frac{dy}{d\mathbf{x}} &= 2\mathbf{A}^\top \mathbf{Ax} - 2\mathbf{A}^\top \mathbf{b} = \mathbf{0} \\
&\Rightarrow \mathbf{A}^\top \mathbf{Ax} = \mathbf{A}^\top \mathbf{b} \\
&\Rightarrow \mathbf{x} = \underbrace{\left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top}_{\mathbf{A}^\dagger} \mathbf{b}
\end{aligned}
$$

Hessian: 2^{nd} derivative

Let $f(\mathbf{x})$ be scalar function of $\mathbf{x} \in \mathbb{R}^n$
Then Hessian:

$$
\mathbf{H} = \frac{d^2 f}{d\mathbf{x}^2} = \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
 & & \vdots & \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n^2}
\end{bmatrix}
$$

Hessian is symmetric.

Positive semi-definite (psd)

A square matrix $\mathbf{A}$ is positive semi-definite if $\mathbf{x}^\top \mathbf{Ax} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$. Positive definite $\mathbf{x}^\top \mathbf{Ax} > 0$

Note: $\mathbf{A}$ is a psd means all eigenvalues ≥ 0.

If a Hessian matrix is psd, then f has minimum point.
e.g. in the pseudoinverse calculation, $\frac{dy}{d\mathbf{x}} = 2\mathbf{A}^\top \mathbf{Ax} - 2\mathbf{A}^\top \mathbf{b}$
So Hessian, $\mathbf{H} = \frac{d}{d\mathbf{x}}\left(\frac{dy}{d\mathbf{x}}\right) = 2\mathbf{A}^\top \mathbf{A}$
Now, for any $\mathbf{x} \neq \mathbf{0}, \mathbf{x}^\top \mathbf{Hx} = 2\mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} = 2\|\mathbf{Ax}\|^2 \geq 0$ since $\|\mathbf{Ax}\|^2$ is the squared norm.
So $\mathbf{H}$ is psd. $\Rightarrow y$ has minimum point. This justifies taking derivatives to find best $\mathbf{x}$