# 运筹学与优化方法

晁国清
计算机科学与技术学院

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

# 课程回顾

- 凸集的定义　对任何$x_1 \in C$，$x_2 \in C$,有$\lambda x_1 + (1-\lambda)x_2 \in C(0 \le \lambda \le 1)$,则称$C$为凸集。
- 常见凸集　超平面；半空间；多面体；球体；椭圆；二阶锥；半定矩阵锥
- 凸集合的性质　投影定理、分割超平面定理和支撑超平面定理

# 凸函数

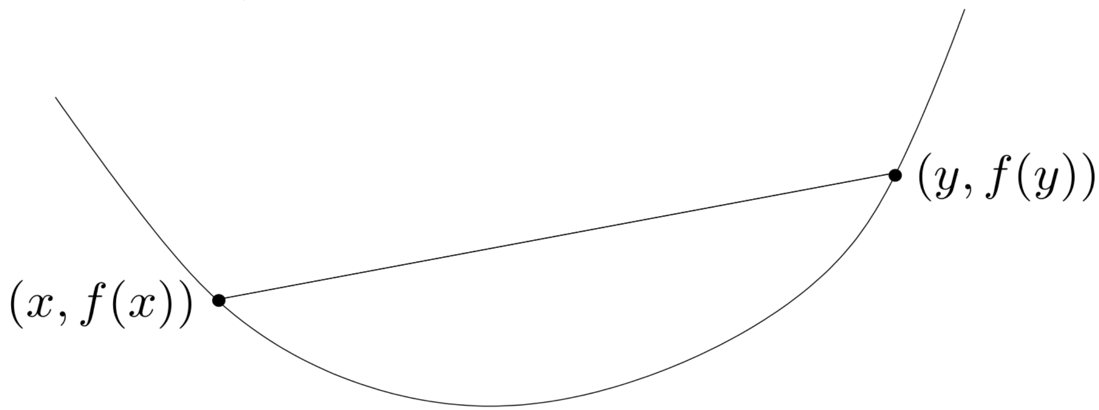
- 凸函数的定义
- 常见凸函数
- 凸函数的性质
- 保持凸性的操作

# 凸函数的定义

- **凸函数**：设$C$是非空凸集，$f$是定义在$C$上的函数，如果对于任意的$x, y \in C$，和$\lambda \in [0,1]$，均有

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y),$$

则称$f$为$C$上的凸函数




通俗讲，函数值位于连接$(x, f(x))$与$(y, f(y))$的线段下方的函数就是凸函数

**凹函数**：凸函数定义中不等号符号相反，即

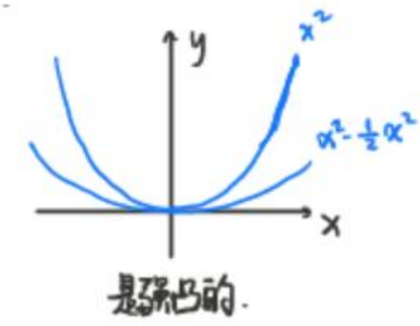
$$f \ concave \Leftrightarrow -f \ convex$$

# 严格凸与强凸

- 严格凸：对于任意的$x \neq y$与$0 < \lambda < 1$, 有 $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$

通俗讲，$f$是凸函数且有一个大于线性函数的曲率

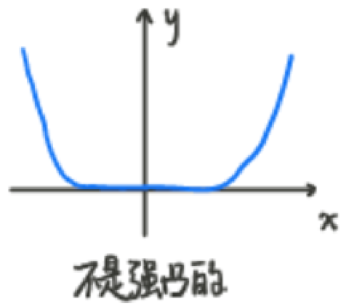
- 强凸：有一个$m > 0$，使得$f - \frac{m}{2}\|x\|_2^2$是凸的，$f$就是一个凸性量度为$m$的强凸函数



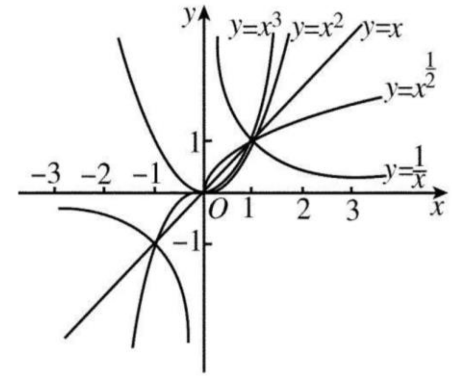

通俗讲：　$f$至少像二次函数凸

注意：强凸函数⇒严格凸函数⇒凸函数

（对于凹函数类似成立）



# 常见的凸函数




- 单变量函数Univariate functions:
  - 指数函数Exponential function: $e^{ax}$ is convex, $\forall a \in R$
  - 幂函数Power function: $x^a$ is convex, for $a \geq 1$ $or$ $a \leq 0$ over $R_+$
  - 幂函数Power function: $x^a$ is concave, for $0 \leq a \leq 1$ over $R_+$
  - 对数函数Logarithmic function: $\log x$ is concave over $R_+$
- 放射函数Affine function: $a^T x + b$
- 二次函数Quadratic function: $f(x) = \frac{1}{2} x^T Q x + b^T x + c, Q \in S_+^n$
- 最小平方损失Least squares loss: $f(x) = \|y - Ax\|_2^2$ $(A^T A \succcurlyeq 0)$

# 常见的凸函数

- 范数Norm：$\|\mathrm{x}\|$is convex for any norm；比如$\ell_p$范数

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

常用的是$\ell_2$和$\ell_1$范数。此外，$\|\mathrm{x}\|_\infty = \max_{i=1,\cdots,n} |x_i|$也是凸函数

算子(谱)范数 Operator(spectral) norm:

$$\|X\|_{op} = \sigma_1(X)$$

迹(核)范数 trace(nuclear)norm:

$$\|X\|_{tr} = \sum_{i=1}^{r} \sigma_r(X)$$

$\sigma_1(X) \geq \cdots \geq \sigma_r(X) \geq 0$ 是矩阵$X$的奇异值

# 凸函数的性质

➢一阶特性：设$C\subset R^n$是非空开凸集，$f(x)$是可微的，则$f(x)$是$C$上的凸函数$\Leftrightarrow$ $f(y)\ge f(x)+\nabla f(x)^T(y-x),\forall x,y\in C$.

证明：充分性：需要证“$f(\lambda x+(1-\lambda)y)\le \lambda f(x)+(1-\lambda)f(y),\forall x,y\in C$”

记$z=\lambda x+(1-\lambda)y$，由已知条件得到

$$f(x)\ge f(z)+\nabla f(z)^T(x-z),\qquad (1)$$

$$f(y)\ge f(z)+\nabla f(z)^T(y-z),\qquad (2)$$

由$(1)*\lambda+(2)*(1-\lambda)$, 得到

$$\lambda f(x)+(1-\lambda)f(y)\ge f(z)+\nabla f(z)^T(\lambda(x-z)+(1-\lambda)(y-z)))$$

$$\ge f(z)=f(\lambda x+(1-\lambda)y),$$ 得证。

必要性：已知$f(x)$为凸函数，$\forall x,y\in C$，则有

$f(\lambda y+(1-\lambda)x)\le \lambda f(y)+(1-\lambda)f(x),\lambda\in(0,1)$，即

$$f(x+\lambda(y-x))\le f(x)+\lambda\big(f(y)-f(x)\big)$$

$$\frac{f(x+\lambda(y-x))-f(x)}{\lambda}\le f(y)-f(x)$$

$$f(x)+\frac{f(x+\lambda(y-x))-f(x)}{\lambda}\le f(y)$$

泰勒展开：$f(x+\lambda(y-x))=f(x)+\lambda\nabla f(x)^T(y-x)+\circ(\lambda\|y-x\|)$, 左右两边同除$\lambda$，然后令$\lambda\to 0$，得到$\frac{f(x+\lambda(y-x))-f(x)}{\lambda}=\nabla f(x)^T(y-x)$，得证。

# 凸函数的性质

➢二阶特性：设$C \subset R^n$是非空开凸集，$f(x)$在$C$上二阶连续可微，则$f(x)$是$C$上的凸函数$\Leftrightarrow$ $f(x)$的海塞Hessian matrix $\nabla^2 f(x) \succcurlyeq 0, \forall x \in C$.

证明:

充分性：已知$\forall x \in C, \nabla^2 f(x) \succcurlyeq 0$, 对$f(y)$在$f(x)$处展开，得到

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

$$\xi = x + \lambda(y-x) \in [x,y], \lambda \in (0,1)$$

由于$\nabla^2 f(x) \succcurlyeq 0$，所以$f(y) \geq f(x) + \nabla f(x)^T (y-x), \forall x, y \in C$ 得证

必要性：任取$x \in C$,对$f(x+\alpha d)$在$f(x)$处泰勒展开，得到

$$f(x+\alpha d) = f(x) + \alpha \nabla f(x)^T d + \frac{\alpha^2}{2} \mathrm{d}^T \nabla^2 f(x)^T \mathrm{d} + \circ(\alpha^2 \|d\|^2)$$

由于$f(x+\alpha d) \geq f(x) + \alpha \nabla f(x)^T d$, 则

$$\frac{1}{2} \mathrm{d}^{\mathrm{T}} \nabla^2 f(x)^T d + \frac{\circ(\alpha^2 \|d\|^2)}{\alpha^2} \geq 0$$

令$\alpha \to 0$, 取极限得到$\mathrm{d}^T \nabla^2 f(x)^T d \geq 0 \Rightarrow \nabla^2 f(x) \succcurlyeq 0$

# 保持凸性的操作(有作业)

- 非负线性组合nonnegative linear combination: $f_1(x),\cdots,f_n(x)$ 都是凸函数，则$g(x)=\lambda_1 f_1(x)+\cdots+\lambda_n f_n(x)$是凸函数，$\lambda_i\geq 0, i=1,\cdots,n$
- 凸函数与仿射函数的复合：$g(x)=f(Ax+b)$
- 凸函数的逐点最大值pointwise maximization: $f(x)=\max(f_1(x),\cdots,f_n(x))$
- 凸函数的部分最小partial minimization:如果 $g(x,y) is\ convex\ in\ x,y$, and C is convex, then $f(x)=\min_{y\in C} g(x,y)\ is\ convex$

# 保持凸性的操作

- 一般的组合General composition: 设$f = h \circ g$, where $g: R^n \to R, h: R \to R, f: R^n \to R$. 则
  - $f$ 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 is concave
  - $f$ is concave if $h$ is concave and nonincreasing, g is convex
- 如何记住这些规则，考虑$n = 1$

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

# 凸集与凸函数的关系

- 函数$f(x)$的水平集 sublevel sets：$L_a=\{x|f(x)\leq a, x\in C\}$
- $f(x)$是凸函数，则其水平集都是凸集

证明：已知$f(x)$是凸函数，那么对于$\forall\, x_1, x_2 \in L_a$ 有

$$f(\lambda x_1+(1-\lambda)x_2)\leq \lambda f(x_1)+(1-\lambda)f(x_2)\leq a,$$

$$\lambda\in(0,1)$$

所以 $\lambda x_1+(1-\lambda)x_2\in L_a$

- 如果水平集是凸集，相应的函数$f(x)$是凸函数么？

哈尔滨工业大学 HARBIN INSTITUTE OF TECHNOLOGY

# 凸集与凸函数的关系

- 函数$f(x)$的上境图epigraph:

$$epi(f)=\{\begin{pmatrix}x\\y\end{pmatrix}|f(x)\le y),x\in C\}$$

- $f(x)$是凸函数$\Leftrightarrow epi(f)$ 为凸集

证明：**必要性**：已知$f(x)$是凸函数，那么$\forall x_1,x_2\in C$, 有

$$f(\lambda x_1+(1-\lambda)x_2)\le\lambda f(x_1)+(1-\lambda)f(x_2)$$

假设$\begin{pmatrix}x_1\\y\end{pmatrix}\in epi(f),\begin{pmatrix}x_2\\y\end{pmatrix}\in epi(f)$,即$f(x_1)\le y,f(x_2)\le y$,则 上式$\le$ $\lambda y+(1-\lambda)y=y$, 得证。

**充分性**：已知$epi(f)$ 为凸集，那么$\forall x_1,x_2\in C$, 有
$f(x_1)\le f(x_1),f(x_2)\le f(x_2)$, 有$(\lambda x_1+(1-\lambda)x_2,\lambda f(x_1)+(1-\lambda)f(x_2))\in epi(f)$
于是$f(\lambda x_1+(1-\lambda)x_2)\le\lambda f(x_1)+(1-\lambda)f(x_2)$，得证。

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

# 参考文献


- 《运筹学与凸优化问题》第二章
- Convex Optimization, Stephen Boyd and Lieven Vandenberghe,Cambridge University Press, 2016.Chapters 2 and 3
- J.P. Hiriart-Urruty and C. Lemarechal (1993), “Fundamentals of convex analysis”, Chapters A and B
- R. T. Rockafellar (1970), “Convex analysis”, Chapters 1 - 10,
- 最优化基础理论与方法 章节1.4
- 网课：最优化基础理论与方法