Convex Optimization Theory and Applications

Topic 2 - Convex Functions

Li Li

Department of Automation
Tsinghua University

Fall, 2009-2021.

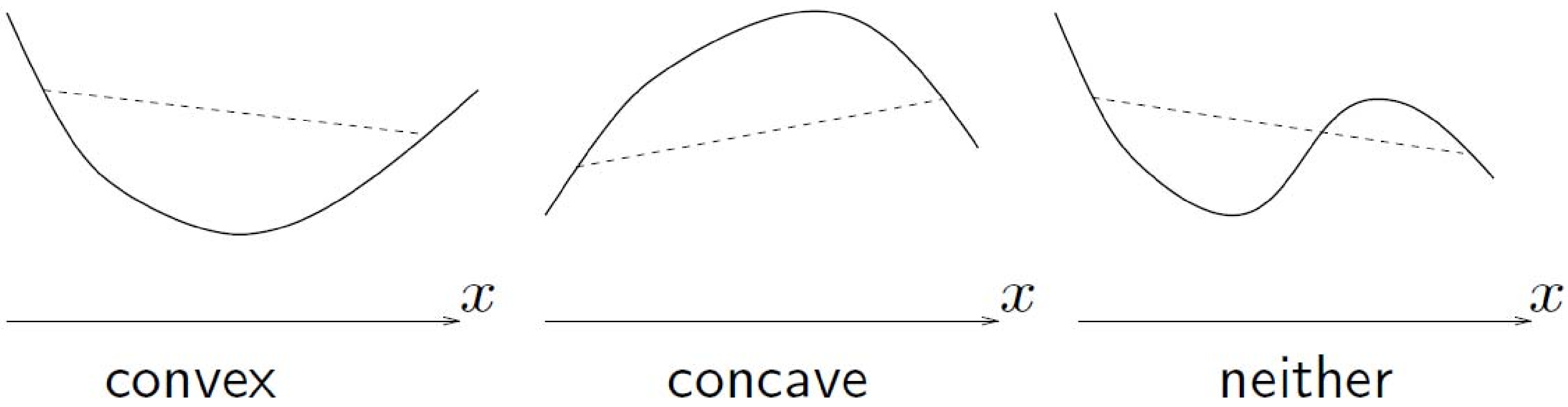
2.0. Outline

2.1. Definitions and Examples

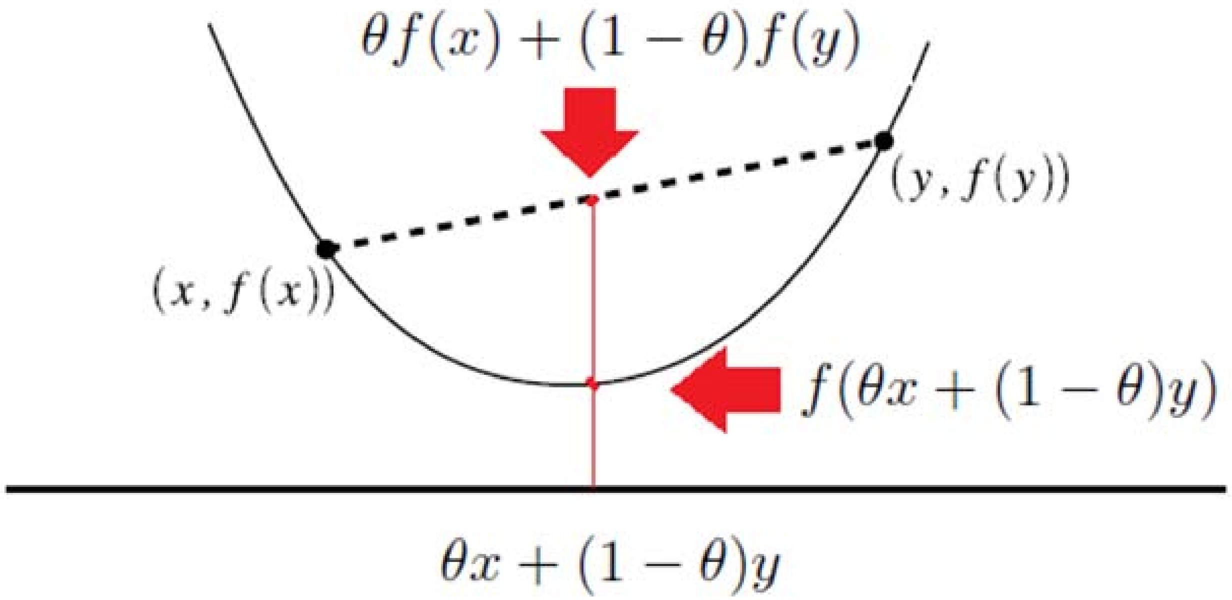
凸函数 f is a real continuous function $f:\Omega\to R$ that is convex, if and only if, for any $p_1\in[0,1]$, we have

$$p_1 f(x_1)+(1-p_1)f(x_2)\geq f\left(p_1x_1+(1-p_1)x_2\right) \qquad (2.1)$$

凹函数 f is concave if $-f$ is convex



2.1. Definitions and Examples



无论是凸函数，还是凹函数，一定要求定义域为凸集

2.1. Definitions and Examples

Jensen's Inequality If p_1 , ..., p_n are positive numbers which sum to 1 and f is a real continuous function that is concave up, then

$$\sum_{i=1}^{n} p_i f(x_i) \geq f\left(\sum_{i=1}^{n} p_i x_i\right) \tag{2.2}$$

How to prove?

2.1. Definitions and Examples

What are the convex functions to prove these inequalities?

$$e^{\frac{x+y}{2}} \le \frac{1}{2}(e^x + e^y) \tag{2.3}$$

$$\frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}} \le \sqrt[n]{a_1 a_2 \cdots a_n} \le \frac{a_1 + a_2 + \cdots + a_n}{n} \tag{2.4}$$

$$\left| \frac{x_1 + x_2 + \cdots + x_n}{n} \right| \le \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}} \tag{2.5}$$

2.1. Definitions and Examples

Consider the concave function in R^+

$$f(x) = \ln x \tag{2.6}$$

We have

$$\ln \sqrt[n]{\prod_{k=1}^{n} x_k} = \frac{1}{n}\sum_{k=1}^{n} \ln x_k = \frac{1}{n}\sum_{k=1}^{n} f(x_k) \le f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) = \ln\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) \tag{2.7}$$

Thus, we prove the right hand side of (2.4).

2.1. Definitions and Examples

Please prove that for a triangle ΔABC, we have

$$\sin A + \sin B + \sin C \leq \frac{3\sqrt{3}}{2} \tag{2.8}$$

2.1. Definitions and Examples

Consider the concave function $f(x)$ in $(0, \pi)$

$$f(x) = \sin x \tag{2.9}$$

We have

$$\frac{\sin A + \sin B + \sin C}{3} = \frac{f(A) + f(B) + f(C)}{3} \leq f\left(\frac{A+B+C}{3}\right) = \sin\frac{\pi}{3} = \frac{\sqrt{3}}{2} \tag{2.10}$$

What is the lower bound of $\sin A + \sin B + \sin C$?

2.1. Definitions and Examples

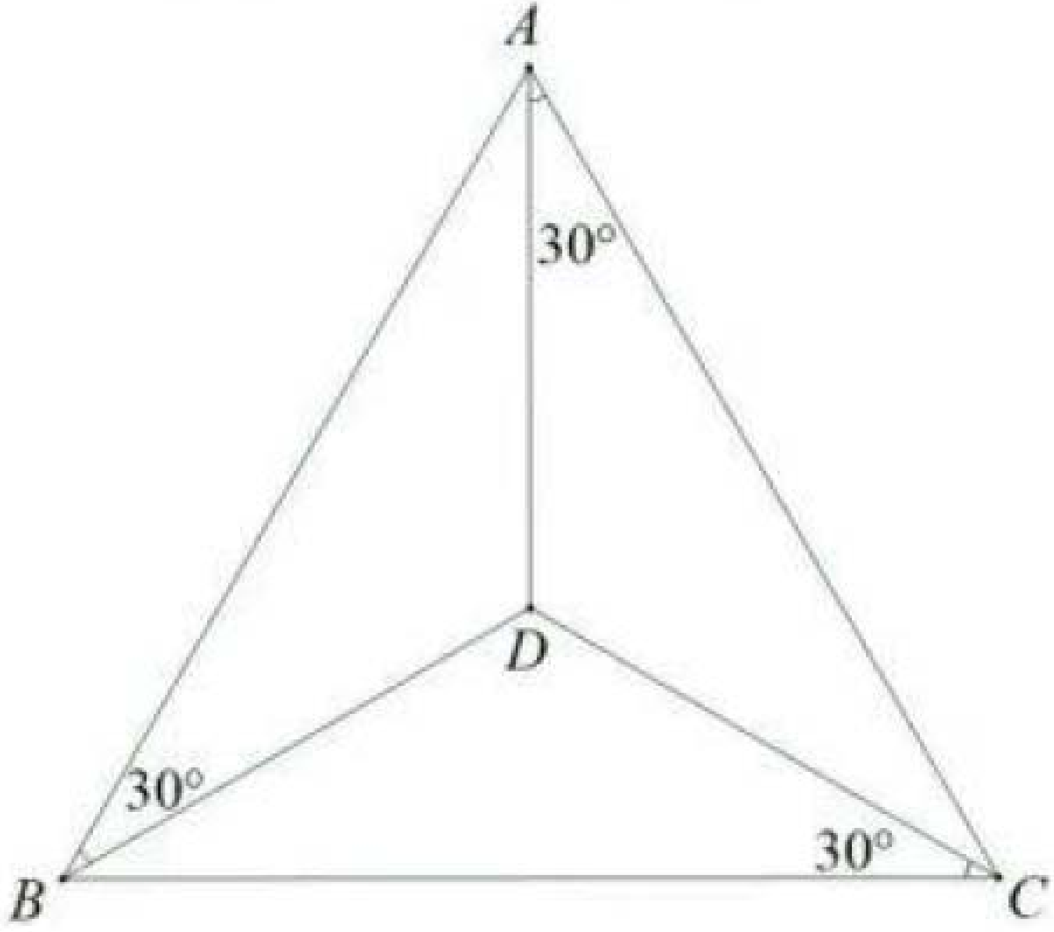
Clearly, we have $\sin A > 0$, $\sin B > 0$, $\sin C > 0$, Thus, we have

$$\begin{aligned} &\sin A + \sin B + \sin C \\ &= \sin A + \sin B + \sin(\pi - (A + B)) \\ &= \sin A + \sin B + \sin(A + B) \end{aligned}$$

Can you get the result now?

2.1. Definitions and Examples

证明：ΔABC是等边三角形



2.1. Definitions and Examples

设顶点在 A, B, C 处、角度尚未获知的那些角分别为 x, y, z. 依角元 Ceva 定理，有

$$\sin x \sin y \sin z = \frac{1}{8}, \text{ 其中 } x + y + z = 90^\circ.$$

再依 Jensen 不等式和均值不等式，又有

$$\frac{3}{2} = 3\sin\left(\frac{x+y+z}{3}\right) \geq \sin x + \sin y + \sin z \geq 3\sqrt[3]{\sin x \sin y \sin z} = \frac{3}{2},$$

于是

$$\sin x + \sin y + \sin z = \frac{3}{2},$$

这表明前述不等式需两端取等，显然必须要 $x = y = z$, 即证。

2.1. Definitions and Examples

设 $a,b,c>0$，$a+b+c=3$，请证明 $\frac{a}{1+b^2}+\frac{b}{1+c^2}+\frac{c}{1+a^2}\geq\frac{3}{2}$（Bulgaria TST 2003）

2.1. Definitions and Examples

如果直接平均不等式，不行

$$\frac{a}{1+b^2}+\frac{b}{1+c^2}+\frac{c}{1+a^2}\le\frac{a}{2b}+\frac{b}{2c}+\frac{c}{2a}\ge\frac{3}{2}$$

但是注意到

$$\frac{a}{1+b^2}=a-\frac{ab^2}{1+b^2}\ge a-\frac{ab^2}{2b}=a-\frac{ab}{2}$$

$$3(ab+bc+ca)\le(a+b+c)^2$$

我们可以得到

$$\frac{a}{1+b^2}+\frac{b}{1+c^2}+\frac{c}{1+a^2}\ge a-\frac{ab}{2}+b-\frac{bc}{2}+c-\frac{ca}{2}\ge\frac{3}{2}$$

2.1. Definitions and Examples

If $a+b+c+d=2$, $a^2+2b^2+3c^2+6d^2=2$, please derive the value of a.

2.1. Definitions and Examples

Since

$$\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{6}\right)\left(a^2+2b^2+3c^2+6d^2\right)\geq\left(a+b+c+d\right)^2=4$$

We have $a^2=4b^2=9c^2=36d^2=1$

So

$a=1$

2.1. Definitions and Examples

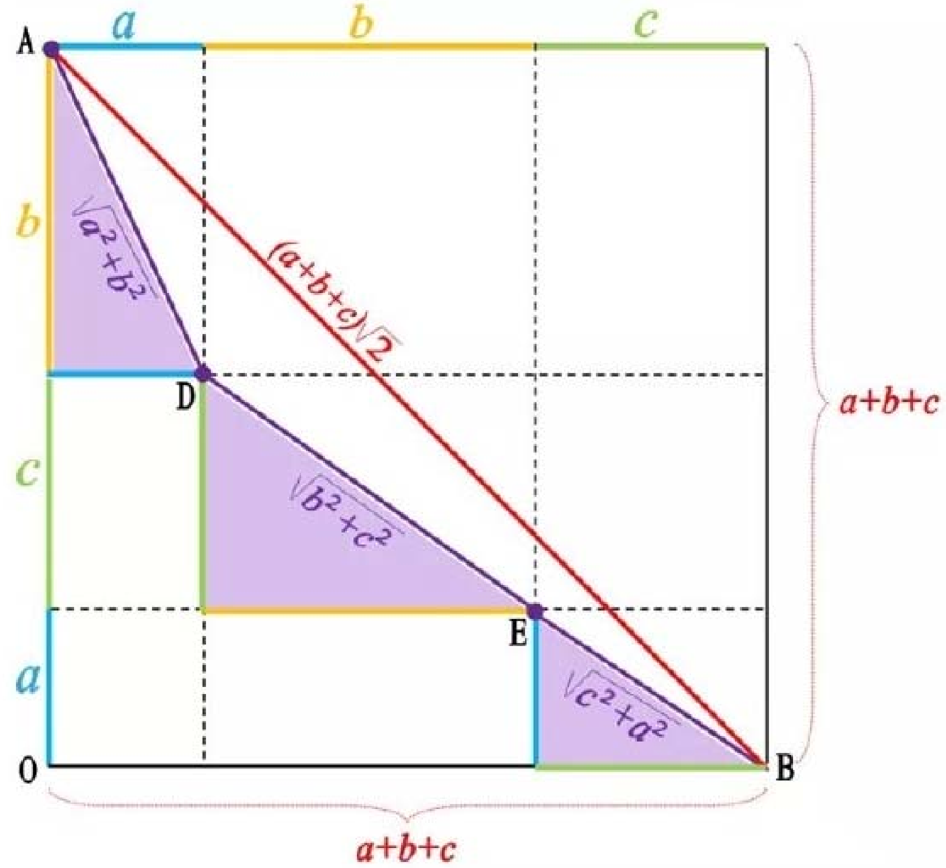
How to prove

$$\sqrt{a^2+b^2}+\sqrt{b^2+c^2}+\sqrt{c^2+a^2} \geq \sqrt{2}\left(a+b+c\right)$$

2.1. Definitions and Examples

Speechless Proof

$$\sqrt{a^2+b^2}+\sqrt{b^2+c^2}+\sqrt{c^2+a^2} \geq (a+b+c)\sqrt{2}$$



2.1. Definitions and Examples

设实数a,b,c,d满足$a \geq b \geq c \geq d > 0$，且$a+b+c+d=1$。证明：$(a+2b+3c+4d)a^a b^b c^c d^d < 1$。(IMO 2020)

2.1. Definitions and Examples

设实数a,b,c,d满足$a\ge b\ge c\ge d>0$，且$a+b+c+d=1$。证明：$(a+2b+3c+4d)a^a b^b c^c d^d<1$。

证明：由算术平均-几何平均不等式（Jensen's Inequality），我们有

$$a^a b^b c^c d^d < a\cdot a+b\cdot b+c\cdot c+d\cdot d=a^2+b^2+c^2+d^2$$

故只需证明

$$(a+2b+3c+4d)(a^2+b^2+c^2+d^2)<\left(\sum_{\text{cyc}} a\right)^3 \qquad (2.11)$$

注意到

$$\left(\sum_{\text{cyc}} a\right)^3=\sum_{\text{cyc}} a^3+3\sum_{\text{sym}} a^2 b+6\sum_{\text{cyc}} abc$$

2.1. Definitions and Examples

及

$$a^3+2ab^2+ad^2\ge a\sum_{\text{cyc}}a^2$$

$$2a^2b+ab^2+b^3+2bc^2+2bd^2\ge 2b\sum_{\text{cyc}}a^2$$

$$3a^2c+3b^2c+3ac^2+3cd^2\ge 3c\sum_{\text{cyc}}a^2$$

$$3a^2d+a^2b+4abd+4acd+4bcd\ge 4d\sum_{\text{cyc}}a^2$$

上述四个不等式相加，其左侧与$\left(\sum_{\text{cyc}}a\right)^3$相比差值为正数，

则(2.11)不等式成立，故原不等式也成立。

2.1. Definitions and Examples

定义在开的凸集上的凸函数必然是连续函数

Let X be a normed space, $x_0 \in X$, $r > 0$, $\varepsilon \in (0, r)$, $m, M \in \mathbb{R}$. Let $f : B^0(x_0, r) \to \mathbb{R}$ be a convex function.

(a) If $f(x) \le m$ on $B^0(x_0, r)$, then $|f(x)| \le |m| + 2|f(x_0)|$ on $B^0(x_0, r)$. 有界则绝对值有界

(b) If $|f(x)| \le M$ on $B^0(x_0, r)$, f is $(2M/\varepsilon)$-Lipschitz on $B^0(x_0, r-\varepsilon)$. Locally Lipschitz 连续

(c) f is locally Lipschitz on C $\Leftrightarrow$ f is continuous on C.

2.1. Definitions and Examples

Proof. By translation, we can suppose that $x_0 = 0$. Denote $B = B^0(0, r)$ and $C = B^0(0, r-\varepsilon)$. 为了简化书写，仅此而已

(a) 要证明第一个命题，我们需要把 $f(x)$ 和 $f(0)$ 联系起来，而手头能用的就是凸函数的基本性质

Since $0 = 0.5x + 0.5(-x)$ $(x \in B)$, according to convexity, we have $f(0) \le 0.5f(x) + 0.5f(-x)$. Consequently, we have

$$f(x) \ge 2f(0) - f(-x) \ge 2f(0) - m$$

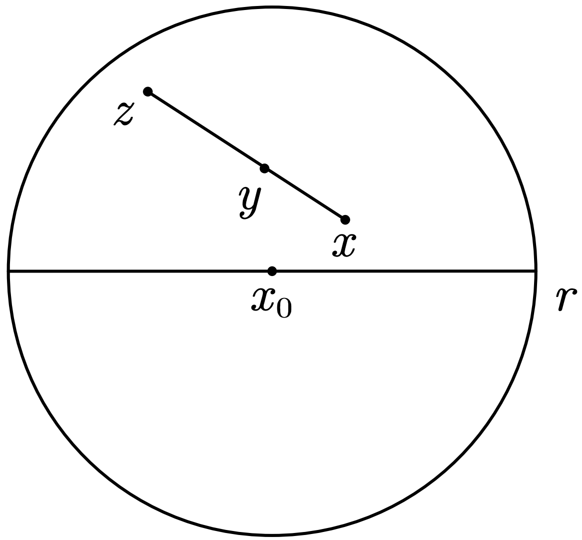
or

$$|f(x)| \le \max\{m, m - 2f(0)\} \le |m| + 2|f(0)| \quad (x \in B)$$

2.1. Definitions and Examples

(b) Consider two distinct points $x, y \in C$. $z = y + \dfrac{\varepsilon}{|y-x|}(y-x)$

belongs to B and $y \in (x, z)$.



An easy calculation shows that

$$y = \frac{\varepsilon}{\varepsilon + |y-x|}x + \frac{|y-x|}{\varepsilon + |y-x|}z \quad \text{(convex combination!)}$$

2.1. Definitions and Examples

Using convexity of f and multiplying by the common denominator, we get

$$\left(\varepsilon+\left|y-x\right|\right) f(y) \leq \varepsilon f(x)+\left|y-x\right| f(z)$$

Then $\varepsilon\left[f(y)-f(x)\right] \leq \left[f(z)-f(y)\right]\left|y-x\right| \leq 2M\left|y-x\right|$. So

$$f(y)-f(x) \leq \frac{2M}{\varepsilon}\left|y-x\right|$$

Interchanging the role of x and y, we obtain that f is $(2M/\varepsilon)$-Lipschitz on C.

2.1. Definitions and Examples

(c) Let $C \subset \mathbb{R}^d$ be open and convex, and $f : C \to \mathbb{R}$ a convex function. Fix $x_0 \in C$. There exist finitely many points $c_1, \dots, c_n \in C$ such that $x_0 \in U := \operatorname{int}[\operatorname{conv}\{c_1, \dots, c_n\}]$ (take, e.g., the vertices of a small d-dimensional cube centered at x_0). By convexity, $f \le \max\{f(c_1), \dots, f(c_n)\}$ on U.

Based on the conclusions of (a) and (b), f is locally Lipschitz on U.

Then, by definitions of uniform continuity, we finalize the proof.

推论：定义在开的凸集上的凸函数必有极小值

2.1. Definitions and Examples

一维凸函数连续性的另外一个证明：函数 f 在一维区间 I 上是凸函数，当且仅当 $\forall(x_1,x_2)\in I$ 及任何 $x\in(x_1,x_2)$ 有

$$\frac{f(x)-f(x_1)}{x-x_1}\le\frac{f(x_2)-f(x_1)}{x_2-x_1}\le\frac{f(x_2)-f(x)}{x_2-x}$$

证明：先证必要性。令

$$\lambda_1=\frac{x_2-x}{x_2-x_1},\quad \lambda_2=\frac{x-x_1}{x_2-x_1}$$

则 $\lambda_1,\lambda_2>0,\ \lambda_1+\lambda_2=1,\ x=\lambda_1x_1+\lambda_2x_2$， $f(x)$ 是凸函数，则

$$f(x)=f(\lambda_1x_1+\lambda_2x_2)\le\lambda_1f(x_1)+\lambda_2f(x_2)$$

又 $f(x)=\lambda_1f(x)+\lambda_2f(x)\le\lambda_1f(x_1)+\lambda_2f(x_2)$

整理得 $\lambda_1[f(x)-f(x_1)]\le\lambda_2[f(x_2)-f(x)]$，带入 λ_1,λ_2 得

$$\frac{x_2-x}{x_2-x_1}[f(x)-f(x_1)]\le\frac{x-x_1}{x_2-x_1}[f(x_2)-f(x)]$$

2.1. Definitions and Examples

即 $\dfrac{f(x)-f(x_1)}{x-x_1}\le\dfrac{f(x_2)-f(x)}{x_2-x}$。根据柯西不等式可知

$$\frac{f(x)-f(x_1)}{x-x_1}\le\frac{\left[f(x)-f(x_1)\right]+\left[f(x_2)-f(x)\right]}{\left(x-x_1\right)+\left(x_2-x\right)}\le\frac{f(x_2)-f(x)}{x_2-x}$$

即 $\dfrac{f(x)-f(x_1)}{x-x_1}\le\dfrac{f(x_2)-f(x_1)}{x_2-x_1}\le\dfrac{f(x_2)-f(x)}{x_2-x}$。

下证充分性。将 $\dfrac{f(x)-f(x_1)}{x-x_1}\le\dfrac{f(x_2)-f(x)}{x_2-x}$ 变形得到

$\dfrac{x_2-x}{x_2-x_1}[f(x)-f(x_1)]\le\dfrac{x-x_1}{x_2-x_1}[f(x_2)-f(x)]$，根据必要性证明的反演可得 $f(\lambda_1x_1+\lambda_2x_2)\le\lambda_1f(x_1)+\lambda_2f(x_2)$，由凸函数定义可知函数 f 是凸函数。证毕。

2.1. Definitions and Examples

上面提到的柯西不等式：如果$b>0$，$d>0$，$\frac{a}{b}\le\frac{c}{d}$，则

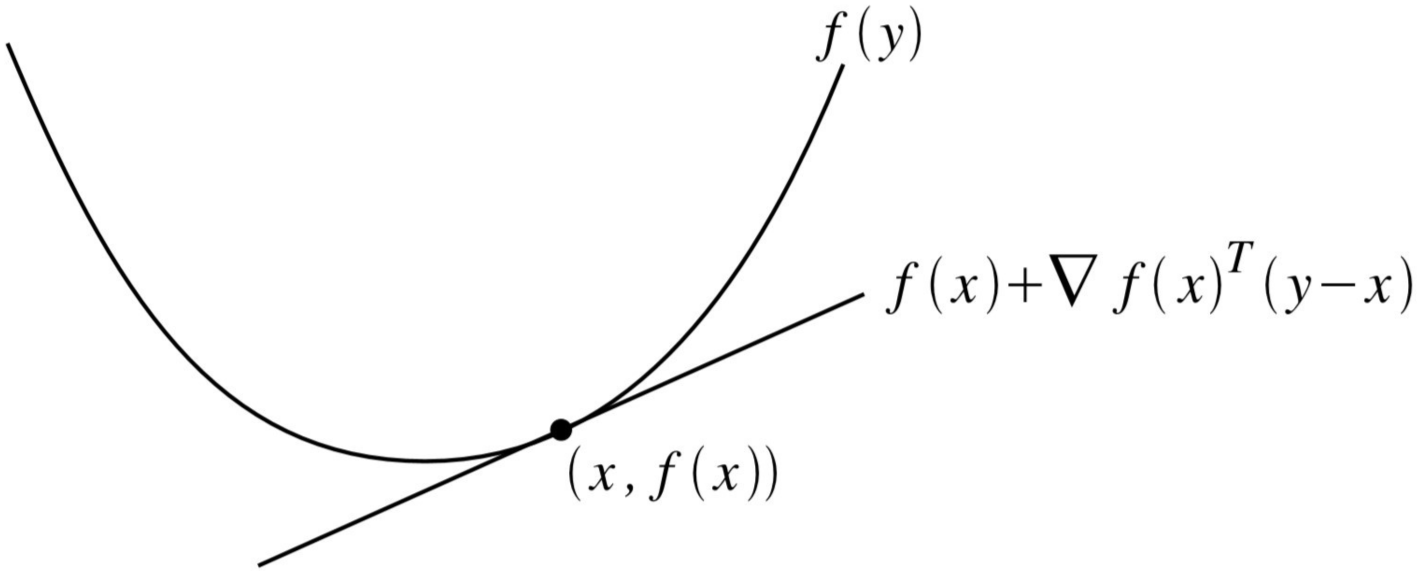
$$\frac{a}{b}\le\frac{a+c}{b+d}\le\frac{c}{d}$$

How to prove?

推论：用上述定理可以证明一维凸函数的连续性

2.1. Definitions and Examples

First Order Condition for Convexity If a differentiable function f satisfies $f(y) \geq f(x) + \nabla f(x)^T (y - x)$ for any x, y in its domain. 凸函数上任一点的切平面永远在函数下方



等价描述 $0 \geq [\nabla f(x) - \nabla f(y)]^T (x - y)$ (Monotone map)

2.1. Definitions and Examples

证明：首先证明一阶条件的必要性。

$$f(tx+(1-t)y) \le tf(x)+(1-t)f(y),\ \forall t \in [0,1],\ \forall x,y \in \mathrm{dom}(f)$$

注意到 $f(tx+(1-t)y) = f(y+t(x-y))$，以及 $tf(x)+(1-t)f(y) = f(y)+t(f(x)-f(y))$，所以可以把上面式子化简为

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

令 $t \to 0$，则不等式右侧恰好为方向导数的定义，也就是说

$$f(x)-f(y) \ge \nabla f(y)^T(x-y)$$

2.1. Definitions and Examples

下面证明一阶条件的充分性。注意到两个定义里面均要求定义域为凸集。

$\forall x \neq y,\ x, y \in \mathrm{dom}(f)$，取 $z = tx + (1-t)y \in \mathrm{dom}(f)$，则有

$$\begin{cases} f(x) \geq f(z) + \nabla f(z)^T (x - z) \\ f(y) \geq f(z) + \nabla f(z)^T (y - z) \end{cases}$$

第一个式子乘以 t 加上第二个式子乘以 $(1-t)$ 得到

$$tf(x) + (1-t) f(y) \geq f(z) = f(tx + (1-t)y)$$

证毕。

一阶条件的重要推论是，如果 $\nabla f(x) = 0$，那么无论定义域内的另外一点 y 是什么，都会有 $f(y) \geq f(x)$。所以，对于凸函数做优化，驻点为0就说明找到了极小值。

2.1. Definitions and Examples

Second Order Condition for Convexity If a twice differentiable function f makes the Hessian matrix $\nabla^2 f(x)$ for any x in its domain a positive semi-definite matrix.

二阶导数是对于一阶导数变化率的衡量

例题： $f(X)=\ln\det X$， $X\in\mathbb{S}_{++}^{n}$是一个凹函数。
证明：直接矩阵求导得到$\nabla^2 f(X)=-X^{-2}$

如果一个函数是严格凸的函数，并不能够推出$\nabla^2 f(x)>0$，一个反例就是 $f(x)=x^4$，它肯定是严格凸的，但是在原点处，其二阶导数并不是一个正数。

2.1. Definitions and Examples

Suppose f is convex and let $x, d \in \mathbb{R}^n$, then by first-order condition we have

$$f(x+td) \geq f(x) + t\nabla f(x)^T d$$

for all $t \in \mathbb{R}$.

Relpacing the left hand side of this inequality with its second-order Taylor expansion yields the inequality

$$f(x) + t\nabla f(x)^T d + \frac{t^2}{2} d^T \nabla^2 f(x) d + o(t^2) \geq f(x) + t\nabla f(x)^T d$$

or equivalently

$$\frac{1}{2} d^T \nabla^2 f(x) d + \frac{o(t^2)}{t^2} \geq 0$$

2.1. Definitions and Examples

Letting $t \to 0$ yields the inequality $d^T \nabla^2 f(x) d \geq 0$. Since d was arbitrary, $\nabla^2 f(x)$ is positive semi-definite at any x.

Conversely, if $x, y \in \mathbb{R}^n$, then by the mean value theorem there is a $\lambda \in (0,1)$ such that

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x_\lambda)(y - x)$$

where $x_\lambda = \lambda y + (1 - \lambda)x$. Hence

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

since $\nabla^2 f(x_\lambda)$ is positive semi-definite. Therefore, f is convex by first-order condition.

2.1. Definitions and Examples

一元凸函数和多元凸函数的重要关系

给任意点 z 和方向向量 v，如果 $g(t)=f(z+tv)$ 是一个关于一维变量 t 的凸函数，那么 $f(x)$ 是凸函数。反之亦然。

证明直接用凸函数的定义即可。

这个定义有的时候也称为一维刻画。因为它可以把一个任意维度的函数 $f(x)$，通过一个给定的方向 v，来降维到一个一维函数 $g(t)=f(z+tv)$，进而通过考虑该一维函数的性质来解决问题。反之，亦可从一个任意维度的函数 $f(x)$ 的一维切片来分析其的特性，特别是凸性。

2.1. Definitions and Examples

典型的凸函数包括：

- 指数函数 e^{ax}，负对数函数
- 仿射函数（同时是凸函数和凹函数）
- 正定或者半正定二次函数 $x^T Ax + b^T x + c$
- 范数（无法利用凸性的一阶条件以及二阶条件进行证明，因为范数本身可能并不是处处可微的）
- 幂函数 x^p，绝对值幂函数 $|x|^p$（$p \geq 1$ 为凸函数，$p \in (0,1)$ 为凹函数）

典型的凹函数包括：

- 对数函数 $\log(x)$

2.1. Definitions and Examples

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

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

convex if $P \succeq 0$

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

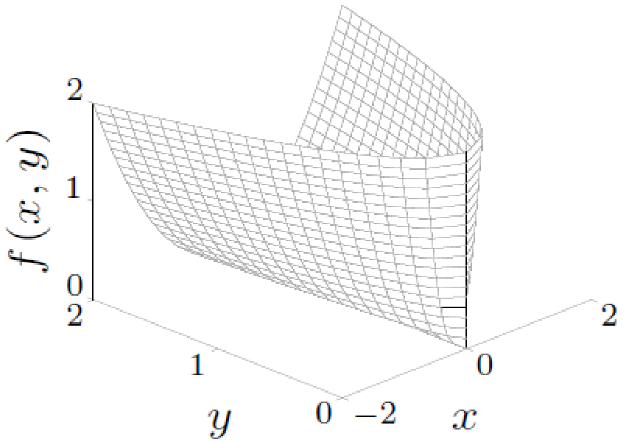
$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A$$

convex (for any A)

quadratic-over-linear: $f(x, y) = x^2/y$

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

convex for $y > 0$



2.1. Definitions and Examples

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

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z} \mathbf{diag}(z) - \frac{1}{(\mathbf{1}^T z)^2} z z^T \qquad (z_k = \exp x_k)$$

to show $\nabla^2 f(x) \succeq 0$, we must verify that $v^T \nabla^2 f(x) v \geq 0$ for all v:

$$v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0$$

since $(\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k)$ (from Cauchy-Schwarz inequality)

geometric mean: $f(x) = (\prod_{k=1}^{n} x_k)^{1/n}$ on $\mathbf{R}_{++}^n$ is concave

(similar proof as for log-sum-exp)

2.1. Definitions and Examples

如果凸函数 f 的一阶和二阶导数都存在，则以下性质等价：

1. $\nabla f(x)$是Lipschitz连续的，且常数为L
2. $[\nabla f(x)-\nabla f(y)]^T(x-y)\le L|y-x|^2,\ \forall x,y$
3. $\nabla^2 f(x)\preceq LI,\ \forall x$
4. $f(y)\le f(x)+\nabla f(x)^T(y-x)+\dfrac{L}{2}|y-x|^2$

证明：1→2可由Cauchy不等式得到，即$\forall x,y$

$$[\nabla f(x)-\nabla f(y)]^T(x-y)\le|\nabla f(x)-\nabla f(y)|\cdot|y-x|\le L|y-x|^2$$

2→3证明方法同上，3→4做二阶Taylor展开即可，4→2交换x,y顺序相加即得。最后3→1，对梯度做Taylor展开得

$$\nabla f(y)=\nabla f(x)+\nabla^2 f[x+\sigma(y-x)](y-x),\ \ \sigma\in[0,1]$$

注意到$\nabla^2 f(x)\preceq LI$恒成立，移项即得证。

到此已完成逻辑闭环，证毕。

2.2. Strong Convexity

若函数 $f(x)-\frac{m}{2}\|x\|_2^2$ 是一个凸函数，那么 $f(x)$ 就是一个凸性量度为 m 的强凸函数。

> Theorem 3:
>
> 以下性质等价
>
> 1. f 强凸，且凸性量度为 m
> 2. $(\nabla f(x)-\nabla f(y))^T(x-y) \geq m\|x-y\|^2, \forall x, y$
> 3. $\nabla^2 f(x) \succeq mI, \forall x$
> 4. $f(y) \geq f(x)+\nabla f(x)^T(y-x)+\frac{m}{2}\|y-x\|^2$

我们先证明 $1 \rightarrow 2$。假如说 f 强凸，那么有 $g(x)=f(x)-\frac{m}{2}\|x\|^2$ 是一个凸函数，那么注意到对于凸函数，我们有

2.2. Strong Convexity

我们先证明$1 \to 2$。假如说f强凸，那么有$g(x) = f(x) - \frac{m}{2}\|x\|^2$是一个凸函数，那么注意到对于凸函数，我们有

> $$(\nabla g(x) - \nabla g(y))^T(x-y) \geq 0, \forall x, y$$

那么我们又可以发现$\nabla g(x) = \nabla f(x) - mx$，所以代入就可以了。

然后我们证明$2 \to 3$。注意到如果设$x = y + tv$，那么会有

> $$(\nabla f(y+tv) - \nabla f(y))^T v \geq mt\|v\|^2$$

两边除以t，并且令$t \to 0$，根据方向导数的公式（但这次用的是计算海塞矩阵的公式，《数值优化》第1节），可以得到

> $$v^T \nabla^2 f(y) v \geq m\|v\|^2$$

这个就是$\nabla^2 f(y) \succeq mI$的意思，因为v是任意的。

2.2. Strong Convexity

接下来证明$3 \to 4$，这个没什么好说的，二阶Taylor展开就可以了，和上面证明凸函数的二阶信息等价性是类似的思路。然后就是$4 \to 1$，这个也是一样，只需要证明$g(x) = f(x) - \frac{m}{2}\|x\|^2$是一个凸函数。而注意到第四个式子想说明的内容是

> $$f(y) - \frac{m}{2}\|y\|^2 \geq f(x) - \frac{m}{2}\|x\|^2 + (\nabla f(x) - mx)^T(y - x)$$

而这个就是凸函数的一阶信息刻画，所以自然也就得到了结论。至此我们已经得到了一条完整的闭环，所以这个等价性就算证明完毕了。

2.3. Operations that Preserve Convexity

practical methods for establishing convexity of a function

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

2. show that $f(x)$ is obtained from simple convex functions by operations that preserve convexity

- nonnegative weighted sum
- composition with affine function
- pointwise maximum and supremum
- composition
- minimization
- perspective

2.3. Operations that Preserve Convexity

nonnegative multiple: αf is convex if f is convex, $\alpha \geq 0$

sum: $f_1 + f_2$ convex if f_1, f_2 convex (extends to infinite sums, integrals)

composition with affine function: $f(Ax + b)$ is convex if f is convex

examples

- log barrier for linear inequalities

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

- (any) norm of affine function: $f(x) = \|Ax + b\|$

2.3. Operations that Preserve Convexity

Pointwise maximum

if $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

examples

- piecewise-linear function: $f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$ is convex
- sum of r largest components of $x \in \mathbf{R}^n$:

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

is convex ($x_{[i]}$ is ith largest component of x)

proof:

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$$

2.3. Operations that Preserve Convexity

Pointwise supremum

if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, then

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

is convex

examples

- support function of a set C: $S_C(x) = \sup_{y \in C} y^T x$ is convex
- distance to farthest point in a set C:

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

- maximum eigenvalue of symmetric matrix: for $X \in \mathbf{S}^n$,

$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^T X y$$

2.3. Operations that Preserve Convexity

Composition with scalar functions复合函数 $f(x)=h(g(x))$，假如它们具有二阶可导的性质，则可以得到

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

因此可得以下结论：Rules for Composite Convex Functions
设 f,g,h 二阶可导，且 $f(x)=h(g(x))$，那么

1. 如果h为凸函数且不降，g为凸函数，那么 f 为凸函数。

2. 如果h为凸函数且不增，g为凹函数，那么 f 为凸函数。

3. 如果h为凹函数且不降，g为凹函数，那么 f 为凹函数。

4. 如果h为凹函数且不增，g为凸函数，那么 f 为凹函数。

2.3. Operations that Preserve Convexity

例题：证明$g(x)=\log\left(\sum_{i=1}^{k}\exp\left(a_i^T x+b_i\right)\right)$是一个凸函数。

这个函数是$\max_i\{a_i^T x+b_i\}$的一个好的近似，所以有时称其为soft-max，神经网络中的softmax层就来源于此。

可以看出这个函数是一个复杂的复合。指数上的$a_i^T x+b_i$可以拆分出来作为一个新的函数，所以实际上内层函数是一个仿射函数，当然是一个凸函数，而外层函数就是

$$f(x)=\log\left(\sum_{i=1}^{n}e^{x_i}\right)，\quad x_i\text{为}x\text{的各个分量}$$

这个函数当然是一个不降的函数。所以根据上面提供的几条准则，便可通过f看出g的凸性。

2.3. Operations that Preserve Convexity

对其求Hessian矩阵

$$\nabla_i f(x) = \frac{e^{x_i}}{\sum_{l=1}^{n} e^{x_l}}, \nabla_{ij}^2 f(x) = \frac{e^{x_i}}{\sum_{l=1}^{n} e^{x_l}} I(i=j) - \frac{e^{x_i} e^{x_j}}{\left(\sum_{l=1}^{n} e^{x_l}\right)^2}$$

将其改写为更加紧凑的形式

$$\nabla^2 f(x) = \mathrm{diag}(z) - zz^T$$

其中$\mathrm{diag}(z)$为对角阵，第i个对角元素为z_i，且有

$$\sum_{i=1}^{n} z_i = 1$$

注意到

$$y^T \nabla^2 f(x) y = \sum_{i=1}^{n} z_i y_i^2 - \left(\sum_{i=1}^{n} z_i y_i\right)^2$$

根据Jensen不等式即可得到其为正定矩阵的结论。

2.3. Operations that Preserve Convexity

我们还可以对$h(x)$做一个拓展，就是在其没有定义的地方人为规定它们都是正无穷或者负无穷。比方$h(x)=\log(x)$就是一个定义域不在全空间的函数。但是我们可以通过延拓，也就是额外设$h(x)=-\infty$，$x\le 0$。这样的话就可以得到一个凸的，又具有单调性的函数$\tilde{h}(x)$。如果可以构造出一个这样的全空间的函数，又不影响原始定义域的值，那就算是一个合理的拓展。

考察$g(x)=x^2$，$h(z)=0$，$\mathrm{dom}(h)=[1,2]$，那么可以得到$f(x)=h\big(g(x)\big)=0$，$x\in[-\sqrt{2},-1]\cup[1,\sqrt{2}]$，很明显这个函数并不是一个凸函数，因为它的定义域都不是一个凸集。错误的原因就在于，如果我们考虑$h(x)$的拓展，会发现无论我们要求它的函数是凸还是凹，都做不到让$\tilde{h}(x)$是单调的。

2.3. Operations that Preserve Convexity

Vector composition

composition of $g : \mathbf{R}^n \to \mathbf{R}^k$ and $h : \mathbf{R}^k \to \mathbf{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

f is convex if
- g_i convex, h convex, $\tilde{h}$ nondecreasing in each argument
- g_i concave, h convex, $\tilde{h}$ nonincreasing in each argument

proof (for $n = 1$, differentiable g, h)

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

examples

- $\sum_{i=1}^m \log g_i(x)$ is concave if g_i are concave and positive
- $\log \sum_{i=1}^m \exp g_i(x)$ is convex if g_i are convex

2.3. Operations that Preserve Convexity

Minimization

if $f(x, y)$ is convex in (x, y) and C is a convex set, then

$$g(x) = \inf_{y \in C} f(x, y)$$

is convex

examples

- $f(x, y) = x^T Ax + 2x^T By + y^T Cy$ with

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0, \qquad C \succ 0$$

  minimizing over y gives $g(x) = \inf_y f(x, y) = x^T (A - BC^{-1}B^T)x$

  g is convex, hence Schur complement $A - BC^{-1}B^T \succeq 0$

- distance to a set: $\mathbf{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex

2.3. Operations that Preserve Convexity

Perspective

the **perspective** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$,

$$g(x,t) = tf(x/t), \qquad \mathbf{dom}\, g = \{(x,t) \mid x/t \in \mathbf{dom}\, f,\ t > 0\}$$

g is convex if f is convex

examples

- $f(x) = x^T x$ is convex; hence $g(x,t) = x^T x/t$ is convex for $t > 0$
- negative logarithm $f(x) = -\log x$ is convex; hence relative entropy $g(x,t) = t\log t - t\log x$ is convex on $\mathbf{R}^2_{++}$
- if f is convex, then

$$g(x) = (c^T x + d) f\left((Ax+b)/(c^T x + d)\right)$$

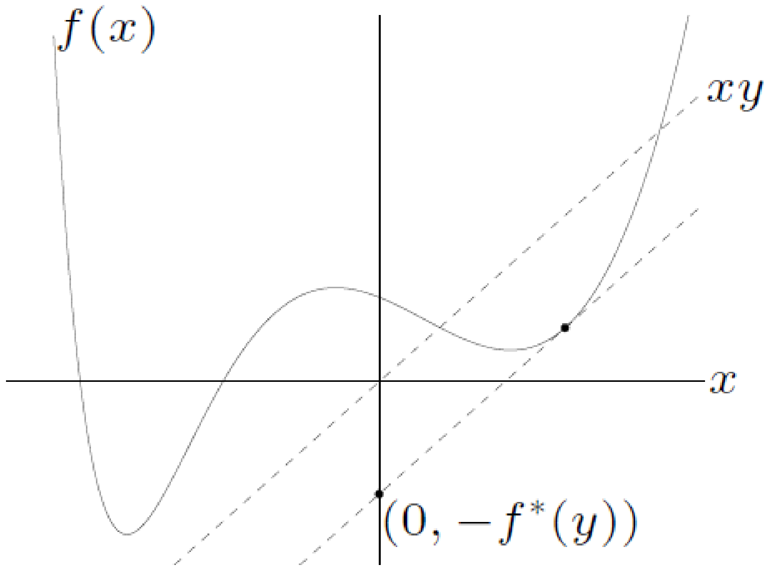
is convex on $\{x \mid c^T x + d > 0,\ (Ax+b)/(c^T x + d) \in \mathbf{dom}\, f\}$

2.3. Operations that Preserve Convexity

The conjugate function

the **conjugate** of a function f is

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$$



- f^* is convex (even if f is not)

2.3. Operations that Preserve Convexity

为了解释conjugate function是怎么来的，我们先讨论一下

对一个严格凸函数，找不到两点的导数相等。为什么？严格凸函数图像的切线（切面）永远严格位于函数的下方，除了切点。写成公式就是

$$\forall x', f(x') \geq f(x) + \langle \nabla f(x), x' - x \rangle$$

假设对于严格凸函数 f，存在两个点 x_1, x_2，它们的梯度相等。根据刚才说的几何意义，得

$$f(x_1) > f(x_2) + \nabla f(x_2)(x_1 - x_2)$$
$$f(x_2) > f(x_1) + \nabla f(x_1)(x_2 - x_1)$$

两式相加，矛盾。

2.3. Operations that Preserve Convexity

这说明$x \mapsto \nabla f(x)$是单射。这启发我们，能否从梯度的角度考虑一个凸函数呢？画图可知，一个凸函数的所有切线构成了一个对原函数的包络。具体来说有如下直观的刻画。

> **定理（凸函数的直线包络）** 对于凸函数f，有
> $f(x) = \max\limits_{a,b:\forall y, f(y) \geq ay+b} ax + b$。

可以说，给定一个凸函数f，由它所有切线的（斜率，截距）信息可以完整地恢复出原来的函数f。可以说，给定斜率y，我们要找的是f的斜率为y的切线的负截距$g(y)$。但我们知道，这个切点的横坐标一定是$x = (\nabla f)^{-1}(y)$。于是切线的方程是

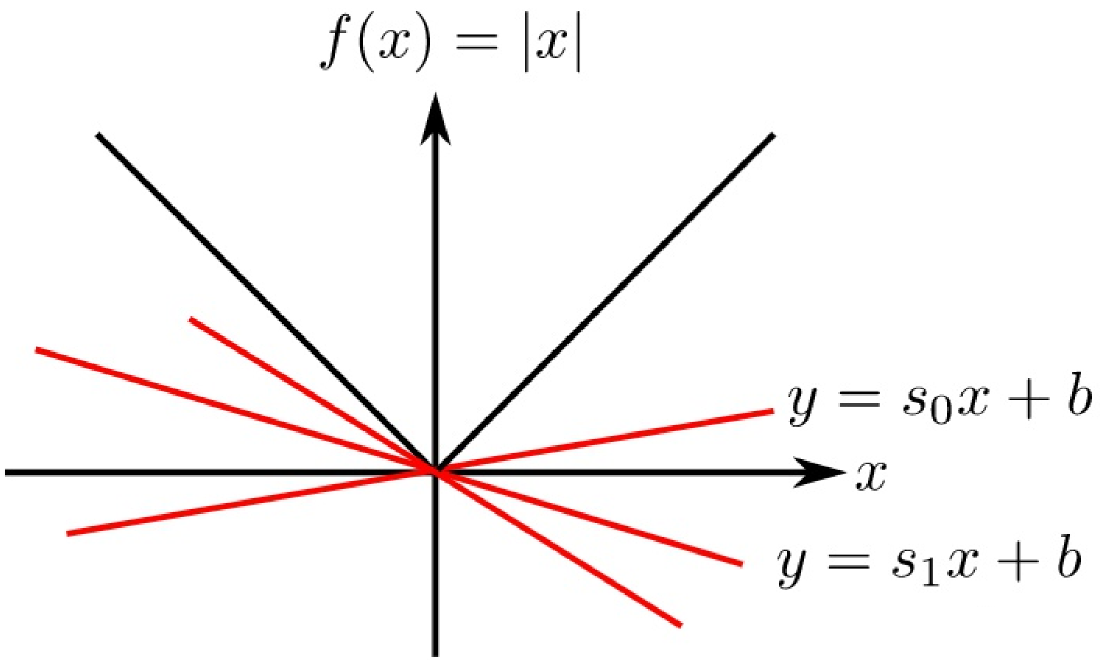
$z = f(x) + \langle \nabla f(x), x' - x \rangle$,因此负截距是

$$
\begin{aligned}
g(y) = -z|_{x'=0} &= -f(x) + \langle \nabla f(x), x \rangle \\
&= -f(x) + \langle y, x \rangle \\
&= -f((\nabla f)^{-1}(y)) + \langle y, (\nabla f)^{-1}(y) \rangle
\end{aligned}
$$

2.3. Operations that Preserve Convexity

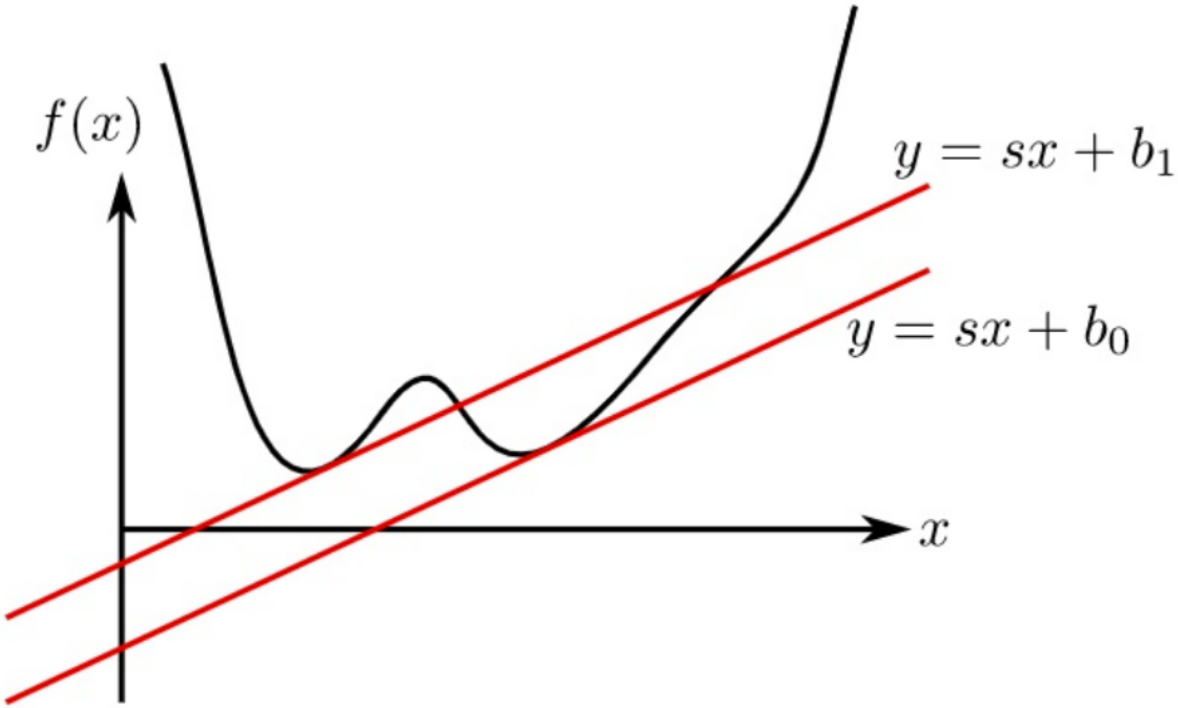
对于凸可微函数，我们把 $(\nabla f)^{-1}(x)$ 或者 $-(\nabla f)^{-1}(x)$ 称为 $f(x)$ 的Legendre变换

Legendre变换本身很有用，但它仅限于凸函数和可微函数, 如果这两个条件有一个不满足, 那么这个变换就无法完成. 考虑 $f(x) = |x|$ 的情形. 函数的斜率在 $x = 0$ 处时是没有定义的, 而其次梯度的范围则为 $s \in [-1, 1]$，如下图所示:



2.3. Operations that Preserve Convexity

同样, 非凸函数在不同的点处其斜率的取值范围与上面的函数类似, 这导致 x 和 s 之间不存在唯一的对应关系. 如下图, 同一个斜率 s 对应着两个 b 的值.



怎么解决？选一组与 $f(x)$ 相交的直线，然后寻找最小截距的那条. 这使得不可微甚至是非凸函数也可以使用这一变换

2.3. Operations that Preserve Convexity

我们注意到负截距有另一种求法。给定斜率 y. 从直线 $z = \langle y, x' \rangle$ 出发，我们知道所求负截距是将此直线向下平移的最小量，使得 f 刚好接触到这条直线。写成数学语言就是下面的优化问题：

$$g(y) = \max_x \langle y, x \rangle - f(x)$$

反过来，我们有

> **定理** 若 f 是凸函数，g 定义如上，则
> $$f(x) = \max_y \langle y, x \rangle - g(y)$$
>
> .

这个定理其实就是凸函数的直线包络结论的推论。注意到对任何斜率 a ,最大的 b 就是 $g(a)$.

这个 g 有时候又被称为 f^* ,是 f 的**Legendre-Fenchel共轭**。

2.3. Operations that Preserve Convexity

- negative logarithm $f(x) = -\log x$

$$
\begin{aligned}
f^*(y) &= \sup_{x>0}(xy + \log x) \\
&= \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & \text{otherwise} \end{cases}
\end{aligned}
$$

- strictly convex quadratic $f(x) = (1/2)x^T Q x$ with $Q \in \mathbf{S}^n_{++}$

$$
\begin{aligned}
f^*(y) &= \sup_{x}(y^T x - (1/2)x^T Q x) \\
&= \frac{1}{2} y^T Q^{-1} y
\end{aligned}
$$

2.3. Operations that Preserve Convexity

Boyd老师组开发的Disciplined Convex Programming通过基本的凸函数原子库（atom library）和凸性演算规则（convexity calculus rules），来推演一个给定的函数是否是凸函数。具体来说，凸性演算规则包括10条顶层法则（top-level rules），无乘积法则（product-free rules），符号法则（sign rules），复合法则（composition rules）。无乘积法则是指避免2个凸函数相乘的表示，符号法则是指避免两个凸函数相减的表示。程序会尝试可能的函数变形，看是否归结为已知的某个凸优化问题，以减少漏判误判。

一般来说，判断一个函数是否是凸函数是NP-hard的。例如判断一个多元四次及以上偶多项式是否是凸的是strongly NP-hard的。

2.4. Quasi-Convexity

下水平集sublevel set和上境图epigraph

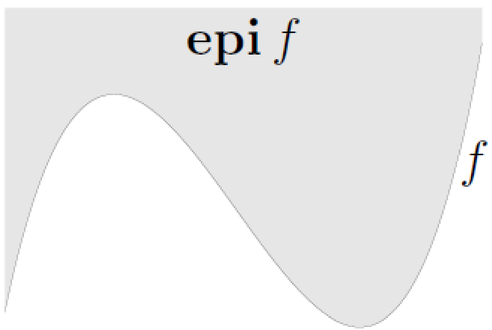
α-**sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$:

$$C_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

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

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

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



f is convex if and only if $\mathbf{epi}\, f$ is a convex set

2.4. Quasi-Convexity

定义1：

Quasi Convex : 对$\forall\alpha, S_\alpha = \{x \in \mathrm{dom} f \mid f(x) \leq \alpha\}$为凸。

即拟凸函数的所有的α-sublevel-set都是凸集。

定义2：

$$\max\{f(x), f(y)\} \geq f\big(\theta x + (1-\theta)y\big)$$
$$\mathrm{dom} f\text{为凸}, \forall x, y \in \mathrm{dom} f, 0 \leq \theta \leq 1$$

类似的有**拟凹函数(quasiconcave)** 的定义。如果一个函数既是拟凸的，又是拟凹的，那么它是**拟线性(quasilinear)** 的。

> **Reamrks**：对于拟线性函数，要求其上水平集和下水平集同时是凸集，因此简单理解，其在某种意义上具有“单调性”。比如 $e^x, \log x$ 等。

2.4. Quasi-Convexity

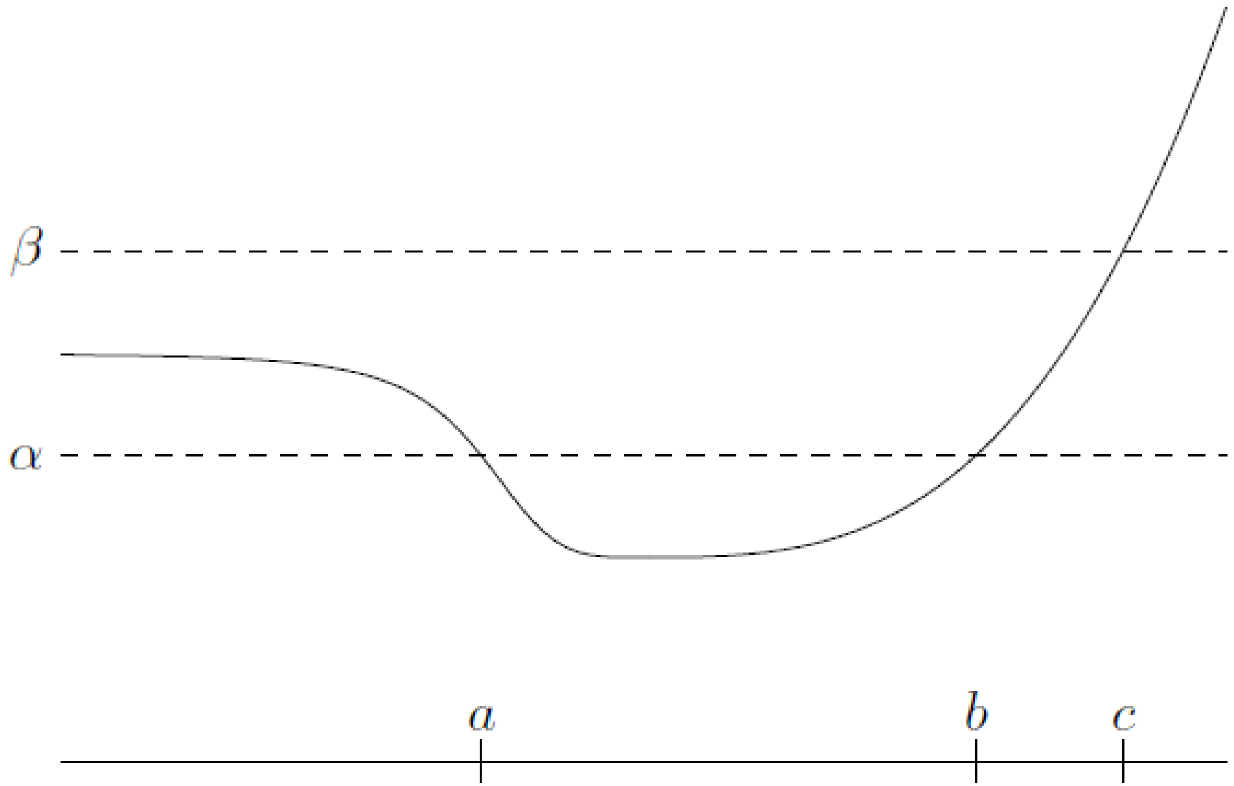


Figure 3.9 A quasiconvex function on $\mathbf{R}$. For each α, the α-sublevel set S_α is convex, *i.e.*, an interval. The sublevel set S_α is the interval $[a, b]$. The sublevel set S_β is the interval $(-\infty, c]$.

2.4. Quasi-Convexity

- $\sqrt{|x|}$ is quasiconvex on $\mathbf{R}$
- $\text{ceil}(x) = \inf\{z \in \mathbf{Z} \mid z \geq x\}$ is quasilinear
- $\log x$ is quasilinear on $\mathbf{R}_{++}$
- $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbf{R}^2_{++}$
- linear-fractional function

$$f(x) = \frac{a^T x + b}{c^T x + d}, \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$$

  is quasilinear
- distance ratio

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}, \qquad \mathbf{dom}\, f = \{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$$

  is quasiconvex

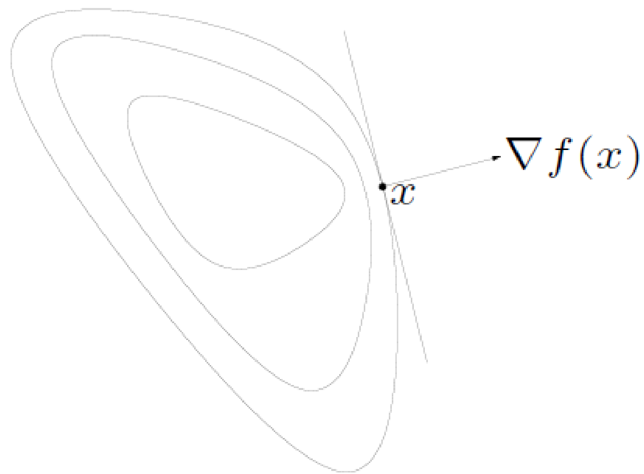
2.4. Quasi-Convexity

modified Jensen inequality: for quasiconvex f

$$0 \leq \theta \leq 1 \quad \Longrightarrow \quad f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\}$$

first-order condition: differentiable f with cvx domain is quasiconvex iff

$$f(y) \leq f(x) \quad \Longrightarrow \quad \nabla f(x)^T (y - x) \leq 0$$



sums of quasiconvex functions are not necessarily quasiconvex

2.4. Quasi-Convexity

等价定义 2：$g(t) = f(x+tv)$ quasiconvex $\iff$ $f(x)$ quasiconvex

证明可以直接用定义。

对于拟凸函数来说，没有二阶的充分必要条件，有充分条件和必要条件。

> **必要条件**：$f(x)$ quasiconvex $\Longrightarrow$ 对任意 $x \in \mathrm{dom} f, y \in R^n$
>
> $$\text{if} \quad y^T \nabla f(x) = 0 \Longrightarrow y^T \nabla^2 f(x) y \geq 0$$
>
> 对于一维函数，只需要 $f'(x) = 0 \Longrightarrow f''(x) \geq 0$
>
> **充分条件**：$f(x)$ quasiconvex $\Longleftarrow$ 对任意
>
> $x \in \mathrm{dom} f, y \in R^n, y \neq 0$
>
> $$\text{if} \quad y^T \nabla f(x) = 0 \Longrightarrow y^T \nabla^2 f(x) y > 0$$
>
> 证明：注意这里对于一维函数 $f: R \to R$ 较简单，因此可以应用"降维打击"的等价定义进行证明。

2.4. Quasi-Convexity

拟凸函数的保凸变换

~~正权重求和~~
与仿射变换复合
最大值/上确界
单调凸函数与凸函数的复合
下确界
~~透射变换~~

2.5. Log-Concave and Log-Convex

a positive function f is log-concave if $\log f$ is concave:

$$f(\theta x + (1-\theta)y) \geq f(x)^\theta f(y)^{1-\theta} \quad \text{for } 0 \leq \theta \leq 1$$

f is log-convex if $\log f$ is convex

- powers: x^a on $\mathbf{R}_{++}$ is log-convex for $a \leq 0$, log-concave for $a \geq 0$
- many common probability densities are log-concave, $e.g.$, normal:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1}(x-\bar{x})}$$

- cumulative Gaussian distribution function Φ is log-concave

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

2.5. Log-Concave and Log-Convex

- twice differentiable f with convex domain is log-concave if and only if

$$f(x)\nabla^2 f(x) \preceq \nabla f(x)\nabla f(x)^T$$

for all $x \in \mathbf{dom}\, f$

- product of log-concave functions is log-concave
- sum of log-concave functions is not always log-concave
- integration: if $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is log-concave, then

$$g(x) = \int f(x, y)\, dy$$

is log-concave (not easy to show)

2.5. Log-Concave and Log-Convex

- convolution $f * g$ of log-concave functions f, g is log-concave

$$(f * g)(x) = \int f(x-y)g(y)dy$$

- if $C \subseteq \mathbf{R}^n$ convex and y is a random variable with log-concave pdf then

$$f(x) = \mathbf{prob}(x + y \in C)$$

  is log-concave

  proof: write $f(x)$ as integral of product of log-concave functions

$$f(x) = \int g(x+y)p(y)\,dy, \qquad g(u) = \begin{cases} 1 & u \in C \\ 0 & u \notin C, \end{cases}$$

  p is pdf of y

2.6. Convexity w.r.t. Generalized Inequalities

$f : \mathbf{R}^n \to \mathbf{R}^m$ is K-convex if $\mathbf{dom}\, f$ is convex and

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

for $x,\ y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$

example $f : \mathbf{S}^m \to \mathbf{S}^m$, $f(X) = X^2$ is $\mathbf{S}^m_+$-convex

proof: for fixed $z \in \mathbf{R}^m$, $z^T X^2 z = \|Xz\|_2^2$ is convex in X, *i.e.*,

$$z^T(\theta X + (1-\theta)Y)^2 z \le \theta z^T X^2 z + (1-\theta) z^T Y^2 z$$

for $X, Y \in \mathbf{S}^m$, $0 \le \theta \le 1$

therefore $(\theta X + (1-\theta)Y)^2 \preceq \theta X^2 + (1-\theta)Y^2$

2.7. Not Exactly Convex but …

IMO 2021

第2题 对任意实数x_1， x_2， …， x_n，证明下述不等式成立

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{\left|x_i - x_j\right|} \le \sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{\left|x_i + x_j\right|}$$

2.8. References

[1] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. http://www.stanford.edu/~boyd/cvxbook/ http://www.ee.ucla.edu/~vandenbe/cvxbook

[2] J. M. Steele, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*, Cambridge University Press, Cambridge, U.K., 2004.

[3] 如何理解勒让德变换 https://www.zhihu.com/question/26050948

[4] Legendre-Fenchel 变换的物理意义是什么 https://www.zhihu.com/question/44860559

[5] 勒让德变换 Legendre Transformation https://zhuanlan.zhihu.com/p/341207902

[6] M. Grant, S. Boyd, Y. Ye, "Disciplined convex programming," *Global Optimization: From Theory to Implementation*, L. Liberti, N. Maculan, eds., pp. 155-210, 2006.