凸函数(续)

王尧

西安交通大学智能决策与机器学习中心
(Email: yao.s.wang@gmail.com)

2022. 4

回顾：凸函数的定义

定义 (凸函数)

$f:\mathbb{R}^n \to \mathbb{R}$ 为适当函数，如果$\mathbf{dom}\, f$ 是凸集，且

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

对所有$x, y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$ 都成立，则称f 是凸函数

- 若f 是凸函数，则$-f$ 是凹函数

回顾：凸函数的定义

Important modifiers:

- Strictly convex: $f\big(tx + (1-t)y\big) < tf(x) + (1-t)f(y)$ for $x \neq y$ and $0 < t < 1$. In words, f is convex and has greater curvature than a linear function

- Strongly convex with parameter $m > 0$: $f - \frac{m}{2}\|x\|_2^2$ is convex. In words, f is at least as convex as a quadratic function

Note: strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex

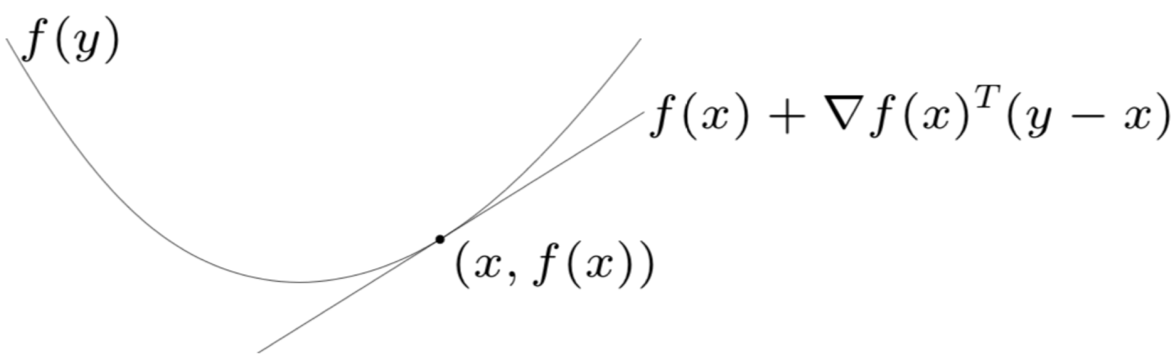
(Analogously for concave functions)

回顾：凸函数的一阶条件

定理

一阶条件：对于定义在凸集上的可微函数f，f 是凸函数当且仅当

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y \in \mathrm{dom}\, f$$



几何直观：f的一阶逼近始终在f的图像下方

梯度的单调性

定理

设f为可微函数，则f为凸函数当且仅当$\mathrm{dom}f$为凸集且∇f为**单调映射**，

$$(\nabla f(x) - \nabla f(y))^{\mathrm{T}}(x - y) \geq 0, \quad \forall\, x, y \in \mathrm{dom} f.$$

Proof.

必要性：若f可微且为凸函数，根据一阶条件，我们有

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x), \\ f(x) &\geq f(y) + \nabla f(y)^{\mathrm{T}}(x - y). \end{aligned}$$

将两式不等号左右两边相加即可得到结论.

梯度的单调性

Proof.

充分性：若∇f为单调映射，构造一元辅助函数

$$g(t)=f(x+t(y-x)),\quad g'(t)=\nabla f(x+t(y-x))^{\mathrm{T}}(y-x)$$

由∇f的单调性可知$g'(t)\geq g'(0),\forall\, t\geq 0$. 因此

$$\begin{aligned}f(y)=g(1)&=g(0)+\int_0^1 g'(t)\mathrm{d}t\\&\geq g(0)+g'(0)=f(x)+\nabla f(x)^{\mathrm{T}}(y-x).\end{aligned}$$

□

向量范数(norm)

norm: a function $\|\cdot\|$ that satisfies

- $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$ 正定性
- $\|tx\| = |t| \, \|x\|$ for $t \in \mathbf{R}$ 齐次性
- $\|x + y\| \leq \|x\| + \|y\|$ 三角不等式

notation: $\|\cdot\|$ is general (unspecified) norm; $\|\cdot\|_{\mathrm{symb}}$ is particular norm

example

Euclidean norm (or 2-norm): $\|\mathbf{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$

Cauchy-Schwarz不等式

定理

设$a, b \in \mathbb{R}^n$, 则

$$|a^{\mathrm{T}}b| \leqslant \|a\|_2 \, \|b\|_2 \, ,$$

且等号成立的条件是a与b线性相关.

注：不等式在估值和放缩等不同的情形下具有重要的作用

向量范数(norm)

norm: a function $\|\cdot\|$ that satisfies

- $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$ 正定性
- $\|tx\| = |t|\,\|x\|$ for $t \in \mathbf{R}$ 齐次性
- $\|x + y\| \leq \|x\| + \|y\|$ 三角不等式

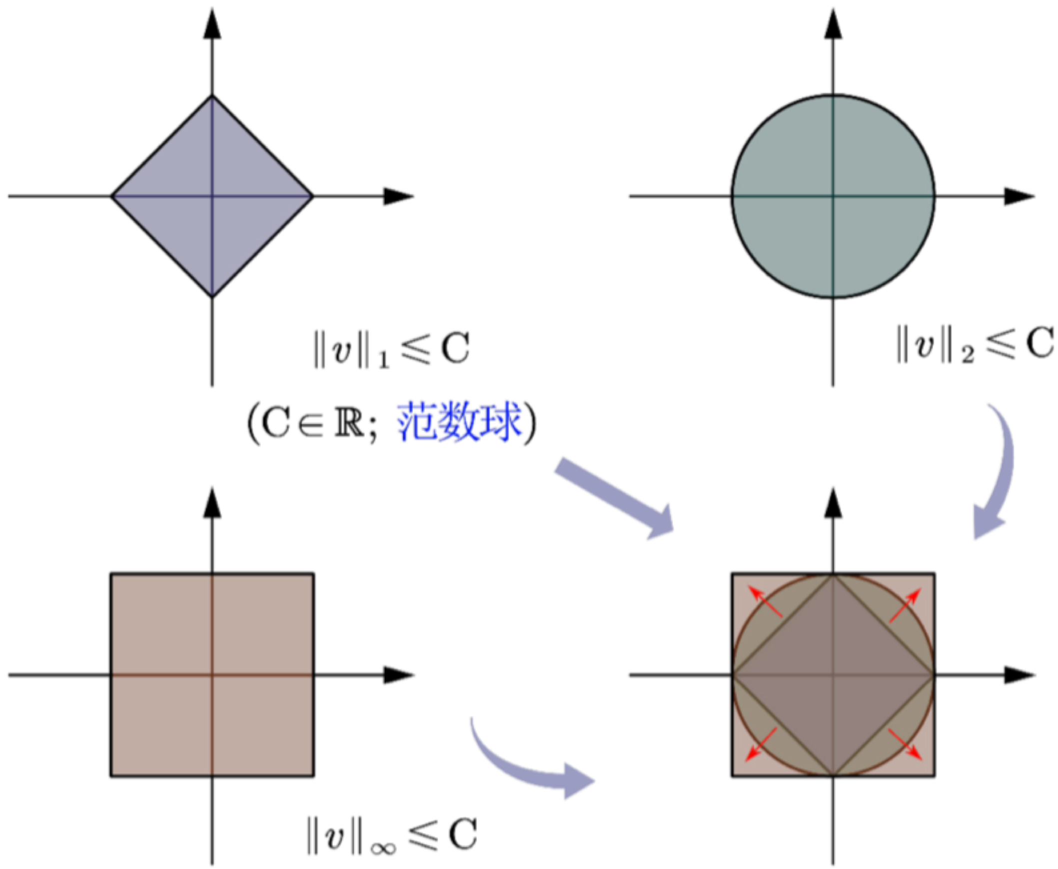
notation: $\|\cdot\|$ is general (unspecified) norm; $\|\cdot\|_{\mathrm{symb}}$ is particular norm

example

Euclidean norm (or 2-norm): $\|\mathbf{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$

课后作业题：证明无穷范数$\|\mathbf{x}\|_\infty := \max_i |x_i|$满足范数的三个性质

范数球(norm ball)



课后作业题：证明向量范数均是凸函数

凸函数限制在直线上

凸函数的一个最基本的判定方式是：先将其限制在任意直线上，然后判断对应的一维函数是否是凸的.

定理

$f:\mathbb{R}^n \to \mathbb{R}$ 是凸函数，当且仅当对每个$x \in \mathbf{dom}\, f, v \in \mathbb{R}^n$，函数$g:\mathbb{R} \to \mathbb{R}$,

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

是关于t的凸函数

example. $f : \mathbf{S}^n \to \mathbf{R}$ with $f(X) = \log\det X$, $\mathbf{dom}\, f = \mathbf{S}^n_{++}$

$$\begin{aligned} g(t) = \log\det(X+tV) &= \log\det X + \log\det(I + tX^{-1/2}VX^{-1/2}) \\ &= \log\det X + \sum_{i=1}^n \log(1+t\lambda_i) \end{aligned}$$

where λ_i are the eigenvalues of $X^{-1/2}VX^{-1/2}$

g is concave in t (for any choice of $X \succ 0$, V); hence f is concave

海瑟(Hessian)矩阵

定义 (海瑟矩阵)

如果函数$f(x):\mathbb{R}^n\to\mathbb{R}$在点$x$处的二阶偏导数$\frac{\partial^2 f(x)}{\partial x_i\partial x_j}$ $i,j=1,2,\cdots,n$都存在，则

$$\nabla^2 f(x)=\begin{bmatrix}\frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1\partial x_2} & \frac{\partial^2 f(x)}{\partial x_1\partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_1\partial x_n}\\ \frac{\partial^2 f(x)}{\partial x_2\partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \frac{\partial^2 f(x)}{\partial x_2\partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_2\partial x_n}\\ \vdots & \vdots & \vdots & & \vdots\\ \frac{\partial^2 f(x)}{\partial x_n\partial x_1} & \frac{\partial^2 f(x)}{\partial x_n\partial x_2} & \frac{\partial^2 f(x)}{\partial x_n\partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2}\end{bmatrix}$$

称为f 在点x 处的海瑟矩阵.

凸函数的二阶条件

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

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

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

2nd-order conditions: for twice differentiable f with convex domain

- f is convex if and only if

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

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

M positive semi-definite $\iff$ $x^\top M x \geq 0$ for all $x \in \mathbb{R}^n$

举例

quadratic function: $f(x) = (1/2)x^T Px + q^T x + r$ (with $P \in \mathbb{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$

least-squares objective: $f(x) = \|Ax - b\|_2^2$

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

convex (for any A)

举例

quadratic function: $f(x) = (1/2)x^T Px + q^T x + r$ (with $P \in \mathbb{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$

least-squares objective: $f(x) = \|Ax - b\|_2^2$

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

convex (for any A)

课后作业：证明 **log-sum-exp**: $f(x) = \log \sum_{k=1}^n \exp x_k$ is convex

保凸运算

practical methods for establishing convexity of a function

1. verify definition (often simplified by restricting to a line)

2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$

3. show that f is obtained from simple convex functions by operations that preserve convexity

非负加权和与仿射函数的复合

非负数乘: 若f 是凸函数，则αf 是凸函数，其中$\alpha \geq 0$.

求和: 若f_1, f_2 是凸函数，则$f_1 + f_2$ 是凸函数.

与仿射函数的复合: 若f 是凸函数，则$f(Ax+b)$ 是凸函数.

例子

- 线性不等式的对数障碍函数

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x), \quad \mathbf{dom}\, f = \{x | a_i^T x < b_i, i = 1, \ldots, m\}$$

- 仿射函数的（任意）范数：$f(x) = \|Ax + b\|$

逐点取上界

若对每个$y \in \mathcal{A}$，$f(x,y)$是关于x的凸函数，则

$$g(x) = \sup_{y \in \mathcal{A}} f(x,y)$$

是凸函数

例子

- 集合C的支撑函数：$S_C(x) = \sup_{y \in C} y^T x$ 是凸函数
- 集合C点到给定点x的最远距离：

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

复合函数

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 $n = 1$:

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

examples

- $\exp g(x)$ is convex if g is convex
- $1/g(x)$ is convex if g is concave and positive

可参看文再文版教材2.5.3中更多保凸运算的介绍

一些优化术语

Reminder: a convex optimization problem (or program) is

$$\begin{aligned} \min_{x \in D} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

where f and g_i, $i = 1, \ldots, m$ are all convex, and the optimization domain is $D = \mathrm{dom}(f) \cap \bigcap_{i=1}^m \mathrm{dom}(g_i)$ (often we do not write D)

- f is called criterion or objective function
- g_i is called inequality constraint function
- If $x \in D$, $g_i(x) \leq 0$, $i = 1, \ldots, m$, and $Ax = b$ then x is called a feasible point
- The minimum of $f(x)$ over all feasible points x is called the optimal value, written $f^\star$

一些优化术语

- If x is feasible and $f(x) = f^\star$, then x is called optimal; also called a solution, or a minimizer
- If x is feasible and $f(x) \le f^\star + \epsilon$, then x is called ϵ-suboptimal
- If x is feasible and $g_i(x) = 0$, then we say g_i is active at x
- Convex minimization can be reposed as concave maximization

$$
\begin{array}{ll}
\min\limits_{x} & f(x) \\
\text{subject to} & g_i(x) \le 0, \\
& \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
\quad \Longleftrightarrow \quad
\begin{array}{ll}
\max\limits_{x} & -f(x) \\
\text{subject to} & g_i(x) \le 0, \\
& \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
$$

Both are called convex optimization problems

凸优化问题的解集

Let X_{opt} be the set of all solutions of convex problem, written

$$\begin{array}{rcll} X_{\text{opt}} & = & \text{argmin} & f(x) \\ & & \text{subject to} & g_i(x) \leq 0, \ i = 1, \ldots, m \\ & & & Ax = b \end{array}$$

Key property 1: X_{opt} is a convex set

Proof: use definitions. If x, y are solutions, then for $0 \leq t \leq 1$,

- $g_i(tx + (1-t)y) \leq t g_i(x) + (1-t) g_i(y) \leq 0$
- $A(tx + (1-t)y) = tAx + (1-t)Ay = b$
- $f(tx + (1-t)y) \leq t f(x) + (1-t) f(y) = f^\star$

Therefore $tx + (1-t)y$ is also a solution

Key property 2: if f is strictly convex, then solution is unique, i.e., X_{opt} contains one element (课后作业，可用反证法证明)

Thank you for your attentions !