Các kiến thức cơ sở

Hoàng Nam Dũng

Khoa Toán - Cơ - Tin học, Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

Tập lồi

Định nghĩa

Tập hợp $S \subseteq \mathbb{R}^n$ *là một* tập lồi *nếu*

$$\lambda x + (1-\lambda)y \in S,\ \forall \lambda \in [0,1], x, y \in S.$$

Nói một cách khác đoạn thẳng nối hai điểm hoàn toàn nằm trong tập hợp nếu hai đầu mút cũng thuộc tập hợp.

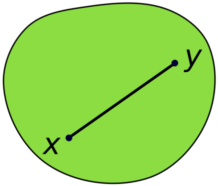
Tập lồi

Định nghĩa

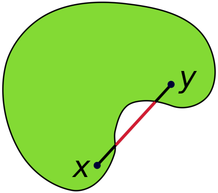
Tập hợp $S \subseteq \mathbb{R}^n$ *là một* tập lồi *nếu*

$$\lambda x + (1-\lambda) y \in S, \ \forall \lambda \in [0,1], x, y \in S.$$

Nói một cách khác đoạn thẳng nối hai điểm hoàn toàn nằm trong tập hợp nếu hai đầu mút cũng thuộc tập hợp.



Tập lồi



Tập không lồi

Ví dụ tập lồi

- Tập rỗng, điểm, đường thẳng, toàn bộ không gian $\mathbb{R}^n$.
- Hình cầu $\{x \in \mathbb{R}^n \mid \|x\| \leq r\}$ với chuẩn $\|\cdot\|$ và bán kín r cho trước.
- Siêu phẳng (hyperplane) $\{x \in \mathbb{R}^n \mid a^T x = b\}$ với $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ cho trước.
- Nửa không gian (halfspace) $\{x \in \mathbb{R}^n \mid a^T x \leq b\}$ với $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ cho trước.
- $\{x \in \mathbb{R}^n \mid Ax = b\}$ với $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ cho trước.
- Đa diện $\{x \in \mathbb{R}^n \mid Ax \leq b\}$ với $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ cho trước.
- ...

Tổ hợp lồi và bao lồi

Định nghĩa

Một tổ hợp lồi *của* $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ *là một tổ hợp tuyến tính*

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k$$

với các hệ số $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$ *thỏa mãn*

$$\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1.$$

Tổ hợp lồi và bao lồi

Định nghĩa

Một tổ hợp lồi *của* $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ *là một tổ hợp tuyến tính*

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k$$

với các hệ số $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$ *thỏa mãn*

$$\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1.$$

Định nghĩa

Bao lồi *của một tập hợp* S, $conv(S)$, *là tập hợp của tất cả các tổ hợp lồi của các phần tử thuộc* S.

Tổ hợp lồi và bao lồi

Định nghĩa

Một tổ hợp lồi *của* $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ *là một tổ hợp tuyến tính*

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k$$

với các hệ số $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$ *thỏa mãn*

$$\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1.$$

Định nghĩa

Bao lồi *của một tập hợp* S, $\mathit{conv}(S)$, *là tập hợp của tất cả các tổ hợp lồi của các phần tử thuộc* S.

$\mathrm{conv}(S)$ là một tập lồi và là tập lồi bé nhất chứa S.

Tổ hợp lồi và bao lồi

Định nghĩa

Một tổ hợp lồi *của* $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ *là một tổ hợp tuyến tính*

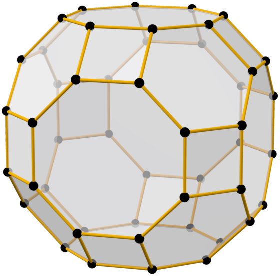
$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k$$

với các hệ số $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$ *thỏa mãn*

$$\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1.$$

Định nghĩa

Bao lồi *của một tập hợp* S, $conv(S)$, *là tập hợp của tất cả các tổ hợp lồi của các phần tử thuộc* S.



$\mathrm{conv}(S)$ là một tập lồi và là tập lồi bé nhất chứa S.

Hàm lồi

Định nghĩa

Một hàm số $f : S \to \mathbb{R}$, $S \subseteq \mathbb{R}^n$, *được gọi là một* hàm lồi *nếu*

- *Miền định nghĩa S là một tập lồi.*
- *Hàm f thỏa mãn*

  $$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \ \forall \lambda \in [0,1], x, y \in S.$$

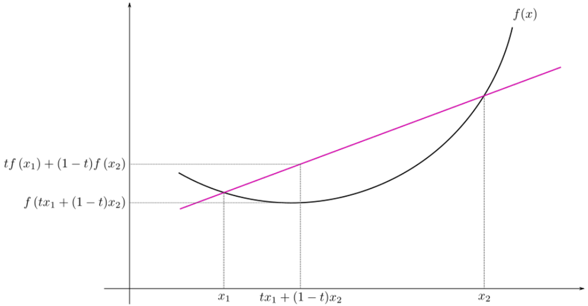
Hàm lồi

Định nghĩa

Một hàm số $f : S \to \mathbb{R}$, $S \subseteq \mathbb{R}^n$, *được gọi là một* hàm lồi *nếu*

- *Miền định nghĩa* S *là một tập lồi.*
- *Hàm* f *thỏa mãn*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \ \forall \lambda \in [0,1], x, y \in S.$$



Hàm lõm

Định nghĩa

Một hàm số $f : S \to \mathbb{R}$, $S \subseteq \mathbb{R}^n$, *được gọi là một* hàm lõm *nếu*

- *Miền định nghĩa* S *là một tập lồi.*
- *Hàm* f *thỏa mãn*

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y), \ \forall \lambda \in [0,1], x, y \in S.$$

$$f \text{ là hàm lõm } \iff -f \text{ là một hàm lồi.}$$

Hàm lồi ngặt và hàm lồi mạnh

Định nghĩa

Một hàm lồi $f : S \to \mathbb{R}$ *được gọi là*

- lồi ngặt *nếu ta có với mọi* $\lambda \in (0, 1), x, y \in S, x \neq y$

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y).$$

  Tức là độ cong của nó lớn hơn độ cong của 1 hàm tuyến tính.

Hàm lồi ngặt và hàm lồi mạnh

Định nghĩa

Một hàm lồi $f : S \to \mathbb{R}$ *được gọi là*

- lồi ngặt *nếu ta có với mọi* $\lambda \in (0,1), x, y \in S, x \neq y$
$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y).$$
Tức là độ cong của nó lớn hơn độ cong của 1 hàm tuyến tính.
- lồi mạnh *nếu tồn tại* $m > 0$ *sao cho* $f - m\|x\|_2^2$ *là một hàm lồi hay tương đương với*
tồn tại $m > 0$ *sao cho ta có với mọi* $\lambda \in [0,1], x, y \in S$
$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)-\frac{1}{2}m\lambda(1-\lambda)\|x-y\|_2^2.$$

Hàm lồi ngặt và hàm lồi mạnh

Định nghĩa

Một hàm lồi $f : S \to \mathbb{R}$ *được gọi là*

- lồi ngặt *nếu ta có với mọi* $\lambda \in (0,1), x, y \in S, x \neq y$

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y).$$

  Tức là độ cong của nó lớn hơn độ cong của 1 hàm tuyến tính.

- lồi mạnh *nếu tồn tại* $m > 0$ *sao cho* $f - m\|x\|_2^2$ *là một hàm lồi hay tương đương với*
  tồn tại $m > 0$ *sao cho ta có với mọi* $\lambda \in [0,1], x, y \in S$

$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)-\frac{1}{2}m\lambda(1-\lambda)\|x-y\|_2^2.$$

Mối quan hệ:

$$\text{Lồi mạnh} \implies \text{lồi ngặt} \implies \text{lồi.}$$

Ví dụ hàm lồi

- Hàm đơn biến
  - e^{ax} trên $\mathbb{R}$ với $a \in \mathbb{R}$
  - x^a trên $\mathbb{R}_{\geq 0}$ với $a \notin (0, 1)$
  - $-x^a$ trên $\mathbb{R}_{\geq 0}$ với $a \in [0, 1]$
  - $-\log x$ trên $\mathbb{R}_{>0}$
- Hàm affine: $a^T x + b$ đồng thời vừa là hàm lồi, vừa là hàm lõm
- Hàm bậc 2: $\frac{1}{2}x^T Ax + b^T x + c$ với A nửa xác định dương ($A \succeq 0$)
- Least square lost: $\|y - Ax\|_2^2$
- Chuẩn $\|x\|$ bất kì, ví dụ chuẩn L_p
- Hàm max: $f(x) = \max\{x_1, x_2, \ldots, x_n\}$
- $\ldots$

Đặc trưng của hàm lồi

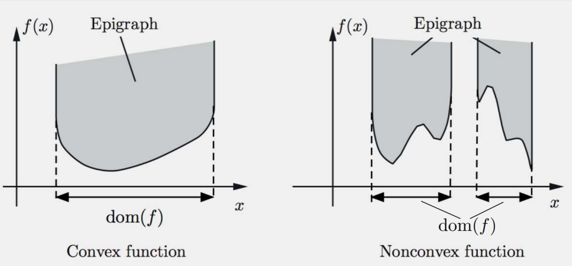
- Epigraph:

$$\mathrm{epi}(f) = \{(x, t) \in \mathrm{dom}(f) \times \mathbb{R} \mid f(x) \leq t\}$$

Đặc trưng của hàm lồi

- Epigraph:

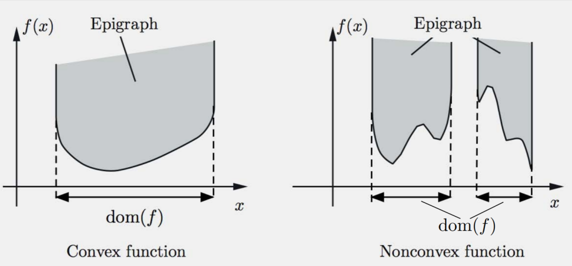
$$\text{epi}(f) = \{(x, t) \in \text{dom}(f) \times \mathbb{R} \mid f(x) \leq t\}$$



Đặc trưng của hàm lồi

- Epigraph:

$$\mathrm{epi}(f) = \{(x, t) \in \mathrm{dom}(f) \times \mathbb{R} \mid f(x) \leq t\}$$

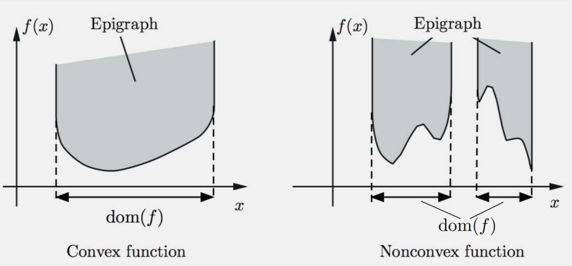


$$f \text{ là hàm lồi} \iff \mathrm{epi}(f) \text{ là một tập lồi.}$$

Đặc trưng của hàm lồi

- Epigraph:

$$\mathrm{epi}(f) = \{(x, t) \in \mathrm{dom}(f) \times \mathbb{R} \mid f(x) \leq t\}$$



$$f \text{ là hàm lồi} \iff \mathrm{epi}(f) \text{ là một tập lồi.}$$

- Tập mức dưới: Nếu f lồi thì tập mức dưới

$$\{x \in \mathrm{dom}(f) \mid f(x) \leq \alpha\}$$

là lồi với mọi $\alpha \in \mathbb{R}$. Tuy nhiên điều ngược lại không đúng.

Đặc trưng của hàm lồi

Định lý (Đặc trưng bậc nhất)

Nếu f khả vi thì f là hàm lồi khi và chỉ khi $dom(f)$ là một tập lồi và

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \ \forall x, y \in dom(f).$$

Đặc trưng của hàm lồi

Định lý (Đặc trưng bậc nhất)

Nếu f khả vi thì f là hàm lồi khi và chỉ khi $dom(f)$ là một tập lồi và

$$f(y) \geq f(x) + \nabla f(x)^T (y - x),\ \forall x, y \in dom(f).$$

Định lý (Đặc trưng bậc hai)

Nếu f khả vi hai lần thì hàm lồi khi và chỉ khi $dom(f)$ là một tập lồi và

$$\nabla^2 f(x) \succeq 0,\ \forall x \in dom(f).$$

Đặc trưng của hàm lồi

Định lý (Đặc trưng bậc nhất)

Nếu f khả vi thì f là hàm lồi khi và chỉ khi $dom(f)$ là một tập lồi và

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \ \forall x, y \in dom(f).$$

Định lý (Đặc trưng bậc hai)

Nếu f khả vi hai lần thì hàm lồi khi và chỉ khi $dom(f)$ là một tập lồi và

$$\nabla^2 f(x) \succeq 0, \ \forall x \in dom(f).$$

Định lý (Bất đẳng thức Jenssen)

Nếu f là một hàm lồi và X là một biến ngẫu nhiên trên dom(f) thì

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

Operations preserving convexity

- Nonnegative linear combination: $f_1, \ldots f_m$ convex implies $a_1 f_1 + \ldots + a_m f_m$ convex for any $a_1, \ldots a_m \geq 0$
- Pointwise maximization: if f_s is convex for any $s \in S$, then $f(x) = \max_{s \in S} \; f_s(x)$ is convex. Note that the set S here (number of functions f_s) can be infinite
- Partial minimization: if $g(x, y)$ is convex in x, y, and C is convex, then $f(x) = \min_{y \in C} \; g(x, y)$ is convex

Example: distances to a set

Let C be an arbitrary set, and consider the maximum distance to C under an arbitrary norm $\|\cdot\|$:

$$f(x) = \max_{y \in C} \|x - y\|$$

Let's check convexity: $f_y(x) = \|x - y\|$ is convex in x for any fixed y, so by pointwise maximization rule, f is convex

Now let C be convex, and consider the minimum distance to C:

$$f(x) = \min_{y \in C} \|x - y\|$$

Let's check convexity: $g(x, y) = \|x - y\|$ is convex in x, y jointly, and C is assumed convex, so apply partial minimization rule

More operations preserving convexity

- Affine composition: if f is convex, then $g(x) = f(Ax + b)$ is convex
- General composition: suppose $f = h \circ g$, where $g : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$. Then:
  - f is convex if h is convex and nondecreasing, g is convex
  - f is convex if h is convex and nonincreasing, g is concave
  - f is concave if h is concave and nondecreasing, g concave
  - f is concave if h is concave and nonincreasing, g convex

  How to remember these? Think of the chain rule when $n = 1$:

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

- Vector composition: suppose that

$$f(x) = h\big(g(x)\big) = h\big(g_1(x), \dots g_k(x)\big)$$

where $g : \mathbb{R}^n \to \mathbb{R}^k$, $h : \mathbb{R}^k \to \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$. Then:

  - f is convex if h is convex and nondecreasing in each argument, g is convex
  - f is convex if h is convex and nonincreasing in each argument, g is concave
  - f is concave if h is concave and nondecreasing in each argument, g is concave
  - f is concave if h is concave and nonincreasing in each argument, g is convex

Example: log-sum-exp function

Log-sum-exp function: $g(x) = \log(\sum_{i=1}^k e^{a_i^T x + b_i})$, for fixed a_i, b_i, $i = 1, \ldots k$. Often called "soft max", as it smoothly approximates $\max_{i=1,\ldots k} \, (a_i^T x + b_i)$

How to show convexity? First, note it suffices to prove convexity of $f(x) = \log(\sum_{i=1}^n e^{x_i})$ (affine composition rule)

Now use second-order characterization. Calculate

$$\begin{aligned}
\nabla_i f(x) &= \frac{e^{x_i}}{\sum_{\ell=1}^n e^{x_\ell}} \\
\nabla^2_{ij} f(x) &= \frac{e^{x_i}}{\sum_{\ell=1}^n e^{x_\ell}} 1\{i = j\} - \frac{e^{x_i} e^{x_j}}{(\sum_{\ell=1}^n e^{x_\ell})^2}
\end{aligned}$$

Write $\nabla^2 f(x) = \mathrm{diag}(z) - zz^T$, where $z_i = e^{x_i} / (\sum_{\ell=1}^n e^{x_\ell})$. This matrix is diagonally dominant, hence positive semidefinite

Bài toán tối ưu

Một bài toán tối ưu hóa gồm có

- x vectơ của các *biến*
- *hàm mục tiêu* f là một hàm (vô hướng) mà chúng ta muốn cực đại hóa hay cực tiểu hóa
- Các *hàm điều kiện* g_i, h_i là các hàm (vô hướng) của x định nghĩa các đẳng thức hay bất đẳng thức mà x phải thỏa mãn.

Bài toán tối ưu

Một bài toán tối ưu hóa gồm có

- x vectơ của các *biến*
- *hàm mục tiêu* f là một hàm (vô hướng) mà chúng ta muốn cực đại hóa hay cực tiểu hóa
- Các *hàm điều kiện* g_i, h_i là các hàm (vô hướng) của x định nghĩa các đẳng thức hay bất đẳng thức mà x phải thỏa mãn.

Bài toán tối ưu có thể viết dưới dạng

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l. \end{aligned}$$

Biến đổi giữa các dạng bài toán tối ưu

max $\longleftrightarrow$ min:

$$\begin{array}{ll} \max & f(x) \\ \text{s.t.} & g_i(x) \le 0, \quad i = 1, \dots, k \\ & h_j(x) = 0, \quad j = 1, \dots, l. \end{array} \iff \begin{array}{ll} \min & -f(x) \\ \text{s.t.} & g_i(x) \le 0, \quad i = 1, \dots, k \\ & h_j(x) = 0, \quad j = 1, \dots, l. \end{array}$$

Biến đổi giữa các dạng bài toán tối ưu

max $\longleftrightarrow$ min:

$$\begin{array}{lll} \max & f(x) & \\ \text{s.t.} & g_i(x) \le 0, & i = 1, \ldots, k \\ & h_j(x) = 0, & j = 1, \ldots, l. \end{array} \iff \begin{array}{lll} \min & -f(x) & \\ \text{s.t.} & g_i(x) \le 0, & i = 1, \ldots, k \\ & h_j(x) = 0, & j = 1, \ldots, l. \end{array}$$

Vế phải khác 0:

$$\begin{array}{lll} \max & f(x) & \\ \text{s.t.} & g_i(x) \le b_i, & i = 1, \ldots, k \\ & h_j(x) = c_j, & j = 1, \ldots, l. \end{array} \iff \begin{array}{lll} \min & -f(x) & \\ \text{s.t.} & \bar{g}_i(x) \le 0, & i = 1, \ldots, k \\ & \bar{h}_j(x) = 0, & j = 1, \ldots, l. \end{array}$$

với $\bar{g}_i = g_i - b_i$, $\bar{h}_j = h_j - c_j$.

Biến đổi giữa các dạng bài toán tối ưu

max $\longleftrightarrow$ min:

$$\begin{array}{lll} \max & f(x) & \\ \text{s.t.} & g_i(x) \le 0, & i = 1,\dots,k \\ & h_j(x) = 0, & j = 1,\dots,l. \end{array} \iff \begin{array}{lll} \min & -f(x) & \\ \text{s.t.} & g_i(x) \le 0, & i = 1,\dots,k \\ & h_j(x) = 0, & j = 1,\dots,l. \end{array}$$

Vế phải khác 0:

$$\begin{array}{lll} \max & f(x) & \\ \text{s.t.} & g_i(x) \le b_i, & i = 1,\dots,k \\ & h_j(x) = c_j, & j = 1,\dots,l. \end{array} \iff \begin{array}{lll} \min & -f(x) & \\ \text{s.t.} & \bar{g}_i(x) \le 0, & i = 1,\dots,k \\ & \bar{h}_j(x) = 0, & j = 1,\dots,l. \end{array}$$

với $\bar{g}_i = g_i - b_i$, $\bar{h}_j = h_j - c_j$.

$\le \longleftrightarrow \ge$:

$$\begin{array}{lll} \min & f(x) & \\ \text{s.t.} & g_i(x) \ge 0, & i = 1,\dots,k \\ & h_j(x) = 0, & j = 1,\dots,l. \end{array} \iff \begin{array}{lll} \min & f(x) & \\ \text{s.t.} & -g_i(x) \le 0, & i = 1,\dots,k \\ & h_j(x) = 0, & j = 1,\dots,l. \end{array}$$

Biến đổi giữa các dạng bài toán tối ưu

$= \longrightarrow \leq$:

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, k \\
& h_j(x) = 0, \quad j = 1, \dots, l.
\end{aligned}
\quad \Longrightarrow \quad
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, k \\
& h_j(x) \leq 0, \quad j = 1, \dots, l \\
& -h_j(x) \leq 0, \quad j = 1, \dots, l.
\end{aligned}
$$

Biến đổi giữa các dạng bài toán tối ưu

$= \longrightarrow \leq$:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \dots, k \\ & h_j(x) = 0, \quad j = 1, \dots, l. \end{array} \Longrightarrow \begin{array}{lll} \min & f(x) & \\ \text{s.t.} & g_i(x) \leq 0, & i = 1, \dots, k \\ & h_j(x) \leq 0, & j = 1, \dots, l \\ & -h_j(x) \leq 0, & j = 1, \dots, l. \end{array}$$

$\leq \longrightarrow =$: Thêm biến bù

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \dots, k \\ & h_j(x) = 0, \quad j = 1, \dots, l. \end{array} \Longrightarrow \begin{array}{lll} \min & f(x) & \\ \text{s.t.} & g_i(x) + s_i = 0, & i = 1, \dots, k \\ & h_j(x) = 0, & j = 1, \dots, l \\ & s_i \geq 0, & i = 1, \dots, k. \end{array}$$

Một số thuật ngữ tối ưu hóa

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l \end{aligned}$$

- $g_i(x) \leq 0$ gọi là *ràng buộc bất đẳng thức*, $h_j(x) = 0$ gọi là *ràng buộc đẳng thức*.

Một số thuật ngữ tối ưu hóa

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l \end{aligned}$$

- $g_i(x) \leq 0$ gọi là *ràng buộc bất đẳng thức*, $h_j(x) = 0$ gọi là *ràng buộc đẳng thức*.
- x thỏa mãn các điều kiện được gọi là một *nghiệm chấp nhận được (feasible solution)*.

Một số thuật ngữ tối ưu hóa

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l \end{aligned}$$

- $g_i(x) \leq 0$ gọi là *ràng buộc bất đẳng thức*, $h_j(x) = 0$ gọi là *ràng buộc đẳng thức*.
- x thỏa mãn các điều kiện được gọi là một *nghiệm chấp nhận được (feasible solution)*.
- Tập hợp các nghiệm CNĐ được gọi là *miền CNĐ (feasible region)*.

Một số thuật ngữ tối ưu hóa

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l \end{aligned}$$

- $g_i(x) \leq 0$ gọi là *ràng buộc bất đẳng thức*, $h_j(x) = 0$ gọi là *ràng buộc đẳng thức*.
- x thỏa mãn các điều kiện được gọi là một *nghiệm chấp nhận được (feasible solution)*.
- Tập hợp các nghiệm CNĐ được gọi là *miền CNĐ (feasible region)*.
- Nếu miền CNĐ bằng rỗng ta nói bài toán *không có nghiệm CNĐ (infeasible)*, ngược lại là *có nghiệm chấp nhận được (feasible)*.

Một số thuật ngữ tối ưu hóa

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l. \end{aligned}$$

- Với một nghiệm CNĐ x, nếu $g_i(x) = 0$ ta nói g_i *active* tại x.

Một số thuật ngữ tối ưu hóa

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\
& h_j(x) = 0, \quad j = 1, 2, \ldots, l.
\end{aligned}
$$

- Với một nghiệm CNĐ x, nếu $g_i(x) = 0$ ta nói g_i *active* tại x.
- Giá trị nhỏ nhất của $f(x)$ trên miền CNĐ gọi là *giá trị tối ưu (optimal value)*, kí hiệu f^*.

Một số thuật ngữ tối ưu hóa

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l. \end{aligned}$$

- Với một nghiệm CNĐ x, nếu $g_i(x) = 0$ ta nói g_i *active* tại x.
- Giá trị nhỏ nhất của $f(x)$ trên miền CNĐ gọi là *giá trị tối ưu (optimal value)*, kí hiệu f^*.
- Nếu x là một nghiệm CNĐ và $f(x) = f^*$ thì ta gọi x là một *nghiệm tối ưu (optimal solution)*.

Một số thuật ngữ tối ưu hóa

$$\begin{array}{rll} \min & f(x) & \\ \text{s.t.} & g_i(x) \leq 0, & i = 1, 2, \ldots, k \\ & h_j(x) = 0, & j = 1, 2, \ldots, l. \end{array}$$

- Với một nghiệm CNĐ x, nếu $g_i(x) = 0$ ta nói g_i *active* tại x.
- Giá trị nhỏ nhất của $f(x)$ trên miền CNĐ gọi là *giá trị tối ưu (optimal value)*, kí hiệu f^*.
- Nếu x là một nghiệm CNĐ và $f(x) = f^*$ thì ta gọi x là một *nghiệm tối ưu (optimal solution)*.
- Nếu x là một nghiệm CNĐ và $f(x) \leq f^* + \varepsilon$ thì ta gọi x là ε-*suboptimal*.

Example: lasso

Given $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$, consider the lasso problem:

$$\begin{array}{ll} \min_\beta & \|y - X\beta\|_2^2 \\ \text{subject to} & \|\beta\|_1 \le s \end{array}$$

Is this convex? What is the criterion function? The inequality and equality constraints? Feasible set? Is the solution unique, when:

- $n \ge p$ and X has full column rank?
- $p > n$ ("high-dimensional" case)?

How do our answers change if we changed criterion to Huber loss:

$$\sum_{i=1}^n \rho(y_i - x_i^T \beta), \quad \rho(z) = \begin{cases} \frac{1}{2} z^2 & |z| \le \delta \\ \delta|z| - \frac{1}{2}\delta^2 & \text{else} \end{cases} \quad ?$$

Example: support vector machines

Given $y \in \{-1, 1\}^n$, $X \in \mathbb{R}^{n \times p}$ with rows $x_1, \ldots x_n$, consider the support vector machine or SVM problem:

$$\begin{aligned} \min_{\beta, \beta_0, \xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & \xi_i \geq 0, \ i = 1, \ldots n \\ & y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i, \ i = 1, \ldots n \end{aligned}$$

Is this convex? What is the criterion, constraints, feasible set? Is the solution (β, β_0, ξ) unique? What if changed the criterion to

$$\frac{1}{2}\|\beta\|_2^2 + \frac{1}{2}\beta_0^2 + C \sum_{i=1}^n \xi_i^{1.01}?$$

For original criterion, what about β component, at the solution?

Các loại bài toán tối ưu

- Khi không có điều kiện ràng buộc thì ta có bài toán tối ưu hóa không ràng buộc.
- Ngược lại ta có bài toán tối ưu hóa có ràng buộc.

Các loại bài toán tối ưu

- Khi không có điều kiện ràng buộc thì ta có bài toán tối ưu hóa không ràng buộc.
- Ngược lại ta có bài toán tối ưu hóa có ràng buộc.

Tiếp theo ta xét các loại bài toán tối ưu hóa

- tuyến tính
- lồi
- rời rạc.

Bài toán tối ưu tuyến tính

Dạng chính tắc

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Dạng chuẩn tắc

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \geq b \\ & x \geq 0. \end{aligned}$$

Dạng tổng quát

$$\begin{aligned} \min \quad & c^T x + d^T y \\ \text{s.t.} \quad & Ax + By = a \\ & Cx + Dy \geq b \\ & x \geq 0. \end{aligned}$$

Bài toán tối ưu lồi

Bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots k \end{aligned}$$

được gọi là là *bài toán tối ưu lồi* nếu f và g_i là các hàm lồi.

Bài toán tối ưu lồi

Bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \dots k \end{aligned}$$

được gọi là là *bài toán tối ưu lồi* nếu f và g_i là các hàm lồi.

Khi đó mỗi điều kiện ứng với một tập mức dưới lồi nên miền chấp nhận được là giao của chúng cũng là một tập lồi.

Bài toán tối ưu lồi

Bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \dots k \end{aligned}$$

được gọi là là *bài toán tối ưu lồi* nếu f và g_i là các hàm lồi.

Khi đó mỗi điều kiện ứng với một tập mức dưới lồi nên miền chấp nhận được là giao của chúng cũng là một tập lồi.

Bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, k \\ & h_j(x) = 0, \quad j = 1, \dots, l. \end{aligned}$$

là lồi nếu

Bài toán tối ưu lồi

Bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots k \end{aligned}$$

được gọi là là *bài toán tối ưu lồi* nếu f và g_i là các hàm lồi.

Khi đó mỗi điều kiện ứng với một tập mức dưới lồi nên miền chấp nhận được là giao của chúng cũng là một tập lồi.

Bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, k \\ & h_j(x) = 0, \quad j = 1, \ldots, l. \end{aligned} \iff \begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, k \\ & h_j(x) \leq 0, \quad j = 1, \ldots, l \\ & -h_j(x) \leq 0, \quad j = 1, \ldots, l \end{aligned}$$

là lồi nếu

- f và g_i là các hàm lồi
- h_j vừa là hàm lồi vừa là hàm lõm.

Bổ đề

$f : \mathbb{R}^n \to \mathbb{R}$ là hàm vừa lồi vừa lõm khi và chỉ khi f là hàm affine.

Bài toán tối ưu lồi

Bổ đề

$f : \mathbb{R}^n \to \mathbb{R}$ là hàm vừa lồi vừa lõm khi và chỉ khi f là hàm affine.

Chứng minh.

f vừa là hàm lồi vừa là hàm lõm thì ta có

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y), \ \forall \lambda \in [0,1], x, y \in \mathbb{R}^n.$$

Bổ đề

$f : \mathbb{R}^n \to \mathbb{R}$ là hàm vừa lồi vừa lõm khi và chỉ khi f là hàm affine.

Chứng minh.

f vừa là hàm lồi vừa là hàm lõm thì ta có

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y), \ \forall \lambda \in [0,1], x, y \in \mathbb{R}^n.$$

Suy ra với mọi $\lambda \in [0,1]$ ta có

$$f(\lambda x) = f(\lambda x + (1-\lambda)0) = \lambda f(x) + (1-\lambda)f(0),$$

Bài toán tối ưu lồi

Bổ đề

$f : \mathbb{R}^n \to \mathbb{R}$ là hàm vừa lồi vừa lõm khi và chỉ khi f là hàm affine.

Chứng minh.

f vừa là hàm lồi vừa là hàm lõm thì ta có

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y), \ \forall \lambda \in [0,1], x, y \in \mathbb{R}^n.$$

Suy ra với mọi $\lambda \in [0,1]$ ta có

$$f(\lambda x) = f(\lambda x + (1-\lambda)0) = \lambda f(x) + (1-\lambda)f(0),$$

hay

$$g(\lambda x) = \lambda g(x)$$

với $g(x) := f(x) - f(0)$.

Bổ đề

$f : \mathbb{R}^n \to \mathbb{R}$ là hàm vừa lồi vừa lõm khi và chỉ khi f là hàm affine.

Chứng minh.

f vừa là hàm lồi vừa là hàm lõm thì ta có

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y),\ \forall \lambda \in [0,1], x, y \in \mathbb{R}^n.$$

Suy ra với mọi $\lambda \in [0,1]$ ta có

$$f(\lambda x) = f(\lambda x + (1-\lambda)0) = \lambda f(x) + (1-\lambda)f(0),$$

hay

$$g(\lambda x) = \lambda g(x)$$

với $g(x) := f(x) - f(0)$. Từ đó ta cũng suy ra với mọi $\lambda > 1$ ta có

$$g(x) = g\left(\frac{1}{\lambda}(\lambda x)\right) = \frac{1}{\lambda} g(\lambda x)$$

hay

$$\lambda g(x) = g(\lambda x).$$

Bài toán tối ưu lồi

Chứng minh.

Ta có

$$g(x+y) = 2g\left(\frac{1}{2}x + \frac{1}{2}y\right) = 2f\left(\frac{1}{2}x + \frac{1}{2}y\right) - 2f(0)$$

$\square$

Chứng minh.

Ta có

$$g(x+y) = 2g\left(\frac{1}{2}x + \frac{1}{2}y\right) = 2f\left(\frac{1}{2}x + \frac{1}{2}y\right) - 2f(0)$$
$$= 2\left(\frac{1}{2}f(x) + \frac{1}{2}f(y)\right) - 2f(0)$$

□

Chứng minh.

Ta có

$$\begin{aligned} g(x+y) &= 2g\left(\frac{1}{2}x + \frac{1}{2}y\right) = 2f\left(\frac{1}{2}x + \frac{1}{2}y\right) - 2f(0) \\ &= 2\left(\frac{1}{2}f(x) + \frac{1}{2}f(y)\right) - 2f(0) \\ &= f(x) + f(y) - 2f(0) = g(x) + g(y). \end{aligned}$$

□

Chứng minh.

Ta có

$$\begin{aligned} g(x+y) &= 2g\left(\frac{1}{2}x + \frac{1}{2}y\right) = 2f\left(\frac{1}{2}x + \frac{1}{2}y\right) - 2f(0) \\ &= 2\left(\frac{1}{2}f(x) + \frac{1}{2}f(y)\right) - 2f(0) \\ &= f(x) + f(y) - 2f(0) = g(x) + g(y). \end{aligned}$$

Suy ra g là hàm tuyến tính và do đó f là hàm affine. $\square$

Bài toán tối ưu lồi

Chứng minh.

Ta có

$$\begin{aligned} g(x+y) &= 2g\left(\frac{1}{2}x + \frac{1}{2}y\right) = 2f\left(\frac{1}{2}x + \frac{1}{2}y\right) - 2f(0) \\ &= 2\left(\frac{1}{2}f(x) + \frac{1}{2}f(y)\right) - 2f(0) \\ &= f(x) + f(y) - 2f(0) = g(x) + g(y). \end{aligned}$$

Suy ra g là hàm tuyến tính và do đó f là hàm affine. $\square$

Như vậy bài toán tối ưu lồi sẽ có dạng

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, k \\ & Ax = b \end{aligned}$$

với f và g_i, $i = 1, 2, \dots k$, là các hàm lồi.

Tối ưu liên tục vs. tối ưu rời rạc

Đôi khi chúng ta đòi hỏi biến phải là số nguyên hay nhị phân. Ví dụ như khi biến của chúng ta là số ôtô cần để vận chuyển, số nhân lực, hay biến quyết định có làm một việc gì đó không.

Khi đó ta sẽ có bài toán tối ưu hóa nguyên hoặc tối ưu hóa nhị phân.

Trong nội dung môn học này chúng ta sẽ không xét đến dạng bài toán trên.

Xét bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l. \end{aligned}$$

Kí hiệu $\mathcal{X}$ là miền CNĐ.

Cực trị địa phương và toàn cục

Xét bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l. \end{aligned}$$

Kí hiệu $\mathcal{X}$ là miền CNĐ.

$\bar{x} \in \mathcal{X}$ được gọi là nghiệm *tối ưu địa phương (locally optimal)* nếu tồn tại $R > 0$ sao cho

$$f(\bar{x}) \le f(x), \ \forall x \in \mathcal{X} : \|x - \bar{x}\|_2 \le R.$$

Cực trị địa phương và toàn cục

Xét bài toán tối ưu

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, k \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, l. \end{aligned}$$

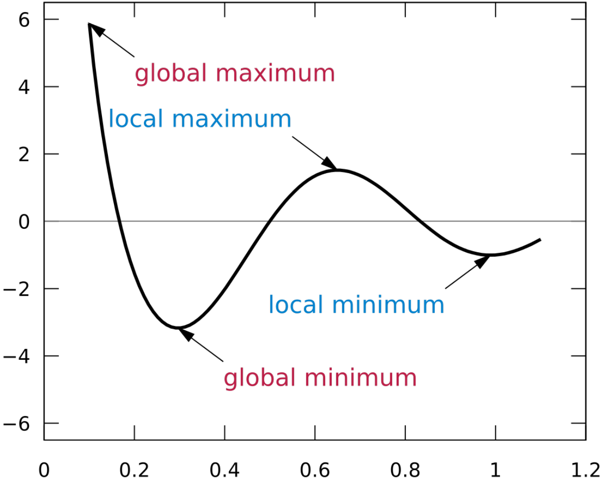
Kí hiệu $\mathcal{X}$ là miền CNĐ.

$\bar{x} \in \mathcal{X}$ được gọi là nghiệm *tối ưu địa phương (locally optimal)* nếu tồn tại $R > 0$ sao cho

$$f(\bar{x}) \leq f(x), \ \forall x \in \mathcal{X} : \|x - \bar{x}\|_2 \leq R.$$

Để phân biệt với nghiệm tối ưu địa phương với nghiệm tối ưu, ta còn gọi nghiệm tối ưu là nghiệm *tối ưu toàn cục (globally optimal)*.

Cực trị địa phương và toàn cục



Cực tiểu địa phương của bài toán tối ưu lồi

Định lý

Nghiệm tối ưu địa phương của một bài toán tối ưu lồi cũng là nghiệm tối ưu toàn cục.

Chứng minh.

Chứng minh phản chứng dựa trên tính lồi của miền CNĐ và hàm mục tiêu. □

Điểm dừng

Điểm dừng (stationary point) x của một hàm khả vi f là một điểm tại đó $\nabla f(x) = 0$.

Điểm dừng

Điểm dừng (stationary point) x của một hàm khả vi f là một điểm tại đó $\nabla f(x) = 0$.

Ta sẽ thấy mỗi điểm cực trị địa phương là một điểm dừng.

Điểm dừng

Điểm dừng (stationary point) x của một hàm khả vi f là một điểm tại đó $\nabla f(x) = 0$.

Ta sẽ thấy mỗi điểm cực trị địa phương là một điểm dừng. Tuy nhiên cần lưu ý rằng không phải mọi điểm dừng đều là cực trị địa phương.

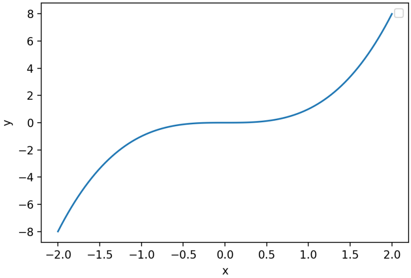
Điểm dừng

Điểm dừng (stationary point) x của một hàm khả vi f là một điểm tại đó $\nabla f(x) = 0$.

Ta sẽ thấy mỗi điểm cực trị địa phương là một điểm dừng. Tuy nhiên cần lưu ý rằng không phải mọi điểm dừng đều là cực trị địa phương.

Ví dụ: $f(x) = x^3$

$f'(0) = 0$ nhưng 0 không phải là một điểm cực trị.



Đường mức

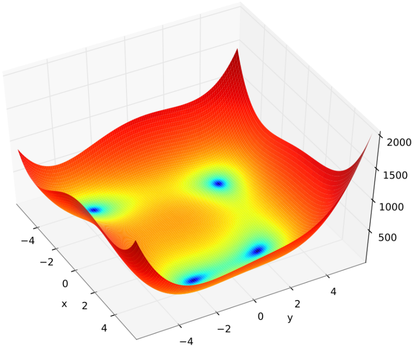
Đường mức của một hàm số $f : \mathbb{R}^n \to \mathbb{R}$ là một tập hợp có dạng

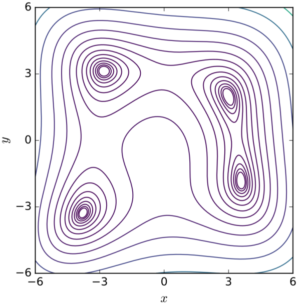
$$L_c(f) = \{x \mid f(x) = c\}.$$

Đường mức

Đường mức của một hàm số $f : \mathbb{R}^n \to \mathbb{R}$ là một tập hợp có dạng

$$L_c(f) = \{x \mid f(x) = c\}.$$





$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$

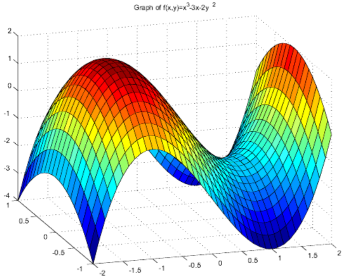
Log-spaced level curve

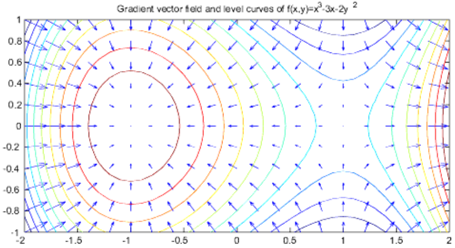
Với hàm f khả vi, gradient tại một điểm hoặc bằng 0 hoặc vuông góc với đường mức tại điểm đó.

Đường mức và gradient của hàm 2 biến

Với hàm f khả vi, gradient tại một điểm hoặc bằng 0 hoặc vuông góc với đường mức tại điểm đó.

Ví dụ: $f(x, y) = x^3 - 3x - 2y^2$





Tài liệu tham khảo

- Chương 2 và chương 3, S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press.
- Chương 1, J. Nocedal and S. Wright, *Numerical Optimization*, Springer.