UCLA ECE236 A/B/C

Chapter 2: Basic Concepts

Chao Wang

SUSTech

Outline

Vector Norms and Matrix Norms

Vector norms 向量范数

零范数

$\|x\|_0 : \#\{x_i \neq 0\}$ 非标准范数，不满足②

e.g. $\|2x\|_0 = \|x\|_0$

Definition

A function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is called a vector norm if it has the following properties:

1. $\|\mathbf{x}\| \geq 0$ for any vector $\mathbf{x} \in \mathbb{R}^n$, and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$;
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for any vector $\mathbf{x} \in \mathbb{R}^n$ and any scalar $\alpha \in \mathbb{R}$;
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. 三角不等式

Notes: The last property is called the triangle inequality. It should be noted that when $n = 1$, the absolute value function is a vector norm.

Vector norms

The most commonly used vector norms belong to the family of p-norms, or ℓ_p-norms, which are defined by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

$\leq |\vec{x}_i| + |\vec{y}_i|$

$\|x+y\|_1 = |\vec{x}_i + \vec{y}_i| + \dots$

皆满足3准则

- $p = 1$: The ℓ_1-norm

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

- $p = 2$: The ℓ_2-norm or Euclidean norm

$\|:\quad \left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)$

向量和之积 $\leq$ 向量积之和.

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\mathbf{x}^T\mathbf{x}}$$

- $p = \infty$: The $\ell_\infty - norm$

绝对最大值.

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

Two important inequalities

- It can be shown that the ℓ_2-norm satisfies the Cauchy-Schwarz inequality

$$\left|\mathbf{x}^T\mathbf{y}\right| \le \|\mathbf{x}\|_2\|\mathbf{y}\|_2$$

for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. This inequality is useful for showing that the ℓ_2-norm satisfies the triangle inequality.

- It is a special case of the Holder inequality

推广

$$\left|\mathbf{x}^T\mathbf{y}\right| \le \|\mathbf{x}\|_p\|\mathbf{y}\|_q, \underline{\frac{1}{p} + \frac{1}{q} = 1}$$

$$\|x+y\|_2^2 = \langle x+y, x+y\rangle$$
$$= \|x\|_2^2 + \|y\|_2^2 + 2\langle x, y\rangle$$
$$\le \|x\|_2^2 + \|y\|_2^2 + 2\|x\|_2\|y\|_2$$
$$= (\|x\|_2 + \|y\|_2)^2$$

三角不等式

$$\Rightarrow \underline{\|x+y\|_2 \le \|x\|_2 + \|y\|_2}$$

Matrix norms 矩阵范数

Definition

Any vector norm induces a matrix norm. It can be shown that given a vector norm, defined appropriately for m-vectors and n-vectors, the function $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ defined by

$$\|A\| = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$$

is a matrix norm. It is called the natural, or induced, matrix norm.

- The ℓ_1-norm:

$$\begin{bmatrix} = \end{bmatrix} \begin{bmatrix} | \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

$$\|A\|_1 = \max_{\|\mathbf{x}\|_1=1} \|A\mathbf{x}\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|.$$

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \end{bmatrix} \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \end{bmatrix} \cdots$$

并找最大的列.

That is, the ℓ_1-norm of a matrix is its maximum column sum.

- The ℓ_∞-norm:

$$\|A\|_\infty = \max_{\|\mathbf{x}\|_\infty=1} \|A\mathbf{x}\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|.$$

That is, the ℓ_∞-norm of a matrix is its maximum row sum.

- The ℓ_2-norm:

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2$$

$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$ （称为A的2－范数）
其中$\lambda_{\max}(A^T A)$表示$A^T A$的最大特征值
即$f(\lambda) = |\lambda E - A^T A| = 0$

$\|A\|_2$: 最大奇异值.

- The Frobenius norm: A 向量化的二范数.

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}. \quad = \sqrt{\mathrm{Tr}(A^T A)}.$$

$$\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1$$

$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n| = 1.$ $\quad \max \begin{bmatrix} \sum a_{1j} x_j \\ \sum a_{2j} x_j \\ \vdots \\ a_{mj} x_j \end{bmatrix},$

$$\therefore \|A\|_1 = \left|\sum_{j=1}^{n} a_{1j} x_j\right| + \cdots + \left|\sum_{j=1}^{n} a_{mj} x_j\right|.$$

$$\le \sum_{j=1}^{n} |a_{1j} x_j| + \cdots + \sum_{j=1}^{n} |a_{mj} x_j|$$

$$= \sum_{j=1}^{n} |a_{1j}||x_j| + \cdots + \sum_{j=1}^{n} |a_{mj}||x_j|$$

$$= \sum_{i=1}^{m} (\underbrace{|a_{i1}|}_{}^{\max})|x_1| + \cdots + \sum_{i=1}^{m} (|a_{in}|)|x_n|.$$

$$\le \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|$$

$$\|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty \qquad Ax = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{bmatrix}$$

$$\|x\|_\infty \Rightarrow \max_{1\le j\le n} |x_j| = 1.$$

$$\Rightarrow \max_{1\le i\le m} \left|\sum_{j=1}^{n} a_{ij}x_j\right|$$

$$\sum_{j=1}^{n} |a_{ij}|\,|x_j| \le \sum_{j=1}^{n} |a_{ij}|$$

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$$

$$\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2} = 1. \qquad x_1^2 + \cdots + x_n^2 = 1.$$

$$\|Ax\|_2^2 = \left(\sum_{j=1}^{n} a_{1j}x_j\right)^2 + \cdots + \sum_{j=1}^{n}$$

$\lambda_1 \cdots \lambda_n$

Linear System

Linear System

$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = b$

Given

$$\left\{\begin{array}{ccccccccc} \underset{a_1}{a_{11}} x_1 & + & \underset{a_2}{a_{12}} x_2 & + & \cdots & + & \underset{a_n}{a_{1n}} x_n & = & b_1, \\ \vdots & & & & & & & & \vdots, \\ a_{m1} x_1 & + & a_{m2} x_2 & + & \cdots & + & a_{mn} x_n & = & b_m. \end{array}\right. \tag{1}$$

In matrix form, we have $\mathbf{Ax} = \mathbf{b}$.

- the *augmented matrix* $\mathbf{A}_b$ of $\mathbf{A}$ that is defined by

增广矩阵

$$\mathbf{A}_b = [\mathbf{A}, \mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n, \mathbf{b}], \tag{2}$$

- two cases:

找不到 $a_1 \cdots a_n$ 来表示 b.

1. rank $(\mathbf{A}) <$ rank $(\mathbf{A}_b)$. The system $\mathbf{Ax} = \mathbf{b}$ has no solutions.
2. rank $(\mathbf{A}) =$ rank $(\mathbf{A}_b) = k$: at least one solution exists

Solution to a linear system

- $\mathbf{A}$ is an m-by-n matrix with $\text{rank}(\mathbf{A}) = m < n$. Suppose that from the m columns of $\mathbf{A}$.

⇒ B可逆 $m \times m$.

- Select a subset $\mathbf{B}$ of m *linearly independent* columns. Then $\mathbf{Ax} = \mathbf{b}$ can be written as

$B x_B = b - N x_N$

$x_B = B^{-1} b - B^{-1} N x_N$

$$\mathbf{Bx}_B + \mathbf{Nx}_N = \mathbf{b}, \tag{3}$$

特殊解: $\begin{cases} x_B: B^{-1} b \\ x_N: 0 \end{cases}$

自由变量

where

$$\mathbf{B} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m],$$

$$\mathbf{N} = \begin{bmatrix} a_{1,m+1} & a_{1,m+2} & \cdots & a_{1n} \\ a_{2,m+1} & a_{2,m+2} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m,m+1} & a_{m,m+2} & \cdots & a_{mn} \end{bmatrix} = [\mathbf{a}_{m+1}, \mathbf{a}_{m+2}, \ldots, \mathbf{a}_n], \tag{4}$$

$$\mathbf{x}_B = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad \text{and} \quad \mathbf{x}_N = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_n \end{bmatrix}.$$

Solution to a linear system

- General solution: $\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix}$ where $\underline{\mathbf{x}_B = \mathbf{B}^{-1}(\mathbf{b} - \mathbf{N}\mathbf{x}_N})$ and $\mathbf{x}_N$ is arbitrary.
- One specific solution: $\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{0} \end{bmatrix}$ where $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b}$ and $\mathbf{x}_N = \mathbf{0}$

Basic solution

Definition

Consider $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{A}$ being $m \times n$ and rank$(\mathbf{A}) = m < n$. Let $\mathbf{B}$ be a matrix formed by choosing m linearly independent columns of A.

- If all the $(n - m)$ variables not associated with columns of $\mathbf{B}$ are set equal to zero, then the solution to the resulting system of equations is called a **basic solution** with respect to $\mathbf{B}$.
- We call the m variables x_i associated with columns of $\mathbf{B}$ the **basic variables**, the other variables are called the **non-basic variables**.
- A basic solution to $\mathbf{Ax} = \mathbf{b}$ is **degenerate** if one or more of the m basic variables vanish. Otherwise, it is said to be **non-degenerate**.

Example: $\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{0} \end{bmatrix}$ is basic solution with $\mathbf{x}_B$ being basic variables. If $\mathbf{x}_B$ has zeros entries, then $\mathbf{x}$ is degenerate, otherwise is non-degenerate.

Derivative

Derivative

$f(x) \in \mathbb{R}^1, \ x \in \mathbb{R}^n$

行向量.

- First-order derivative for multi-variable functions: row vector

$$Df := \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right].$$

梯度 列向量

- Gradient of $\nabla f = (Df)^T$, which is a column vector.
- A gradient represents the slope of the tangent of the graph of function. It gives the linear approximation of f at a point. It points toward the greatest rate of increase.

梯度表现了函数增长最快的方向.

Derivative

- Hessian (i.e., second-derivative) of f :

  矩阵 (导数的雅可比)

  $$F(x) := D^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ & \vdots & \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

  which is a symmetric matrix.

  $\frac{\partial^2 f}{\partial x_n \partial x_m} = \frac{\partial^2 f}{\partial x_m \partial x_n}$

- For one-dimensional function $f(x)$ where $x \in \mathbb{R}$, it reduces to $f''(x)$.
- $F(x)$ is the Jacobian of $\nabla f(x)$, that is, $F(x) = J(\nabla f(x))$.
- Alternative notation: $H(x)$ and $\nabla^2 f(x)$ are also used for Hessian.
- Gradient and Hessian are local properties that help us recognize local solutions and determine a direction to move at toward the next point.

Example

$$\nabla f = \begin{bmatrix} 3x_1^2+2x_1-x_2+5 \\ -x_1+2x_2+8 \end{bmatrix} \quad F(x) = \begin{bmatrix} 6x_1+2 & -1 \\ -1 & 2 \end{bmatrix}$$

Consider

$$f(x_1, x_2) = x_1^3 + x_1^2 - x_1x_2 + x_2^2 + 5x_1 + 8x_2 + 4$$

Then,

$$\nabla f(x) = \begin{bmatrix} 3x_1^2 + 2x_1 - x_2 + 5 \\ -x_1 + 2x_2 + 8 \end{bmatrix} \in \mathbb{R}^2$$

and

$$F(x) = \begin{bmatrix} 6x_1 + 2 & -1 \\ -1 & 2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

二次方函数.

Observation: if f is a quadratic function (remove x_1^3 in the above example), $\nabla f(x)$ is a linear vector and $F(x)$ is a symmetric constant matrix for any x.

降为1次. 降为常数.

$$f(x_1, x_2) = x_1^3 + x_1^2 - x_1x_2 + x_2^2 + 5x_1 + 8x_2 + 4$$

$$Df = [3x_1^2 + 2x_1 - x_2 + 5,\ -x_1 + 2x_2 + 8].$$

$$F(x) = D^2 f = \begin{bmatrix} 6x_1 + 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Suppose $\phi \in \mathcal{C}^m$(m times continuously differentiable). The Taylor expansion of ϕ at a point a is

$$\phi(a+h) = \phi(a) + \phi'(a)h + \frac{\phi''(a)}{2!}h^2 + \cdots + \frac{\phi^m(a)}{m!}h^m + o\left(h^m\right).$$

There are other ways to write the last two terms.

Example: Consider $x, d \in \mathbb{R}^n$ and $f \in \mathcal{C}^2$. Define $\phi(\alpha) = f(x + \alpha d)$. Then,

$$\phi'(\alpha) = \nabla f(x+\alpha d)^T d$$
$$\phi''(\alpha) = d^T F(x+\alpha d) d$$

令$\alpha=0$.

$\phi(0)=f(x)$,

Hence,

$$f(x+\alpha d) = f(x) + \left(\nabla f(x)^T d\right)\alpha + o(\alpha)$$
$$= f(x) + \left(\nabla f(x)^T d\right)\alpha + \frac{d^T F(x) d}{2}\alpha^2 + o\left(\alpha^2\right)$$

Convex set

General optimization problem

General optimization problem:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.,} \quad & f_i(x) \le b_i, i = 1, \ldots, m. \end{aligned}$$

- $x = (x_1, \ldots, x_n)$: optimization variables
- $f_0 : \mathbb{R}^n \to \mathbb{R}$: objective function
- $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$: constraint functions

General optimization problem

General optimization problem

- very difficult to solve
- methods involve some issues: long computation time, or can not always find the global solution

Exceptions: certain problem classes can be solved efficiently and reliably

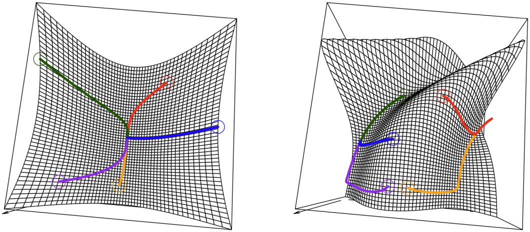
- linear programming problems
- convex optimization problems 优点：局部解即为全局解
- …

Convex optimization

Definition of convex optimization problem:

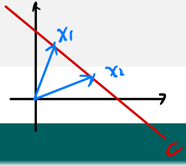
$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.,} \quad & f_i(x) \le b_i, i = 1, \ldots, m \end{aligned}$$

①.
- feasible region is a convex set. $f_i(x)$.
  ②.
- objective and constraint functions are convex functions $f_0(x)$
- For the convex optimization problem, local optima are global optima!



Convex   Nonconvex

Affine set



Definition

A set C is affine if the line through between any two points in C lies in C, i.e., if for any $x_1, x_2 \in C$ and any θ, we have

$$\theta x_1 + (1-\theta)x_2 \in C$$

- example: solution set of linear equations $C = \{x \mid Ax = b\}$

Proof.

$\forall x_1, x_2 \in C, \theta \in \mathbb{R}$,

$$A(\theta x_1 + (1-\theta)x_2) = \theta A x_1 + (1-\theta)Ax_2 = \theta b + (1-\theta)b = b.$$

$\square$

$C = \{x \mid Ax = b\}$

$\forall x_1, x_2 \in C$ then $Ax_1 = b$

$\forall \theta$ $Ax_2 = b.$

$\Rightarrow \theta A x_1 + (1-\theta) A x_2 = \theta b + (1-\theta) b$

$\Rightarrow A(\theta x_1 + (1-\theta) x_2) = b.$

Convex set

Definition

A set C is convex if the line segment between any two points in C lies in C, i.e., if for any $x_1, x_2 \in C$ and any θ with $0 \leq \theta \leq 1$, we have

$$\theta x_1 + (1-\theta)x_2 \in C$$

- Generate a convex set based on given points: Convex hull ($conv(S)$): Set of all convex combinations of points in S: $x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k, \quad \sum_{i=1}^{k} \theta_i = 1, \theta_i \geq 0.$

Convex cone

Definition

Conic combination of x_1 and x_2 : any point of the form

$$x = \theta_1 x_1 + \theta_2 x_2 \in C, \text{ for all } \theta_1 \geq 0, \theta_2 \geq 0$$

- Convex set is one specific type of convex cone, i.e., convex set $\Rightarrow$ convex cone
- Summary:
  - Affine: $\theta_1 = 1 - \theta_2$
  - Convex: $\theta_1 = 1 - \theta_2, \theta_i \in [0, 1]$
  - Cone: $\theta_i \geq 0$

Example of convex set

- hyperplane: set of form $\{x|a^T x = b\}$ $(a \neq 0)$
  Proof: hyperplane is an affine set, so it is convex.
- halfspace: set of the form $C = \{x|a^T x \leq b\}$ $(a \neq 0$ a is the normal vector for all $\theta \in [0,1])$
  Proof: $\forall x_1, x_2 \in C, \theta \in [0,1]$

$$\begin{aligned} a^T(\theta x_1 + (1-\theta)x_2) &= \theta a^T x_1 + (1-\theta)a^T x_2 \\ &\leq \theta b + (1-\theta)b = b \end{aligned}$$

Example of convex set

- Norm ball with center x_c and radius r :

圆心

$$B\left(x_c, r\right)=\left\{x \mid \left\|x-x_c\right\| \leq r\right\}=\left\{x_c+r u \mid \|u\| \leq 1\right\}$$

Proof: $\forall x_1, x_2 \in B(x_c, r), \theta \in [0,1]$

$$\begin{aligned}\left\|\theta x_1+(1-\theta) x_2-x_c\right\| &= \left\|\theta(x_1-x_c)+(1-\theta)(x_2-x_c)\right\| \\ &\leq \theta\left\|x_1-x_c\right\|+(1-\theta)\left\|x_2-x_c\right\|=r\end{aligned}$$

- $\mathbf{S}_{+}^{n}=\left\{X \in \mathbf{S}^{n} \mid X \succeq 0\right\}$ denotes the set of positive semidefinite $n \times n$ matrices

$$X \in S_{+}^{n} \Leftrightarrow z^{\top} X z \geq 0, \text{ for all } z$$

Proof: $\forall x_1, x_2 \in \mathbf{S}_{+}^{n}, \theta \in [0,1], \forall z$

$$z^{T}(\theta X+(1-\theta) Y) z \geq 0$$

Operations that preserve convexity

a practical method for establishing convexity of a set C

- apply definition

$$x_1, x_2 \in C, 0 \leq \theta \leq 1 \Rightarrow \theta x_1 + (1-\theta)x_2 \in C$$

- show that C is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, ...) by operations that preserve convexity
  - intersection
  - affine functions
  - perspective function
  - linear-fractional functions

Affine function

Definition

suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine ($f(x) = Ax + b$ with $A \in \mathbb{R}^{m \times n}$ $b \in \mathbb{R}^m$)

- the image of convex set under f is convex

$$S \subseteq \mathbb{R}^n \text{ convex } \Rightarrow f(S) = \{f(x) \mid x \in S\} \text{ convex}$$

- the inverse image $f^{-1}(C)$ of a convex set under f is convex

$$C \subseteq \mathbb{R}^m \text{ convex } \Rightarrow f^{-1}(C) = \{x \in \mathbb{R}^n \mid f(x) \in C\} \text{ convex}$$

$S \subseteq \mathbb{R}^n$ convex.

$\forall x_1, x_2 \in S, \ \forall \theta \in [0,1]. \quad \theta x_1 + (1-\theta) x_2 \in S.$

$f(x_1) = Ax_1 + b \qquad f(x_2) = Ax_2 + b.$

$$f(\theta x_1 + (1-\theta) x_2) = A(\theta x_1 + (1-\theta) x_2) + b$$

$$= \theta A x_1 + \theta b + (1-\theta) A x_2 + (1-\theta) b$$

$$= \theta f(x_1) + (1-\theta) f(x_2).$$

$C \in R^m$, $\forall x_1, x_2 \in C$.

$\forall \theta \in [0,1]$ $\theta x_1 + (1-\theta)x_2 \in C$.

$f^{-1}(x_1) = y_1$ then $Ay_1 + b = x_1$

$f^{-1}(x_2) = y_2$. $Ay_2 + b = x_2$.

then $A(\theta y_1 + (1-\theta)y_2) + b \in C$.

$\therefore \theta y_1 + (1-\theta)y_2) \in f^{-1}(C)$.

Affine function

- the image of convex set under f is convex
  Proof:
  $$\begin{aligned}\theta f(x_1) + (1-\theta) f(x_2) &= \theta(Ax_1 + b) + (1-\theta)(Ax_2 + b) \\ &= A(\theta x_1 + (1-\theta)x_2) + b \\ &= f(\theta x_1 + (1-\theta)x_2)\end{aligned}$$

- the inverse image $f^{-1}(C)$ of a convex set under f is convex
  Proof:
  $x', y' \in C, i.e., x = f^{-1}(x'), y = f^{-1}(y'), \Rightarrow Ax + b = x', Ay + b = y',$
  Based on the convexity of C, we have $\theta x' + (1-\theta)y' \in C$, i.e.,
  $A(\theta x + (1-\theta)y) + b \in C, \Rightarrow \theta x + (1-\theta)y \in f^{-1}(C)$

Perspective function

Definition

Perspective function $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$

非意味着 P(x,t) 是 convex function；映射后的集合是 convex 的.

$$P(x,t) = x/t, \qquad \text{dom } P = \{(x,t) \mid t > 0\}$$

- images and inverse images of convex sets under perspective are convex

Proof: Previous idea: $\forall \hat{x} = (x,t), \hat{y} = (y,s) \in C$, prove $\theta P(x,t) + (1-\theta)P(y,s) \in P(C)$. Here, we prove it from the opposite side.

$$P(\theta\hat{x} + (1-\theta)\hat{y}) = \frac{\theta x + (1-\theta)y}{\theta t + (1-\theta)s} = \frac{\theta t}{\theta t + (1-\theta)s} \cdot \frac{x}{t} + \frac{(1-\theta)s}{\theta t + (1-\theta)s} \cdot \frac{y}{s}.$$

Let $\mu = \frac{\theta t}{\theta t + (1-\theta)s}$, we have
$\mu P(\hat{x}) + (1-\mu)P(\hat{y}) = P(\theta\hat{x} + (1-\theta)\hat{y}) \in P(C)$

" $\Rightarrow$ "

$\hat{x} = (x, t)$, $\hat{y} = (y, s)$ $\in S$ convex.

then $\theta\hat{x} + (1-\theta)\hat{y} \in S$. $P(\hat{x}) = \frac{x}{t}$ $P(\hat{y}) = \frac{y}{s}$.

$$P(\theta\hat{x} + (1-\theta)\hat{y}) = P((\theta x + (1-\theta)y, \theta t + (1-\theta)s)) = \frac{\theta x + (1-\theta)y}{\theta t + (1-\theta)s} = \overbrace{\frac{\theta t}{\theta t + (1-\theta)s}}^{\mu}\frac{x}{t} + \overbrace{\frac{(1-\theta)s}{\theta t + (1-\theta)s}}^{1-\mu}\frac{y}{s}$$

$$= \mu P(\hat{x}) + (1-\mu)P(\hat{y}) \in C.$$

" $\Leftarrow$ " C is convex.

$\forall \hat{x}_1, \hat{x}_2 \in f^{-1}(C)$ $\hat{x}_1 = (x_1, m)$ $\hat{x}_2 = (x_2, n)$.

then $P(\hat{x}_1) = \frac{x_1}{m}$, $P(\hat{x}_2) = \frac{x_2}{n}$ $\in C$.

$\forall \theta \in [0,1]$. $\theta P(\hat{x}_1) + (1-\theta)P(\hat{x}_2) \in C$.

$$= \frac{\theta x_1}{m} + \frac{(1-\theta)x_2}{n} \in C.$$

Suppose $\theta = \frac{\mu m}{\mu m + (1-\mu)n}$

$$\Rightarrow \frac{\mu x_1 + (1-\mu)x_2}{\mu m + (1-\mu)n} \in C.$$

$$P((\mu x_1 + (1-\mu)x_2, \mu m + (1-\mu)n))$$

$$= P(\mu\hat{x}_1 + (1-\mu)\hat{x}_2) \in C.$$

$\therefore$ $\mu\hat{x}_1 + (1-\mu)\hat{x}_2 \in P^{-1}(C)$.

Linear-fractional functions

Definition

linear-fractional function $f : \mathbb{R}^n \to \mathbb{R}^m$:

$$f(x) = \frac{Ax + b}{c^\top x + d}, \quad \operatorname{dom} f = \left\{ x \mid c^\top x + d > 0 \right\}$$

- Images and inverse images of convex sets under linear-fractional are convex
- A linear-fractional function is formed by composing the perspective function with an affine function. Suppose $g : \mathbf{R}^n \to \mathbf{R}^{m+1}$ is affine, *i.e.*

$$g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}$$

$S \in \mathbb{R}^n$ convex. to prove $f(S)$ is convex.

$\forall x_1, x_2 \in S$. $\forall \theta \in [0,1]$ $\theta x_1 + (1-\theta) x_2 \in S$.

then $f(x_1) = \dfrac{Ax_1 + b}{C^T x_1 + d}$

$f(x_2) = \dfrac{Ax_2 + b}{C^T x_2 + d}$

$\theta f(x_1) + (1-\theta) f(x_2) = \dfrac{\theta A x_1 + \theta b}{C^T x_1 + d} + \dfrac{(1-\theta) A x_2 + (1-\theta) b}{C_2^T x_2 + d.}$

$f(\theta x_1 + (1-\theta) x_2) = \dfrac{A(\theta x_1 + (1-\theta) x_2) + b}{C^T(\theta x_1 + (1-\theta) x_2) + d.}$

$\dfrac{\theta (Ax_1 + b)}{\cdots} + \dfrac{(1-\theta)(Ax_2 + b)}{\cdots}$

$a^T x = b$

Theorem (Separation hyperplane theorem)

if C and D are disjoint convex sets, then there exists $a \neq 0$ and b such that

$$a^\top x \leq b \text{ for } x \in C, \quad a^\top x \geq b \text{ for } x \in D$$

the hyperplane $\{x \mid a^\top x = b\}$ separates C and D

下界　欧几里德距离

Proof: $\|c - d\|_2 = dist(C, D) = \inf\{\|u - v\|_2, |u \in C, v \in D\}$, Let $f(x) = a^T x + b$, where $a = d - c$, $b = \frac{\|d\|^2 - \|c\|^2}{2}$.

$(d-c)^T x$　　$\frac{\|d\|^2 - \|c\|^2}{2}$

$d^T x - c^T x + \frac{\quad}{2}$

d^T

strongly convex function:

$\exists \mu > 0 \quad f(x) - \frac{\mu}{2}\|x\|_2^2$ is convex.

2-nd: $[\nabla^2 f(x) - \mu I] \succeq 0$.

特征根不能小于 μ.

$Hx = \lambda x$

$x^T H x = \lambda x^T x$

$x^T (H - \mu I) x = (\lambda - \mu) x^T x$

$\Rightarrow \lambda - \mu \geq 0$

$H - \mu I \succeq 0$.

Convex function

Convex function

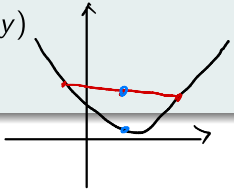
Definition

$f : \mathbb{R}^n \to \mathbb{R}$ is convex if dom f is a convex set and

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

Convex set

for all $x, y \in$ dom f, $0 \leq \theta \leq 1$



- f is concave if $-f$ is convex
- f is strictly convex if $\mathrm{dom} f$ is a convex set and

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$$

for all $x, y \in$ dom f, $x \neq y$, $0 < \theta < 1$

数归

"⇒" induction hypothesis assume $x_1, \dots, x_k$ convex combination $\in S$.

consider $x_1, \dots, x_k, x_{k+1}$ let λ_i $\sum \lambda_i = 1$ $\underline{\sum_{i=1}^{k+1} \lambda_i x_i \in S}$

$$\sum_{i=1}^{k+1} \lambda_i x_i = \lambda_{k+1} x_k + \sum_{i=1}^{k} \lambda_i x_i$$

$$= \lambda_{k+1} x_k + (1-\lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i$$

$$\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} = 1.$$

2. $\forall x, y \in \operatorname{conv}(S)$ $\begin{cases} x = \sum \alpha_i x_i & \alpha_i \geq 0 \quad \sum \alpha_i = 1. \\ y = \sum \beta_i y_i & \beta_i \geq 0 \quad \sum \beta_i = 1. \end{cases}$

$\forall \theta \in (0,1)$ $\underline{\theta x + (1-\theta) y} = \theta (\sum \alpha_i x_i) + (1-\theta)(\sum \beta_i y_i) = \sum (\theta \alpha_i) x_i + \sum (1-\theta) \beta_i y_i$.

$\in \operatorname{conv}(S)$.

$\sum \theta \alpha_i + \sum (1-\theta) \beta_i = 1$

$\forall x \in conv(S)$ $\quad x = \sum \lambda_i x_i \quad \lambda_i \geq 0 \quad \sum \lambda_i = 1 \quad x_i \in S$

$\because S \subseteq T \quad x_i \in T$.

since T is convex $\Rightarrow x = \sum_{i=1}^{n} \lambda_i x_i \in T$

$\Rightarrow conv(S) \subseteq T$.

set: 属于

function: [illegible] or 不[illegible]

$$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$$

- affine function $f(x) = a^T x + b$ convex and concave.

  Proof: $a^T(\theta x + (1-\theta)y) + b = \theta(a^Tx + b) + (1-\theta)(a^Ty + b)$

- norm: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_k |x_k|$

  Proof: use triangular inequality

$$\|\theta x + (1-\theta) y\| \le \|\theta x\| + \|(1-\theta) y\| = \theta\|x\| + (1-\theta)\|y\| \qquad \theta,\ 1-\theta > 0.$$

First-order condition

*前提

- f is differentiable if dom f is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$$

  exists at each $x \in \operatorname{dom} f$

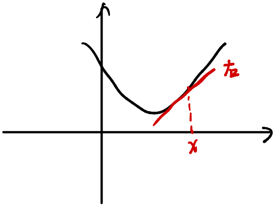
- First-order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \qquad \text{for all } x, y \in \operatorname{dom} f$$

  y.

  First-order approximation of f is global underestimator

整个函数为
在某个点切线
的上方.



Proof of First-order condition

f is convex $\Leftrightarrow f(y) \geq f(x) + \nabla f(x)^\top (y-x)$

Proof: "$\Rightarrow$"

For any $t \in (0,1)$,

$f((1-t)x+ty) \leq (1-t)f(x)+tf(y)$

$tf(y) \geq f(x+t(y-x)) - f(x) + f(x)$

$$
\begin{aligned}
f(x+t(y-x)) &= f((1-t)x+ty) \\
&\leq (1-t)f(x) + tf(y)
\end{aligned}
$$

Divide by t,

导数概念式.

$$
f(y) \geq f(x) + \frac{f(x+t(y-x)) - f(x)}{t}.
$$

The conclusion follows when $t \to 0$

"$\Leftarrow$"

$z = \theta x + (1-\theta)y$.

$$
\begin{aligned}
\theta f(x) &\geq \theta f(z) + \theta \nabla f(z)^T (x-z) \quad ① \\
(1-\theta) f(y) &\geq (1-\theta) f(z) + (1-\theta) \nabla f(z)^T (y-z) \quad ②
\end{aligned}
\tag{5}
$$

① + ② $\Rightarrow$ convex.

The first inequality multiple θ and plus the 2nd inequality with $(1-\theta)$ times.

second-order condition

- f is twice differentiable if $\mathbf{dom}\, f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, i, j = 1, \ldots, n$$

  exists at each $x \in \mathbf{dom}\, f$

- Second-order conditions: for twice differentiable f with convex domain f is convex if and only if

半正定.

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex

$f(x) = x^2$

$f'(x) = 2x$

$f''(x) = 2 \geq 0$

Examples

- quadratic function: $f(x) = (1/2)x^\top Px + q^\top x + r$ (with $P \in \mathbf{S}^n_+$)
  Proof:
  $$\nabla f(x) = Px + q, \quad \nabla^2 f(x) = P$$
  $\nabla f(x) = Px + q$
  $\nabla^2 f(x) = P$
  convex if $P \succeq 0$
  (增加 $\frac{1}{2}$，可去除系数)
- least-squares objective: $f(x) = \|Ax - b\|_2^2$
  Proof:
  $\nabla f = 2A^\top(Ax - b)$
  $\nabla^2 f = 2A^\top A$
  $$\nabla f(x) = 2A^\top(Ax - b), \quad \nabla^2 f(x) = 2A^\top A$$
- quadratic-over-linear: $f(x, y) = x^2/y, (y > 0)$
  Proof:
  $$\nabla^2 f(x, y) = \frac{2}{y^3} \begin{pmatrix} y \\ -x \end{pmatrix} \begin{pmatrix} y \\ -x \end{pmatrix}^\top \succeq 0$$
  convex for any $y > 0$
  $\nabla f = \begin{bmatrix} \frac{2x}{y} \\ -\frac{x^2}{y^2} \end{bmatrix}$ $\quad \nabla^2 f = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix}$.

: $f(x) = (1/2)x^\top P x + q^\top x + r$ (with $P \in \mathbf{S}_+^n$)

$\nabla f(x) = Px + q$

$\nabla^2 f(x) = P \succeq 0 \Rightarrow$ convex.

$f(x) = \|Ax - b\|_2^2$

$\nabla f(x) = 2A^T(Ax - b)$.

$\nabla^2 f(x) = 2A^TA \succeq 0$. $x^TA^TAx = (Ax)^T Ax$

$f(x, y) = x^2/y, (y > 0)$

$\nabla f(x,y) = \begin{bmatrix} \frac{2x}{y} & -\frac{x^2}{y^2} \end{bmatrix}$

$\nabla^2 f(x,y) = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3}\left(\begin{bmatrix} y \\ -x \end{bmatrix}\right)\left(\begin{bmatrix} y \\ -x \end{bmatrix}\right)^T \succeq 0$.

Restriction of a convex function to a line

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function $g : \mathbb{R} \to \mathbb{R}$

$$g(t) = f(x + tv), \quad \operatorname{dom} g = \{t \mid x + tv \in \operatorname{dom} f\}$$

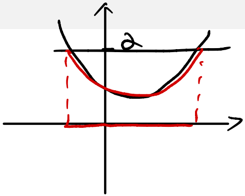
is convex (in t) for any $x \in \operatorname{dom} f, v \in \mathbb{R}^n$

Proof: "f convex $\Rightarrow$ g convex"

$$\begin{aligned} g(\theta t + (1-\theta)s) &= f(x + (\theta t + (1-\theta)s)v) \\ &= f(\theta(x + tv) + (1-\theta)(x + sv)) \\ &\leq \theta f(x + tv) + (1-\theta) f(x + sv) \end{aligned}$$

"g convex $\Rightarrow$ f convex"
Set $v = y - x$, hence $g(1) = f(y), g(0) = f(x)$

$$\begin{aligned} \theta f(x) + (1-\theta) f(y) &= \theta g(0) + (1-\theta) g(1) \\ &\geq g(\theta 0 + (1-\theta)1) = f(\theta x + (1-\theta)y) \end{aligned}$$



- α-sublevel set of $f : \mathbb{R}^n \to \mathbb{R}$:

$$C_\alpha = \{x \in \operatorname{dom} f \mid f(x) \le \alpha\}$$

sublevel sets of convex functions are convex (converse is false)

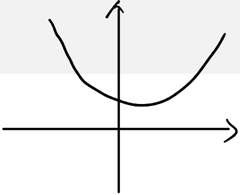
Proof: $\forall x_1, x_2 \in C_\alpha$

$$f(\underbrace{\theta x_1 + (1-\theta)x_2}_{\in C_\alpha}) \le \theta f(x_1) + (1-\theta) f(x_2) \le \alpha$$

$\alpha\theta + (1-\theta)\alpha$.

$\forall x_1, x_2 \in C_\alpha$

Epigraph



- epigraph of $f : \mathbb{R}^n \to \mathbb{R}$:

$$\text{epi } f = \{(x,t) \in \mathbb{R}^{n+1} \mid x \in \text{dom } f, f(x) \le t\}$$

f is convex if and only if epi f is a convex set

Proof: "$\Rightarrow$"

$\forall (x,t),(y,s) \in \text{epi}f$

$\Rightarrow \ \forall (x_1, t_1), (x_2, t_2) \in \text{epi}f.$

$f(\theta x_1 + (1-\theta) x_2) \le \theta f(x_1) + (1-\theta) f(x_2) \le \theta t_1 + (1-\theta)$

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y) \le \theta t + (1-\theta)s,$$

so $\theta(x,t) + (1-\theta)(y,s) \in \text{epi}f$.

"$\Leftarrow$"

$\forall (x, f(x)), (y, f(y)) \in \text{epi}f$ then

$(\theta x + (1-\theta)y, \theta f(x) + (1-\theta)f(y)) \in \text{epi}f$.

$\Rightarrow$: $f(\theta x+(1-\theta)y) \le \theta f(x)+(1-\theta)f(y) \le \theta t+(1-\theta)t.$

and $f(x) \le t.$ x, y

$\Leftarrow$: $\mathrm{epi}\, f$ is convex. $(x,t), (y,t) \in \mathrm{epi}\, f.$

$(\theta x+(1-\theta)y, t) \in \mathrm{epi}\, f.$

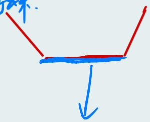
$f(\theta x.$

Convexity and global minimum

Theorem

Theorem 1. Consider an optimization problem

特殊.



$$\min f(x) \text{ s.t. } x \in \Omega,$$

where f is a convex function and Ω is a convex set. Then, any local minimum is also a global minimum.

Proof: Let $\bar{x}$ be a local minimum, 反证.

区域半径.

$$\Rightarrow \bar{x} \in \Omega \text{ and } \underline{\exists \epsilon > 0} \text{ s.t. } f(\bar{x}) \le f(x), \forall x \in B(\bar{x}, \epsilon).$$

Suppose for the sake of contradiction that $\exists z \in \Omega$ with

$$f(z) < f(\bar{x}).$$

Because of convexity of Ω, we have

$$\lambda \bar{x} + (1 - \lambda) z \in \Omega, \forall \lambda \in [0, 1].$$

Convexity and global minimum

By convexity of f, we have

$$\begin{aligned} f(\lambda\bar{x} + (1-\lambda)z) &\leq \lambda f(\bar{x}) + (1-\lambda)f(z) \\ &< \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x}). \end{aligned}$$

But, as $\lambda \to 1, (\lambda\bar{x} + (1-\lambda)z) \to \bar{x}$ and the previous inequality contradicts local optimality of $\bar{x}$.

This theorem, as simple as it is, is one of the most important theorems in convex analysis.

①. separation hyperplane the

$C \cap D = \{\phi\}$ C, D convex

$\Rightarrow \exists\, a \neq 0$ $a^T x \le b$ for $x \in C$

$a^T x \ge b$ for $x \in D$.

梯度的替代品.

Subgradient

(non-differentiable convex function)

Basic inequality

The basic inequality for differentiable convex functions:

first-order condition

$$f(y) \geq f(x) + \underbrace{\nabla f(x)}_{\text{at } x.}{}^T (y - x) \quad \text{for all } y \in \text{dom } f$$

- The first-order approximation of f at x is a global lower bound
- $\nabla f(x)$ defines non-vertical supporting hyperplane to epigraph of f at $(x, f(x))$:

$a \qquad z \qquad b. \qquad f(y) \leq t.$

$$\begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \text{ for all } (y, t) \in \text{ epi } f$$

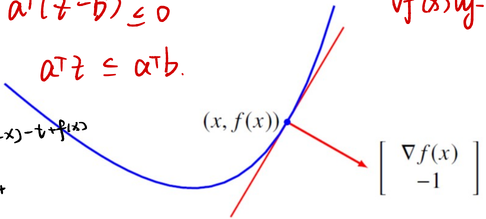
$\Rightarrow \nabla f^T(x)(y-x) - (t - f(x)) \leq 0$

$a^T(z-b) \leq 0$

$a^T z \leq a^T b.$

$[\nabla f(x) \; -1] \begin{bmatrix} y-x \\ t-f(x) \end{bmatrix} = \nabla f(x)(y-x) - t + f(x)$

$f(x)+$



Subgradient

- g is a subgradient of a convex function f at $x \in \mathrm{dom}\, f$ if

1xn n×1.

$$f(y) \geq f(x) + g^T(y - x) \text{ for all } y \in \mathrm{dom}\, f$$

- the subdifferential $\partial f(x)$ of f at x is the set of all subgradients:

$$\partial f(x) = \left\{ g | g^T(y - x) \leq f(y) - f(x), \forall y \in \mathrm{dom}\, f \right\}$$

$$= \bigcap_{z \in \mathrm{dom}\, f} \left\{ g \mid f(z) \geq f(x) + g^T(z - x) \right\}$$

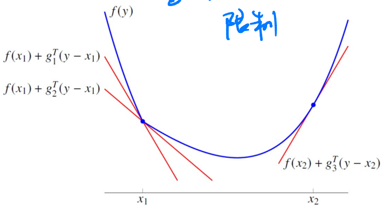
! 对z没有任何限制

$\{g \mid a^T g \leq b\}$

convex

$a: z - x$

$b: f(z) - f(x)$



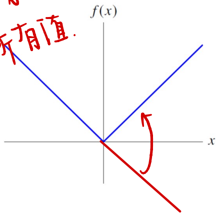
g_1, g_2 are subgradients at x_1; g_3 is a subgradient at x_2 and the subdifferential is the intersection of an infinite set of halfspaces

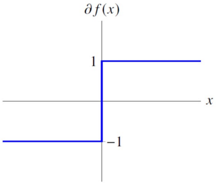
- Absolute value $f(x) = |x|$

$f(y) \geq f(x) + g^T(y-x)$

左导到右导

逆时针所有值.





$1\times n \; n\times 1$

yy^T: $n\times 1 \; 1\times n$

$\|y\|_2 \geq g^T y$

$f(y) \geq g^T y.$

$y^T y \geq g^T(y^T y).$

- Euclidean norm $f(x) = \|x\|_2$

$$\partial f(x) = \left\{ \frac{1}{\|x\|_2} x \right\} \text{ if } x \neq 0, \partial f(x) = \{g \mid \|g\|_2 \leq 1\} \text{ if } x = 0$$

$\|x\|_2 = \sqrt{x^T x}.$

suppose $u = x^T x.$

$(\sqrt{u})' = \frac{1}{2\sqrt{u}} \cdot 2x. = \frac{1}{\|x\|_2} x.$

Subgradient calculus

- Weak subgradient calculus: rules for finding one subgradient
  - sufficient for most nondifferentiable convex optimization algorithms
  - if you can evaluate $f(x)$, you can usually compute a subgradient
- Strong subgradient calculus: rules for finding $\partial f(x)$ (all subgradients)
  - some algorithms, optimality conditions, etc., need entire subdifferential
  - can be quite complicated

we will assume that $x \in \text{int dom } f$

Basic rules

$\not\!\!$ exercise

- Differentiable functions: $\partial f(x) = \{\nabla f(x)\}$ if f is differentiable at x
- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$ $\quad af(y) \geq af(x) + ag^T(y-x)$
- Addition: $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$
- Affine composition: if $g(x) = f(Ax + b)$, then

chain rule

$$\partial g(x) = A^T \partial f(Ax + b)$$

- Finite pointwise maximum: if $f(x) = \max_{i=1,\ldots,m} f_i(x)$, define $I(x) = \{i \mid f_i(x) = f(x)\}$, the 'active' functions at x.
  - Weak result to compute a subgradient at x, choose any $k \in I(x)$, any subgradient of f_k at x
  - Strong result

$$\partial f(x) = \operatorname{conv} \bigcup_{i \in I(x)} \partial f_i(x)$$

- the convex hull of the union of subdifferentials of 'active' functions at x

- if f_i 's are differentiable, $\partial f(x) = \operatorname{conv}\{\nabla f_i(x) | i \in I(x)\}$

<1> 证明: ① $\{\nabla f(x)\} \subseteq \partial f(x)$.

1 st. order condition of convex function : $f(y) \geq f(x) + \nabla f(x)^T (y-x)$

② $\partial f(x) \subseteq \{\nabla f(x)\}$

$\forall v \in \partial f(x_0)$ Give x_0 (固定某点)

$\forall x \quad f(x) \geq f(x_0) + \langle v, x-x_0 \rangle$

set $x = x_0 + \lambda z$ $\lambda > 0$ $\forall z$.

$$\frac{f(x_0+\lambda z) - f(x_0)}{\lambda} \geq \langle v, z \rangle$$

λ 作变量, $\lambda \to 0$ 即 $f(x)$ 在 z 方向上的方向导数.

$g(\lambda) := f(x_0 + \lambda z)$ $\lambda \to 0$ $\Rightarrow \nabla f(x_0)^T z \geq v^T z$.

$\Rightarrow \langle \nabla f(x_0) - v, z \rangle \geq 0 \quad \forall z$.

$z = -\nabla f(x_0) + v$

$\Rightarrow -\|\nabla f(x_0) - v\|_2^2 \geq 0$

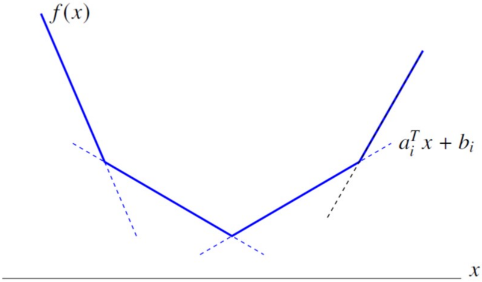
Example

- Piecewise-linear function: $f(x) = \max_{i=1,\ldots,m} \left(a_i^T x + b_i\right)$
- the subdifferential at x is a polyhedron

$$\partial f(x) = \operatorname{conv}\left\{a_i | i \in I(x)\right\}$$

with $I(x) = \left\{i \quad | \quad a_i^T x + b_i = f(x)\right\}$

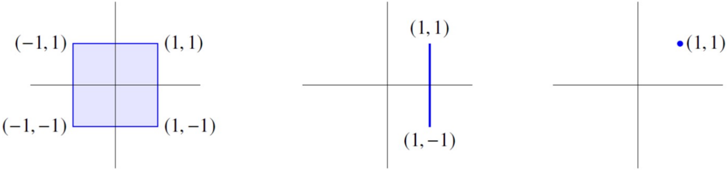
x_0 $\operatorname{conv}\{a_i | i \in I(x)\}$.



Example

- L_1 norm: $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^n} s^T x$
- the subdifferential is a product of intervals

$$\partial f(x) = J_1 \times \cdots \times J_n, \quad J_k = \begin{cases} [-1,1] & x_k = 0 \\ \{1\} & x_k > 0 \\ \{-1\} & x_k < 0 \end{cases}$$



$\partial f(0,0) = [-1,1] \times [-1,1]$ $\quad \partial f(1,0) = \{1\} \times [-1,1]$ $\quad \partial f(1,1) = \{(1,1)\}$

Basic rules

- Minimization function: $f(x) = \inf_y h(x, y)$
- Weak result: to find a subgradient at $\hat{x}$,
  - find $\hat{y}$ that minimizes $h(\hat{x}, y)$ (assuming minimum is attained)
  - find subgradient $(g, 0) \in \partial h(\hat{x}, \hat{y})$

Proof: for all x, y,

$$\begin{aligned} h(x, y) &\geq h(\hat{x}, \hat{y}) + g^T(x - \hat{x}) + 0^T(y - \hat{y}) \\ &= f(\hat{x}) + g^T(x - \hat{x}) \end{aligned}$$

therefore

$$f(x) = \inf_y h(x, y) \geq f(\hat{x}) + g^T(x - \hat{x})$$

Example

- Problem: explain how to find a subgradient of

$$f(x) = \inf_{y \in C} \|x - y\|_2$$

  where C is a closed convex set

- Solution: to find a subgradient at $\hat{x}$,
  - if $f(\hat{x}) = 0$ (that is, $\hat{x} \in C$), take $g = 0$
  - if $f(\hat{x}) > 0$, find projection $\hat{y} = P(\hat{x})$ on C and take

$$g = \frac{1}{\|\hat{y} - \hat{x}\|_2}(\hat{x} - \hat{y}) = \frac{1}{\|\hat{x} - P(\hat{x})\|_2}(\hat{x} - P(\hat{x}))$$

Basic rule

- Composition function:
  $f(x) = h(f_1(x), \ldots, f_k(x))$, $\quad h$ convex and nondecreasing, f_i convex
- Weak result: to find a subgradient at $\hat{x}$,
  - find $z \in \partial h(f_1(\hat{x}), \ldots, f_k(\hat{x}))$ and $g_i \in \partial f_i(\hat{x})$
  - then $g = z_1 g_1 + \cdots + z_k g_k \in \partial f(\hat{x})$ reduces to standard formula for differentiable h, f_i

Proof:

$$\begin{aligned}
f(x) &\geq h\left(f_1(\hat{x}) + g_1^T(x-\hat{x}), \ldots, f_k(\hat{x}) + g_k^T(x-\hat{x})\right) \\
&\geq h(f_1(\hat{x}), \ldots, f_k(\hat{x})) + z^T\left(g_1^T(x-\hat{x}), \ldots, g_k^T(x-\hat{x})\right) \\
&= h(f_1(\hat{x}), \ldots, f_k(\hat{x})) + (z_1 g_1 + \cdots + z_k g_k)^T(x-\hat{x}) \\
&= f(\hat{x}) + g^T(x-\hat{x})
\end{aligned}$$

Some properties

Theorem

Let f be convex with a nonempty $\mathbf{dom}\, f$. Then:

(a) $\partial f(x)$ *is a closed convex set (possibly empty)*

(b) *the subdifferential of a convex function is a monotone operator:*

$$(u-v)^T(x-y) \geq 0 \quad \text{for all } x, y, u \in \partial f(x), v \in \partial f(y)$$

Proof:
(a) this follows from the definition: $\partial f(x)$ is an intersection of halfspaces
(c) by definition

$$f(y) \geq f(x) + u^T(y-x), \quad f(x) \geq f(y) + v^T(x-y)$$

combining the two inequalities shows monotonicity

Summary

- Convex set:
  - example:hyperplane, norm ball...
  - difference between convex hull, affine set
  - operations that preserve convexity: affine function, ...
  - separation hyperplane theorem
- Convex function:
  - example:affine function...
  - 1st, 2nd order condition
  - sublevel set and epigraph
- Subgradient:
  - motivation: extend the conceipt of gradient in non-differential but convex functions
  - rules on its calculus
  - monotone property

T2.

1. (f_i convex , g convex , $g \uparrow$)

$\forall x, y, \quad \lambda \in [0,1]$

$h(\lambda x + (1-\lambda) y) = g(f(\lambda x + (1-\lambda) y)) = g(f_1(\lambda x + (1-\lambda) y), f_2(\lambda x + (1-\lambda) y) \dots)$

$f_i(\lambda x + (1-\lambda) y) \le \lambda f_i(x) + (1-\lambda) f_i(y)$

$\le g(\lambda f_1(x) + (1-\lambda) f_1(y), \lambda f_2(x) + (1-\lambda) f_2(y) \dots)$

2. $f(x) = \|Ax - b\|_2 + \|x\|_2$

1). $Ax - b \ne 0$, $x \ne 0$ $\quad \frac{A^T(Ax-b)}{\|Ax-b\|_2} + \frac{x}{\|x\|_2}$

2). $Ax - b \ne 0$, $x = 0$ $\quad \frac{A^T(Ax-b)}{\|Ax-b\|_2}$

3). $Ax - b = 0$, $x \ne 0$ $\quad \frac{x}{\|x\|_2}$

4). $Ax - b = 0$, $x = 0$ $\quad 0$