Chapter 05
Linear Least-squares

Lin ZHANG
School of Computer Science and Technology
Tongji University

Outline

- Matrix Differentiation
- Lagrange Multiplier
- Least-squares for Homogeneous Linear Systems
- Least-squares for Inhomogeneous Linear Systems
  - An SVD-based approach

Matrix differentiation

- Function is a vector and the variable is a scalar

$$\mathbf{f}(t) = \left[f_1(t), f_2(t), \ldots, f_n(t) \right]^T$$

Definition

$$\frac{d\mathbf{f}}{dt} = \left[\frac{df_1(t)}{dt}, \frac{df_2(t)}{dt}, \ldots, \frac{df_n(t)}{dt} \right]^T$$

Matrix differentiation

- Function is a matrix and the variable is a scalar

$$\mathbf{F}(t)=\begin{bmatrix} f_{11}(t)\ f_{12}(t),\dots,f_{1m}(t) \\ f_{21}(t)\ f_{22}(t),\dots,f_{2m}(t) \\ \vdots \\ f_{n1}(t)\ f_{n2}(t),\dots,f_{nm}(t) \end{bmatrix}=\left[f_{ij}(t)\right]_{n\times m}$$

Definition

$$\frac{d\mathbf{F}}{dt}=\begin{bmatrix} \frac{df_{11}(t)}{dt}\ \frac{df_{12}(t)}{dt},\dots,\frac{df_{1m}(t)}{dt} \\ \frac{df_{21}(t)}{dt}\ \frac{df_{22}(t)}{dt},\dots,\frac{df_{2m}(t)}{dt} \\ \vdots \\ \frac{df_{n1}(t)}{dt}\ \frac{df_{n2}(t)}{dt},\dots,\frac{df_{nm}(t)}{dt} \end{bmatrix}=\left[\frac{df_{ij}(t)}{dt}\right]_{n\times m}$$

Matrix differentiation

- Function is a scalar and the variable is a vector

$$f(\mathbf{x}), \mathbf{x} = (x_1, x_2, \ldots, x_n)^T$$

Definition

$$\frac{df}{d\mathbf{x}} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^T$$

In a similar way,

$$f(\mathbf{x}), \mathbf{x} = (x_1, x_2, \ldots, x_n)$$

$$\frac{df}{d\mathbf{x}} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]$$

Matrix differentiation

- Function is a vector and the variable is a vector

$$\mathbf{x} = \left[x_1, x_2, \ldots, x_n\right]^T, \mathbf{y} = \left[y_1(\mathbf{x}), y_2(\mathbf{x}), \ldots, y_m(\mathbf{x})\right]^T$$

Definition

$$\frac{d\mathbf{y}}{d\mathbf{x}^T} = \begin{bmatrix} \frac{\partial y_1(\mathbf{x})}{\partial x_1}, \frac{\partial y_1(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial y_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial y_2(\mathbf{x})}{\partial x_1}, \frac{\partial y_2(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial y_2(\mathbf{x})}{\partial x_n} \\ \vdots \\ \frac{\partial y_m(\mathbf{x})}{\partial x_1}, \frac{\partial y_m(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial y_m(\mathbf{x})}{\partial x_n} \end{bmatrix}_{m \times n}$$

Matrix differentiation

- Function is a vector and the variable is a vector

$$\mathbf{x} = [x_1, x_2, ..., x_n]^T, \mathbf{y} = [y_1(\mathbf{x}), y_2(\mathbf{x}), ..., y_m(\mathbf{x})]^T$$

In a similar way,

$$\frac{d\mathbf{y}^T}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial y_1(\mathbf{x})}{\partial x_1}, \frac{\partial y_2(\mathbf{x})}{\partial x_1}, ..., \frac{\partial y_m(\mathbf{x})}{\partial x_1} \\ \frac{\partial y_1(\mathbf{x})}{\partial x_2}, \frac{\partial y_2(\mathbf{x})}{\partial x_2}, ..., \frac{\partial y_m(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial y_1(\mathbf{x})}{\partial x_n}, \frac{\partial y_2(\mathbf{x})}{\partial x_n}, ..., \frac{\partial y_m(\mathbf{x})}{\partial x_n} \end{bmatrix}_{n \times m}$$

Matrix differentiation

- Function is a vector and the variable is a vector

Example:

$$\mathbf{y} = \begin{bmatrix} y_1(\mathbf{x}) \\ y_2(\mathbf{x}) \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, y_1(\mathbf{x}) = x_1^2 - x_2, y_2(\mathbf{x}) = x_3^2 + 3x_2$$

$$\frac{d\mathbf{y}^T}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial y_1(\mathbf{x})}{\partial x_1} & \frac{\partial y_2(\mathbf{x})}{\partial x_1} \\ \frac{\partial y_1(\mathbf{x})}{\partial x_2} & \frac{\partial y_2(\mathbf{x})}{\partial x_2} \\ \frac{\partial y_1(\mathbf{x})}{\partial x_3} & \frac{\partial y_2(\mathbf{x})}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 & 0 \\ -1 & 3 \\ 0 & 2x_3 \end{bmatrix}$$

Matrix differentiation

- Function is a scalar and the variable is a matrix

$$f(\mathbf{X}), \mathbf{X} \in \mathbb{R}^{m \times n}$$

Definition

$$\frac{df(\mathbf{X})}{d\mathbf{X}} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} \frac{\partial f}{\partial x_{12}} \cdots \frac{\partial f}{\partial x_{1n}} \\ \cdots \\ \frac{\partial f}{\partial x_{m1}} \frac{\partial f}{\partial x_{m2}} \cdots \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

Matrix differentiation

- Useful results

(1) $\mathbf{x}, \mathbf{a} \in \mathbb{R}^{n\times 1}$

Then,

$$\frac{d\mathbf{a}^T\mathbf{x}}{d\mathbf{x}} = \mathbf{a}, \frac{d\mathbf{x}^T\mathbf{a}}{d\mathbf{x}} = \mathbf{a}$$

How to prove?

Matrix differentiation

- Useful results

(2) $\mathbf{x} \in \mathbb{R}^{n\times 1}$ Then, $\dfrac{d\mathbf{x}^T\mathbf{x}}{d\mathbf{x}} = 2\mathbf{x}$

(3) $\mathbf{y}(\mathbf{x}) \in \mathbb{R}^{m\times 1}$, $\mathbf{x} \in \mathbb{R}^{n\times 1}$, $\dfrac{d\mathbf{y}^T(\mathbf{x})}{d\mathbf{x}} = \left(\dfrac{d\mathbf{y}(\mathbf{x})}{d\mathbf{x}^T}\right)^T$

(4) $A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^{n\times 1}$ Then, $\dfrac{dA\mathbf{x}}{d\mathbf{x}^T} = A$

(5) $A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^{n\times 1}$ Then, $\dfrac{d\mathbf{x}^T A^T}{d\mathbf{x}} = A^T$

(6) $A \in \mathbb{R}^{n\times n}, \mathbf{x} \in \mathbb{R}^{n\times 1}$ Then, $\dfrac{d\mathbf{x}^T A\mathbf{x}}{d\mathbf{x}} = (A + A^T)\mathbf{x}$

(7) $\mathbf{X} \in \mathbb{R}^{m\times n}, \mathbf{a} \in \mathbb{R}^{m\times 1}, \mathbf{b} \in \mathbb{R}^{n\times 1}$ Then, $\dfrac{d\mathbf{a}^T\mathbf{X}\mathbf{b}}{d\mathbf{X}} = \mathbf{a}\mathbf{b}^T$

Matrix differentiation

- Useful results

(8) $\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{a} \in \mathbb{R}^{m \times 1}, \mathbf{b} \in \mathbb{R}^{n \times 1}$ Then, $\dfrac{d\mathbf{a}^T \mathbf{X}^T \mathbf{b}}{d\mathbf{X}} = \mathbf{b}\mathbf{a}^T$

(9) $\mathbf{X} \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times m}$ Then, $\dfrac{d\left(tr\mathbf{X}B\right)}{d\mathbf{X}} = B^T$

(10) $\mathbf{X} \in \mathbb{R}^{n \times n}, \mathbf{X}$ is invertible, $\dfrac{d\left|\mathbf{X}\right|}{d\mathbf{X}} = \left|\mathbf{X}\right|\left(\mathbf{X}^{-1}\right)^T$

Outline

- Matrix Differentiation
- Lagrange Multiplier
- Least-squares for Homogeneous Linear Systems
- Least-squares for Inhomogeneous Linear Systems
    - An SVD-based approach

Lagrange multiplier

- Single-variable function

$f(x)$ is differentiable in (a, b). At $x_0 \in (a, b)$, $f(x)$ achieves an extremum

$$\Rightarrow \frac{df}{dx}\Big|_{x_0} = 0$$

- Two-variables function

$f(x, y)$ is differentiable in its domain. At (x_0, y_0), $f(x, y)$ achieves an extremum

$$\Rightarrow \frac{\partial f}{\partial x}\Big|_{(x_0, y_0)} = 0, \frac{\partial f}{\partial y}\Big|_{(x_0, y_0)} = 0$$

Lagrange multiplier

- In general case

If $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^{n\times 1}$ achieves a local extremum at $\mathbf{x}_0$ and it is derivable at $\mathbf{x}_0$, then $\mathbf{x}_0$ is a stationary point of $f(\mathbf{x})$, i.g.,

$$\frac{\partial f}{\partial x_1}|_{\mathbf{x}_0} = 0, \frac{\partial f}{\partial x_2}|_{\mathbf{x}_0} = 0, \ldots, \frac{\partial f}{\partial x_n}|_{\mathbf{x}_0} = 0$$

Or in other words,

$$\nabla f(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{0}$$

Lagrange multiplier

- Lagrange multiplier is a strategy for finding **all the possible** extremum points of a function subject to equality constraints

Problem: find **all the possible** extremum points for $y = f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^{n\times 1}$

under m constraints $g_k(\mathbf{x}) = 0, k = 1, 2, \ldots, m$

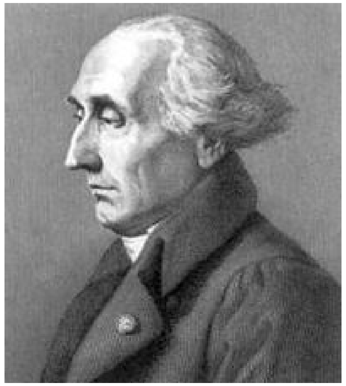
Solution: $F(\mathbf{x}; \lambda_1, \ldots, \lambda_m) = f(\mathbf{x}) + \sum_{k=1}^{m} \lambda_k g_k(\mathbf{x})$

If $\mathbf{x}_0$ is an extremum point of $f(\mathbf{x})$ under constraints

$\Downarrow$

$\exists \lambda_{10}, \lambda_{20} \ldots, \lambda_{m0},$ making $(\mathbf{x}_0, \lambda_{10}, \lambda_{20} \ldots, \lambda_{m0})$ a stationary point of F

Thus, by identifying the stationary points of F, we can get all the possible extremum points of $f(\mathbf{x})$ under equality constraints



Joseph-Louis Lagrange
Jan. 25, 1736~Apr.10, 1813

Lagrange multiplier

- Lagrange multiplier is a strategy for finding **all the possible** extremum points of a function subject to equality constraints

Problem: find **all the possible** extremum points for $y = f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^{n\times 1}$

under m constraints $g_k(\mathbf{x}) = 0, k = 1, 2, \ldots, m$

Solution: $F(\mathbf{x}; \lambda_1, \ldots, \lambda_m) = f(\mathbf{x}) + \sum_{k=1}^{m} \lambda_k g_k(\mathbf{x})$

$(\mathbf{x}_0, \lambda_{10}, \ldots, \lambda_{m0})$ is a stationary point of F $\rightarrow$

$$\underbrace{\frac{\partial F}{\partial x_1} = 0, \frac{\partial F}{\partial x_2} = 0, \ldots, \frac{\partial F}{\partial x_n} = 0, \frac{\partial F}{\partial \lambda_1} = 0, \frac{\partial F}{\partial \lambda_2} = 0, \ldots, \frac{\partial F}{\partial \lambda_m} = 0}_{n + m \text{ equations!}}$$

at that point

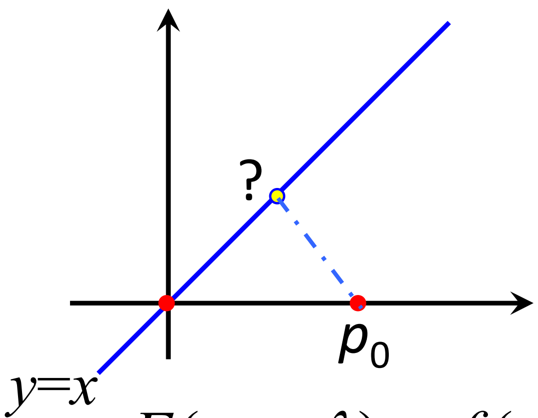
$\downarrow$

$\mathbf{x}_0$ is a possible extremum point of $f(\mathbf{x})$ under equality constraints

Lagrange multiplier

- Example

Problem: for a given point p_0 = (1, 0), among all the points lying on the line $y=x$, identify the one having the least distance to p_0.



The distance is

$$f(x,y)=(x-1)^2+(y-0)^2$$

Now we want to find the global minimizer of $f(x, y)$ under the constraint

$$g(x,y)=y-x=0$$

According to Lagrange multiplier method, construct the Lagrange function

$$F(x,y,\lambda)=f(x,y)+\lambda g(x,y)=(x-1)^2+y^2+\lambda(y-x)$$

Find the stationary point of $F(x,y,\lambda)$

Lagrange multiplier

- Example $\begin{cases} \dfrac{\partial F}{\partial x} = 0 \\ \dfrac{\partial F}{\partial y} = 0 \\ \dfrac{\partial F}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} 2(x-1)+\lambda = 0 \\ 2y-\lambda = 0 \\ x-y=0 \end{cases} \Rightarrow \begin{cases} x = 0.5 \\ y = 0.5 \\ \lambda = 1 \end{cases}$

Thus, $(0.5, 0.5, 1)$ is the only stationary point of $F(x, y, \lambda)$

$(0.5,0.5)$ is the only possible extremum point of $f(x,y)$ under constraints

The global minimizer of $f(x,y)$ under constraints exists

$(0.5,0.5)$ is the global minimizer of $f(x,y)$ under constraints

Outline

- Matrix Differentiation
- Lagrange Multiplier
- Least-squares for Homogeneous Linear Systems
- Least-squares for Inhomogeneous Linear Systems
  - An SVD-based approach

Least-squares for homogeneous linear system

Let's consider a system of m linear equations with n unknowns

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0 \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0 \end{cases} \Leftrightarrow A_{m\times n}\boxed{\mathbf{x}_{n\times 1}} = \mathbf{0}$$

unknowns

We consider the case: m>=n, and rank(A)=n

Theoretically, there is only a trivial solution: $\mathbf{x}$ = 0

We can add a constraint $\|\mathbf{x}\|_2 = 1$ to avoid the trivial solution

Least-squares for homogeneous linear system

We want to minimize $E(\mathbf{x}) = \|A\mathbf{x}\|_2^2$, subject to $\|\mathbf{x}\|_2 = 1$

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} E(\mathbf{x}),\ s.t.,\ \|\mathbf{x}\|_2 = 1 \qquad (1)$$

Construct the Lagrange function,

$$L(\mathbf{x}, \lambda) = \|A\mathbf{x}\|_2^2 + \lambda\left(1 - \|\mathbf{x}\|_2^2\right) \qquad (2)$$

Solving the stationary point $(\mathbf{x}_0, \lambda_0)$ of $L(\mathbf{x}, \lambda)$,

$$\begin{cases} \dfrac{\partial\left[\|A\mathbf{x}\|_2^2 + \lambda\left(1 - \|\mathbf{x}\|_2^2\right)\right]}{\partial \mathbf{x}} = \mathbf{0} \\ \dfrac{\partial\left[\|A\mathbf{x}\|_2^2 + \lambda\left(1 - \|\mathbf{x}\|_2^2\right)\right]}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} A^T A\mathbf{x}_0 = \lambda_0 \mathbf{x}_0 \\ \mathbf{x}_0^T \mathbf{x}_0 = 1 \end{cases}$$

Note: the stationary point of $L(\mathbf{x}, \lambda)$ is not unique

Least-squares for homogeneous linear system

Suppose that $(\mathbf{x}_i, \lambda_i)$ is a stationary point of L, then $\mathbf{x}_i$ is a possible extremum point of $E(\mathbf{x})$ under the equality constraint and we have

$$E(\mathbf{x}_i) = \|A\mathbf{x}_i\|_2^2 = \mathbf{x}_i^T A^T A\mathbf{x}_i = \mathbf{x}_i^T \lambda_i \mathbf{x}_i = \lambda_i$$

The global minimum of $E(\mathbf{x})$ is $\min\{\lambda_i\}$ and the global minimizer of $E(\mathbf{x})$ is the unit eigen-vector of A^TA associated with its least eigen value

Outline

- Matrix Differentiation
- Lagrange Multiplier
- Least-squares for Homogeneous Linear Systems
- Least-squares for Inhomogeneous Linear Systems
    - An SVD-based approach

Least-squares for inhomogeneous linear system

Consider the following linear equations system

$$\begin{cases} x_1 + x_2 = 3 \\ 2x_1 + x_2 = 4 \end{cases} \Leftrightarrow \underbrace{\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 3 \\ 4 \end{bmatrix}}_{\mathbf{b}}$$

Matrix form: $A\mathbf{x} = \mathbf{b}$

It can be easily solved $\begin{cases} x_1 = 1 \\ x_2 = 2 \end{cases}$

Least-squares for inhomogeneous linear system

How about the following one?

$$\begin{cases} x_1 + x_2 = 3 \\ 2x_1 + x_2 = 4 \\ x_1 + 2x_2 = 6 \end{cases} \Leftrightarrow \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix}$$

It does not have a solution!

What is the condition for a linear equation system $A\mathbf{x} = \mathbf{b}$ can be solved?

Carl Friedrich Gauss

Can we solve it in an approximate way?

A: we can use least squares technique!

Least-squares for inhomogeneous linear system

Let's consider a system of m linear equations with n unknowns

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases} \Leftrightarrow A_{m\times n}\underbrace{\mathbf{x}_{n\times 1}}_{\text{unknowns}} = \mathbf{b}_{m\times 1} \left(\mathbf{b} \neq \mathbf{0}\right)$$

We consider the case: rank(A)=n, and rank([A; $\mathbf{b}$])=$n+1$

Theoretically, there is no solution!

Instead, we want to find a vector $\mathbf{x}$ that minimizes the error:

$$E(\mathbf{x}) \equiv \sum_{i=1}^{m} (a_{i1}x_1 + \ldots + a_{in}x_n - b_i)^2 = \|A\mathbf{x} - \mathbf{b}\|_2^2$$

LS for Inhomogeneous Linear System

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} E(\mathbf{x}) = \arg\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2$$

The stationary point of $E(\mathbf{x})$ is $\mathbf{x}_s = \left(A^T A\right)^{-1} A^T \mathbf{b}$

Since $E(\mathbf{x})$ is a **convex** function, its stationary point is the global minimizer[1]

$$\mathbf{x}^* = \mathbf{x}_s = \boxed{\left(A^T A\right)^{-1} A^T} \mathbf{b}$$

Pseudoinverse of A

[1] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004, pp. 69

An SVD-based approach

SVD decomposition theorem: Any matrix $A_{m\times n}$ can be decomposed as the following form,

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}^{T}$$

where U and V are two orthogonal matrices, $r(A)=r$,

$$\Sigma_{m\times n} = \begin{bmatrix} \Sigma_r & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n} = \begin{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n}$$

$\sigma_1, \sigma_2, \ldots, \sigma_r > 0$ are called the **singular values** of A

In general case, $\Sigma_{m\times n}$ is not unique. However, if $\{\sigma_i\}_{i=1}^{r}$ are arranged in order, $\Sigma_{m\times n}$ is uniquely determined by A. In the following, we require that $\sigma_1 \geq \sigma_2 \geq, \ldots, \geq \sigma_r > 0$

An SVD-based approach

Linear least squares is a general idea for solving inhomogeneous linear equations,

$$A_{m\times n}\mathbf{x}_{n\times 1} = \mathbf{b}_{m\times 1}, \mathbf{b} \neq \mathbf{0} \quad (1)$$

Using the idea of least squares, Eq. 1 is equivalent to the following problem,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left\| A_{m\times n}\mathbf{x}_{n\times 1} - \mathbf{b}_{m\times 1} \right\|_2^2 \quad (2)$$

Eq. 2 can be solved by finding the stationary point $\mathbf{x}^*$ of $\left\| A_{m\times n}\mathbf{x}_{n\times 1} - \mathbf{b}_{m\times 1} \right\|_2^2$, i.e. $\mathbf{x}^*$ should satisfy,

$$A^T A\mathbf{x}^* = A^T\mathbf{b} \quad (3)$$

In Eq. 3, when $rank(A) = n$ (the columns of A are linearly independent) ,

$rank(A^T A) = n$ ➡ $A^T A$ is invertible ➡ $\mathbf{x}^*$ is uniquely determined as $\mathbf{x}^* = (A^T A)^{-1} A^T\mathbf{b}$

How about when rank(A)<n?

An SVD-based approach

- For solving the linear least squares numerically with a computer, usually we do not use the form of Eq. (3) (though it is elegant) for two reasons
  - When rank(A)<n, $\mathbf{x}^*$ can not be determined
  - Even though A^TA is invertible, the formation of A^TA can dramatically degrade the accuracy of the computation
- Instead, we can use the technique of SVD

An SVD-based approach

Suppose the SVD form of A is,

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T$$

$$\Downarrow$$

$$A\mathbf{x}-\mathbf{b} = U\Sigma V^T\mathbf{x}-\mathbf{b} = U\left(\Sigma V^T\mathbf{x}\right)-U\left(U^T\mathbf{b}\right) \triangleq U\left(\Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}\right)$$

where $\mathbf{y}_{n\times 1} = V^T\mathbf{x}, \mathbf{c}_{m\times 1} = U^T\mathbf{b}$

Since U is an orthogonal matrix,

$$\|A\mathbf{x}-\mathbf{b}\| = \left\|U\left(\Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}\right)\right\| = \left\|\Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}\right\|$$

Then, our objective is to identify $\mathbf{y}$ that can make $\left\|\Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}\right\|$ have minimum length

An SVD-based approach

$$\Sigma\mathbf{y}_{n\times 1}=\begin{bmatrix}\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)}\end{bmatrix}_{m\times n}\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}=\begin{bmatrix}\sigma_1 y_1\\ \sigma_2 y_2\\ \vdots\\ \sigma_r y_r\\ 0\\ \vdots\\ 0\end{bmatrix}_{m\times 1} \Rightarrow \Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}=\begin{bmatrix}\sigma_1 y_1-c_1\\ \sigma_2 y_2-c_2\\ \vdots\\ \sigma_r y_r-c_r\\ -c_{r+1}\\ \vdots\\ -c_m\end{bmatrix}_{m\times 1}$$

Then, we simply let $y_i=\frac{c_i}{\sigma_i}, 1\le i\le r$; then, $\left\|\Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}\right\|$ can get the minimum length $\sqrt{\sum_{i=r+1}^{m}c_i^2}$

Note that y_{r+1}~y_n can be arbitrary

An SVD-based approach

The operation $y_i = \dfrac{c_i}{\sigma_i}, 1 \le i \le r$ can be simply completed by a matrix multiplication,

$$\mathbf{y} = \begin{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_r} \end{bmatrix} & \boldsymbol{O}_{r\times(m-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(m-r)} \end{bmatrix}_{n\times m} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}_{m\times 1} = \begin{bmatrix} c_1/\sigma_1 \\ c_2/\sigma_2 \\ \vdots \\ c_r/\sigma_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1} \triangleq \Sigma^{+}\mathbf{c}_{m\times 1}$$

where Σ^{+} means transposing Σ and inverting all non-zero diagonal entries

Finally,

$$\mathbf{x} = V\mathbf{y}_{n\times 1} = V\Sigma^{+}\mathbf{c}_{m\times 1} = V\Sigma^{+}U^{T}\mathbf{b}$$

Moore-Penrose inverse

An SVD-based approach

- Some notes about the generalized inverse used in linear least squares
  - It does not have requirements for the rank of A
  - It can guarantee that the obtained solution can make $\|A\mathbf{x}-\mathbf{b}\|$ having the minimum length; but **the solution may be not unique**


Q+A