Định nghĩa 1. Hàm số $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ được gọi là **lồi** (convex) nếu với mọi $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ và $\theta \in \mathbb{R}$ sao cho $0 \leq \theta \leq 1$, ta có

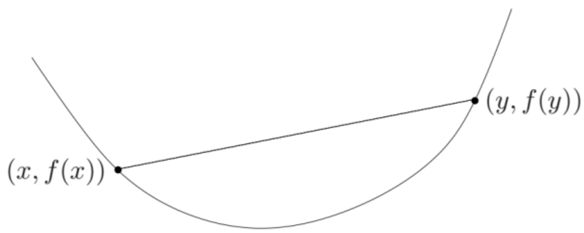
$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) \quad (*).$$

Tương tự, ta nói f là:

- **lồi ngặt** (strictly convex) nếu dấu $\leq$ trong $(*)$ được thay bằng dấu $<$,
- **lõm** (concave) nếu dấu $\leq$ trong $(*)$ được thay bằng dấu $\geq$,
- **lõm ngặt** (strictly concave) nếu dấu $\leq$ trong $(*)$ được thay bằng dấu $>$.

Nhận xét 2.

- Nếu f là hàm lồi (lõm) thì $-f$ là hàm lõm (lồi).
- Về mặt hình học, $(*)$ có nghĩa là đoạn thẳng nối $(\boldsymbol{x}, f(\boldsymbol{x}))$ và $(\boldsymbol{y}, f(\boldsymbol{y}))$ luôn nằm trên đồ thị của hàm số f.



Ví dụ 3. Hàm affine vừa lồi vừa lõm.

Mệnh đề 4. Cho $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ có đạo hàm liên tục đến cấp 2, các phát biểu sau là tương đương:

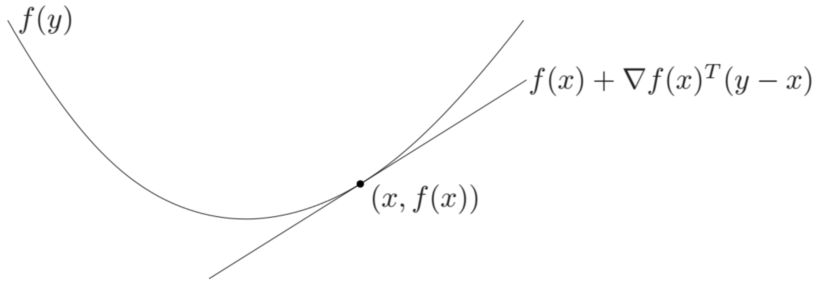
(i) f là hàm lồi.
(ii) $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x})$, với mọi $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.
(iii) $\nabla^2 f(\boldsymbol{x}) \geq 0$ ($\nabla^2 f(\boldsymbol{x})$ là ma trận nửa xác định dương), với mọi $\boldsymbol{x} \in \mathbb{R}^n$.

Tương tự, các phát biểu sau là tương đương:

(i) f là hàm lõm.
(ii) $f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x})$, với mọi $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.
(iii) $\nabla^2 f(\boldsymbol{x}) \leq 0$ ($\nabla^2 f(\boldsymbol{x})$ là ma trận nửa xác định âm), với mọi $\boldsymbol{x} \in \mathbb{R}^n$.

Nhận xét 5.

- Ta đã biết $f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x})$ là xấp xỉ Taylor bậc nhất của f tại $\boldsymbol{x}$. Do đó (ii) có nghĩa là xấp xỉ Taylor bậc nhất luôn là **ước lượng dưới** (underestimator) của hàm f.
- Cũng từ (ii) nếu $\nabla f(\boldsymbol{x})^T = \boldsymbol{0}$ thì $f(\boldsymbol{y}) \geq f(\boldsymbol{x})$ với mọi $\boldsymbol{y} \in \mathbb{R}^n$ nên $\boldsymbol{x}$ là một **điểm cực tiểu toàn cục** (global minimum point) mà tại đó hàm f nhận **giá trị nhỏ nhất** (minimum value).



- (iii) cho phép ta kiểm tra tính lồi (lõm) của hàm số bằng các kĩ thuật của đại số tuyến tính.

Ví dụ 6. Cho $A \in \mathbb{R}^{n\times n}$ là ma trận đối xứng ($A^T = A$), $\boldsymbol{b} \in \mathbb{R}^n, c \in \mathbb{R}$, hàm số $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ được xác định bởi

$$f(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$$

được gọi là **hàm bậc hai** (quadratic function).

Vì

$$\nabla f(\boldsymbol{x}) = 2A\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) = 2A$$

Nên ta có

- f lồi khi và chỉ khi A nửa xác định dương ($A \geq 0$),
- f lồi ngặt khi và chỉ khi A xác định dương ($A > 0$),
- f lõm khi và chỉ khi A nửa xác định âm ($A \leq 0$),
- f lõm ngặt khi và chỉ khi A xác định âm ($A < 0$).

Ví dụ 7. Hàm số $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ được xác định bởi

$$f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$$

là hàm lồi ngặt vì ta có

$$f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{x}^T I_n \boldsymbol{x}$$

với I_n là ma trận đơn vị cấp n nên $\nabla^2 f(\boldsymbol{x}) = 2I_n > 0$. Điểm cực tiểu duy nhất của hàm số chính là nghiệm duy nhất của hệ PTTT

$$\nabla f(\boldsymbol{x}) = 2I_n\boldsymbol{x} = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}.$$

Ví dụ 8. Khảo sát tính lồi/lõm và tìm các điểm cực trị toàn cục (nếu có) của hàm số 3 biến sau

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - x_1x_2 + x_1x_3 + 2x_1 - 6x_3 + 5.$$

Định nghĩa 9. Cho $A \in \mathbb{R}^{m\times n}$ và $\boldsymbol{b} \in \mathbb{R}^m$, bài toán tìm $\widehat{\boldsymbol{x}} \in \mathbb{R}^n$ để

$$\text{minimize } \|A\boldsymbol{x} - \boldsymbol{b}\|^2$$

được gọi là **bài toán bình phương tối tiểu** (least squares problem). Vector $\boldsymbol{r} = A\boldsymbol{x} - \boldsymbol{b} \in \mathbb{R}^m$ được gọi là **phần dư** (residual) và $\widehat{\boldsymbol{x}}$ được gọi là nghiệm của bài toán.

Nhận xét 10.

- $\|A\boldsymbol{x}-\boldsymbol{b}\|^2 = \|\boldsymbol{r}\|^2 = r_1^2 + r_2^2 + \cdots + r_m^2 = (a_1^T\boldsymbol{x} - b_1)^2 + (a_2^T\boldsymbol{x} - b_2)^2 + \cdots + (a_m^T\boldsymbol{x} - b_m)^2$, với $a_i^T \in \mathbb{R}^n$ là dòng thứ i của ma trận A.
- $A\boldsymbol{x} = x_1a_1 + x_2a_2 + \cdots + x_na_n$, với $a_j \in \mathbb{R}^m$ là cột thứ j của ma trận A ($A\boldsymbol{x}$ là một tổ hợp tuyến tính của các cột của A).
- Khi $\boldsymbol{b}$ là một tổ hợp tuyến tính của các cột của A thì $\widehat{\boldsymbol{x}}$ là nghiệm của hệ PTTT $A\boldsymbol{x} = \boldsymbol{b}$, khi đó $\widehat{\boldsymbol{r}} = A\widehat{\boldsymbol{x}} - \boldsymbol{b} = \boldsymbol{0}$ và $\|A\widehat{\boldsymbol{x}} - \boldsymbol{b}\|^2 = 0$.
- Khi $\boldsymbol{b}$ không là một tổ hợp tuyến tính của các cột của A thì hệ PTTT $A\boldsymbol{x} = \boldsymbol{b}$ vô nghiệm, khi đó $\|A\widehat{\boldsymbol{x}} - \boldsymbol{b}\|^2 > 0$ nhưng $A\widehat{\boldsymbol{x}}$ là "xấp xỉ tốt nhất" của $\boldsymbol{b}$ ($A\widehat{\boldsymbol{x}} \approx \boldsymbol{b}$) theo nghĩa phần dư $\widehat{\boldsymbol{r}} = A\widehat{\boldsymbol{x}} - \boldsymbol{b}$ có chuẩn nhỏ nhất.

Ví dụ 11. Cho

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Hệ PTTT $A\boldsymbol{x} = \boldsymbol{b}$ vô nghiệm vì $\boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ không là tổ hợp tuyến tính của $\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$.

Bài toán bình phương tối tiểu

$$\text{minimize } \|A\boldsymbol{x} - \boldsymbol{b}\|^2 \text{ là minimize } (2x_1 - 1)^2 + (-x_1 + x_2)^2 + (2x_2 + 1)^2$$

có nghiệm là (tìm bằng công thức bên dưới) $\widehat{\boldsymbol{x}} = \left(\frac{1}{3}, -\frac{1}{3}\right)$ với phần dư $\widehat{\boldsymbol{r}} = A\widehat{\boldsymbol{x}} - \boldsymbol{b} = \left(-\frac{1}{3}, -\frac{2}{3}, \frac{1}{3}\right)$ và $\|\widehat{\boldsymbol{r}}\|^2 = \frac{2}{3}$.

Không có trường hợp $\boldsymbol{x}$ nào cho phần dư có chuẩn nhỏ hơn $\|\widehat{\boldsymbol{r}}\| = \sqrt{\frac{2}{3}}$. Chẳng hạn, với $\widetilde{\boldsymbol{x}} = \left(\frac{1}{2}, -\frac{1}{2}\right)$ ta có $\widetilde{\boldsymbol{r}} = A\widetilde{\boldsymbol{r}} - \boldsymbol{b} = (0, -1, 0)$ và $\|\widetilde{\boldsymbol{r}}\| = 1$.

Mệnh đề 12. Nghiệm $\widehat{\boldsymbol{x}}$ của bài toán bình phương tối tiểu

$$\text{minimize } \|A\boldsymbol{x} - \boldsymbol{b}\|^2$$

thỏa

$$A^TA\widehat{\boldsymbol{x}} = A^T\boldsymbol{b}.$$

Trường hợp các cột của A độc lập tuyến tính (A có hạng là n) thì A^TA khả nghịch nên

$$\widehat{\boldsymbol{x}} = (A^TA)^{-1}A^T\boldsymbol{b} = A^\dagger\boldsymbol{b}.$$

là nghiệm duy nhất của bài toán.

Ma trận $A^\dagger = (A^TA)^{-1}A^T$ được gọi là ma trận **giả nghịch đảo** (pseudo-inverse) của A. Nhận xét, khi A là ma trận vuông khả nghịch thì $A^\dagger = (A^TA)^{-1}A^T = A^{-1}(A^T)^{-1}A^T = A^{-1}$ và $\widehat{\boldsymbol{x}} = A^\dagger\boldsymbol{b} = A^{-1}\boldsymbol{b}$ là nghiệm duy nhất của hệ PTTT $A\boldsymbol{x} = \boldsymbol{b}$.

Chứng minh 13. Với $A \in \mathbb{R}^{m\times n}$ và $\boldsymbol{b} \in \mathbb{R}^m$, xét hàm số $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ được xác định bởi

$$f(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2.$$

Ta có

$$\begin{aligned} f(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 &= (A\boldsymbol{x} - \boldsymbol{b})^T(A\boldsymbol{x} - \boldsymbol{b}) = (\boldsymbol{x}^T A^T - \boldsymbol{b}^T)(A\boldsymbol{x} - \boldsymbol{b}) \\ &= \boldsymbol{x}^T A^T A\boldsymbol{x} - \boldsymbol{x}^T A^T \boldsymbol{b} - \boldsymbol{b}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{b} = \boldsymbol{x}^T(A^T A)\boldsymbol{x} - (2\boldsymbol{b}^T A)\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{b}. \end{aligned}$$

Do đó

$$\nabla f(\boldsymbol{x}) = 2(A^T A)\boldsymbol{x} - (2\boldsymbol{b}^T A)^T = 2(A^T A)\boldsymbol{x} - 2A^T\boldsymbol{b} \text{ và } \nabla^2 f(\boldsymbol{x}) = 2(A^T A).$$

Do $\boldsymbol{h}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{h} = 2(\boldsymbol{h}^T(A^T A)\boldsymbol{h}) = 2\big((\boldsymbol{h}^T A^T)(A\boldsymbol{h})\big) = 2\big((A\boldsymbol{h})^T(A\boldsymbol{h})\big) = 2\|A\boldsymbol{h}\|^2 \geq 0$ với mọi $\boldsymbol{h} \in \mathbb{R}^n$ nên $\nabla^2 f(\boldsymbol{x})$ là ma trận nửa xác định dương với mọi $\boldsymbol{x} \in \mathbb{R}^n$. Như vậy f là hàm lồi nên đạt giá trị nhỏ nhất tại $\widehat{\boldsymbol{x}}$ là nghiệm của hệ PTTT

$$\nabla f(\boldsymbol{x}) = 2(A^T A)\boldsymbol{x} - 2A^T\boldsymbol{b} = \boldsymbol{0} \Leftrightarrow A^T A\widehat{\boldsymbol{x}} = A^T\boldsymbol{b}.$$

Ví dụ 14 (tiếp Ví dụ 11). Với

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

Ta có

$$A^T A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}, (A^T A)^{-1} = \begin{bmatrix} \frac{5}{24} & \frac{1}{24} \\ \frac{1}{24} & \frac{5}{24} \end{bmatrix},$$

$$A^\dagger = (A^T A)^{-1}A^T = \begin{bmatrix} \frac{5}{24} & \frac{1}{24} \\ \frac{1}{24} & \frac{5}{24} \end{bmatrix}\begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} \frac{5}{12} & -\frac{1}{6} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{6} & \frac{5}{12} \end{bmatrix},$$

$$\widehat{\boldsymbol{x}} = A^\dagger\boldsymbol{b} = \begin{bmatrix} \frac{5}{12} & -\frac{1}{6} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{6} & \frac{5}{12} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}.$$

Ứng dụng 15 (polynomial fitting). Trong mặt phẳng cho 5 điểm có tọa độ

Điểm i	x_i	y_i
1	-2	9
2	-1	-1
3	0	1
4	1	3
5	2	17

Tìm đa thức có bậc không quá k "khớp" các điểm trên.

Giải: (xem file Notebook)

Ta tìm các hệ số a_i của đa thức

$$\hat{y}(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k$$

để $\hat{y}(x)$ “khớp tốt nhất” $n = 5$ điểm trên, nghĩa là

$$\hat{y}(x_i) = a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_k x_i^k \approx y_i \text{ tức là } \hat{r_i} = \hat{y}(x_i) - y_i \approx 0, i = 1, 2, \dots, n.$$

Đặt

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^k \end{bmatrix} \in \mathbb{R}^{n \times (k+1)}, \qquad \boldsymbol{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} \in \mathbb{R}^{k+1}, \qquad \boldsymbol{r} = A\boldsymbol{a} - \boldsymbol{y} \in \mathbb{R}^n.$$

Ta có bài toán bình phương tối tiểu, cụ thể, ta cần tìm vector hệ số $\boldsymbol{a}$ sao cho tổng bình phương “lỗi” sau là nhỏ nhất

$$\sum_{i=1}^{n} (\hat{y}(x_i) - y_i)^2 = \sum_{i=1}^{n} r_i^2 = \|\boldsymbol{r}\|^2 = \|A\boldsymbol{a} - \boldsymbol{y}\|^2$$

Khi các hoành độ $x_i (i = 1, 2, \dots, n)$ phân biệt, ta có các cột của ma trận A là độc lập tuyến tính, do đó vector các hệ số cần tìm là

$$\hat{\boldsymbol{a}} = (A^T A)^{-1} A^T \boldsymbol{y} = A^\dagger \boldsymbol{y}.$$

<u>$k = 0$</u>:

$$A = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad (A^T A) = n, \qquad (A^T A)^{-1} = \frac{1}{n}, \qquad A^\dagger = (A^T A)^{-1} A^T = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix},$$

$$\hat{\boldsymbol{a}} = A^\dagger \boldsymbol{y} = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} \begin{bmatrix} y_1 \\ y_1 \\ \vdots \\ a_n \end{bmatrix} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} = 5.8$$

Đa thức bậc không (hằng số) khớp tốt nhất (tổng bình phương lỗi tối tiểu) là

$$\hat{y}(x) = \bar{y} = 5.8.$$

<u>$k = 1$</u>:

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad (A^T A) = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}, \qquad (A^T A)^{-1} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix},$$

$$\hat{\boldsymbol{a}} = A^\dagger \boldsymbol{y} = \begin{bmatrix} 5.8 \\ 2 \end{bmatrix}$$

Đa thức bậc nhất (đường thẳng) khớp tốt nhất là

$$\hat{y}(x) = 5.8 + 2x.$$

$\underline{k = 2}$:

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} 9 \\ -1 \\ 1 \\ 3 \\ 17 \end{bmatrix}$$

Đa thức bậc hai (parabol) khớp tốt nhất là

$$\hat{y}(x) = -1.05714286 + 2x + 3.42857143x^2.$$

Các đa thức bậc cao hơn (xem Notebook).

Vì có 5 điểm nên đa thức bậc 4 sau “đi qua” 5 điểm (không lỗi)

$$\hat{y}(x) = 1 + 2x - x^2 + x^4.$$