凸优化初步

郭兰哲

2016.10.25

主要内容

1.凸集基本概念

2.凸函数基本概念

3.凸优化问题一般提法

4.对偶

仿射集(Affine Set)

- 仿射集：集合内连接任意两点的直线也在集合内。即：

  $\forall x_1, x_2 \in C, \theta \in R \quad \Longrightarrow \quad \theta x_1 + (1-\theta) x_2 \in C$

- 仿射组合：设k个点 $x_1, x_2 \ldots \ldots x_k$，$\theta_1, \theta_2, \ldots \ldots \theta_k \in R$，$\theta_1 + \theta_1 + \cdots \theta_k = 1$，则称$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k \ x_k$为这$k$个点的仿射组合。

- 例子：直线，平面，超平面……

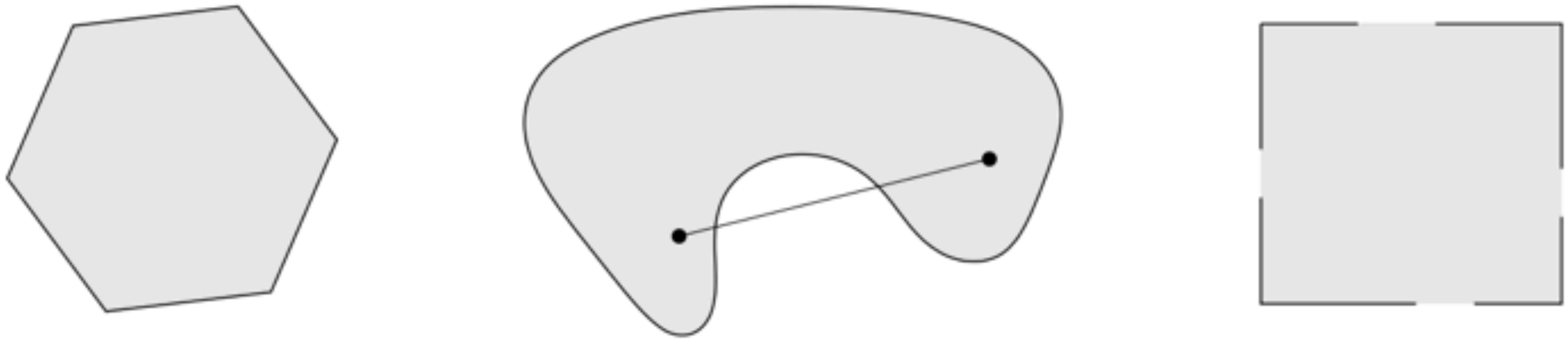
凸集(Convex Set)

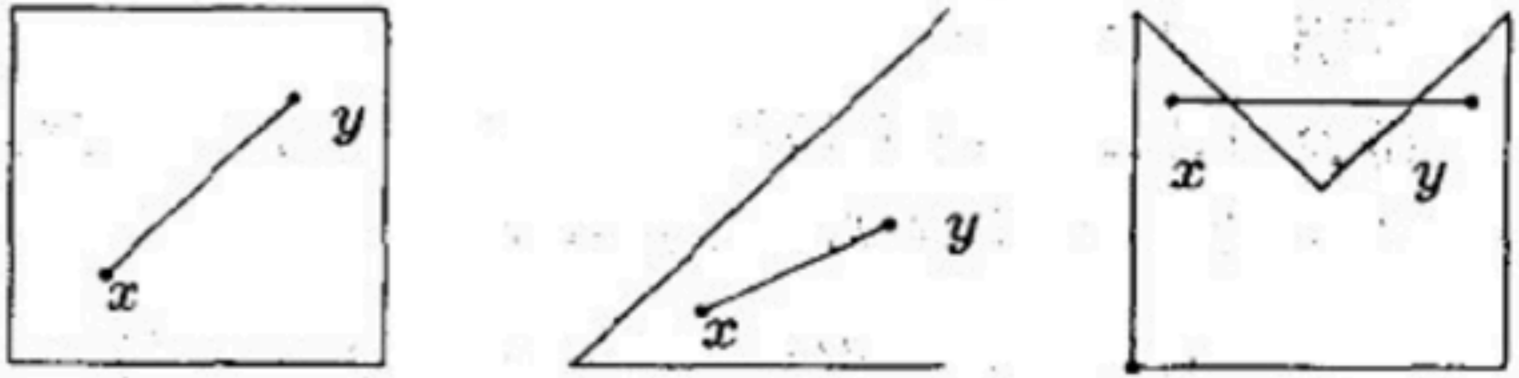
- 集合C内任意两点间的线段均在集合C内，则称集合C为凸集。

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

- 凸组合：设k个点 $x_1, x_2 \dots\dots x_k$， $\theta_1, \theta_2, \dots\dots \theta_k \in [0,1]$，$\theta_1 + \theta_1 + \cdots \theta_k = 1$，则称$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k\ x_k$为这$k$个点的凸组合。

- 凸包：包含集合C的最小的凸集称为凸包。





(a) 有界凸集　(b) 无界凸集　(c) 非凸集

超平面与半空间

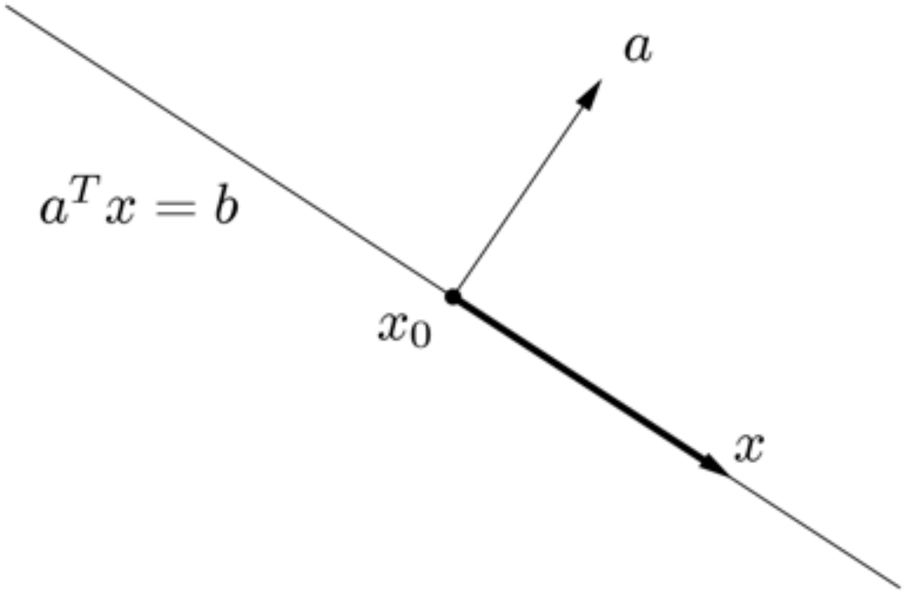
- 超平面 hyperplane

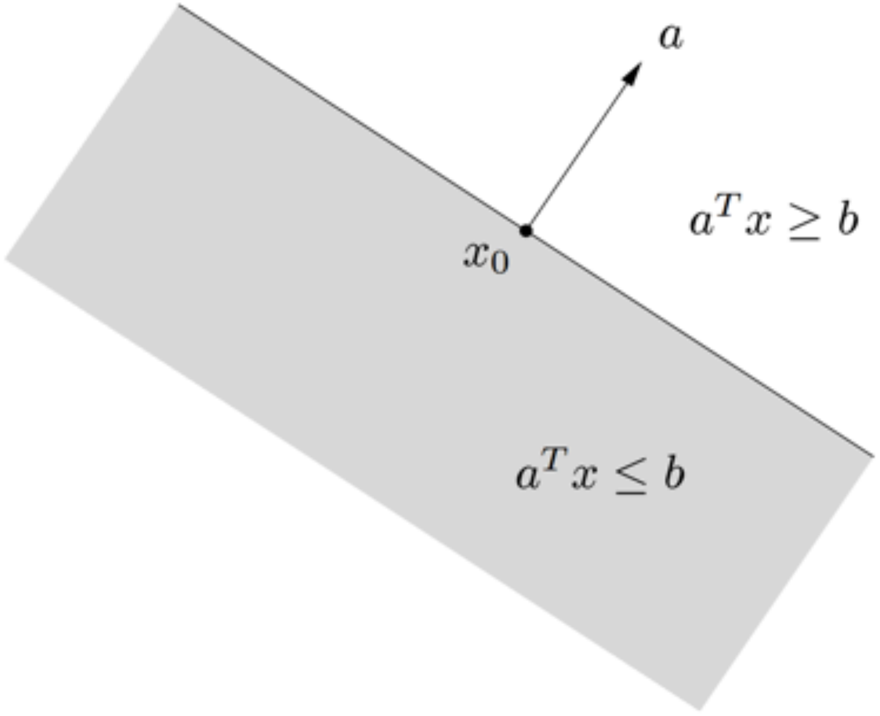
$$\{x \mid a^T x = b\}$$

- 半空间 halfspace

$$\{x \mid a^T x \le b\} \qquad \{x \mid a^T x \ge b\}$$

超平面和半空间





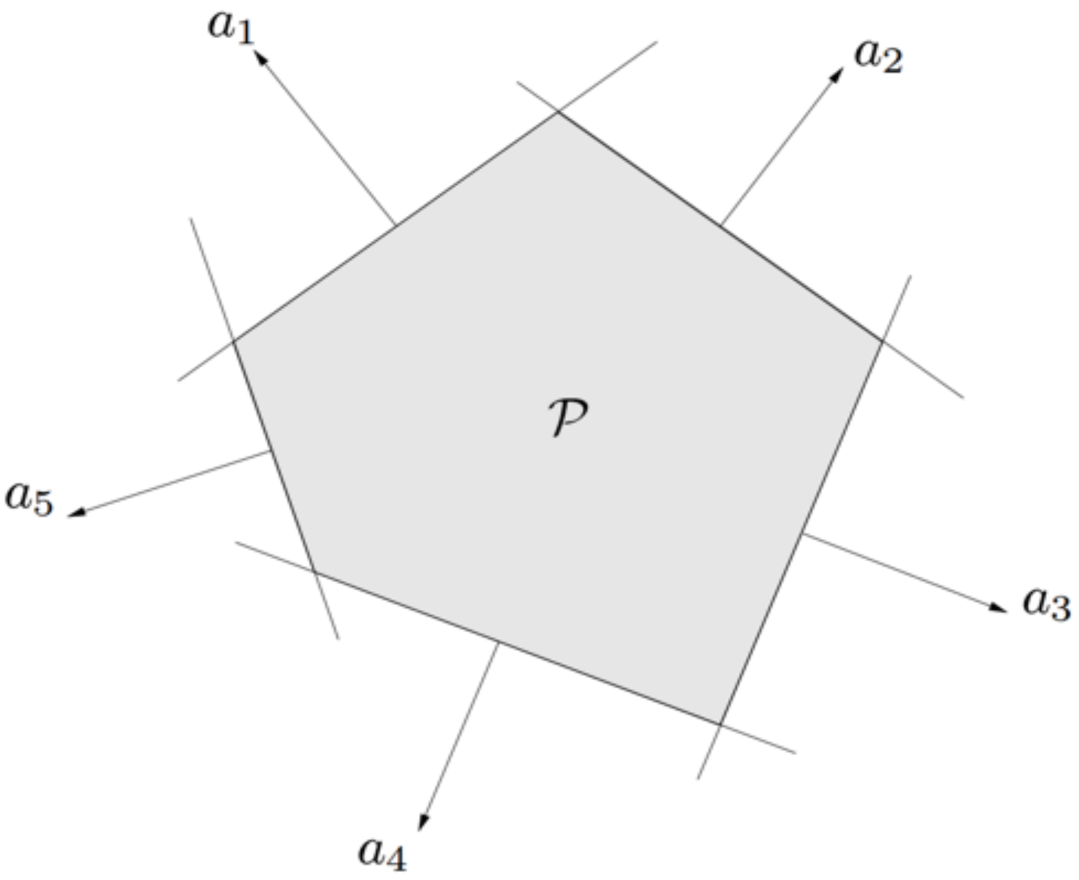
多面体

- 多面体：有限个半空间和超平面的交集。

$$P = \{x \mid a_j^T x \le b_j, c_i^T x = d_i\}$$

- 仿射集(如超平面、直线)、射线、线段、半空间都是多面体。
- 多面体是凸集。

多面体



保持凸性的运算

- 集合交运算
  - 根据定义容易证明。

- 仿射变换
  - 函数为f＝Ax + b的形式，称函数是仿射的，即线性函数加常数的形式。

- 透视变换

- 投射变换(线性分式变换)

仿射变换

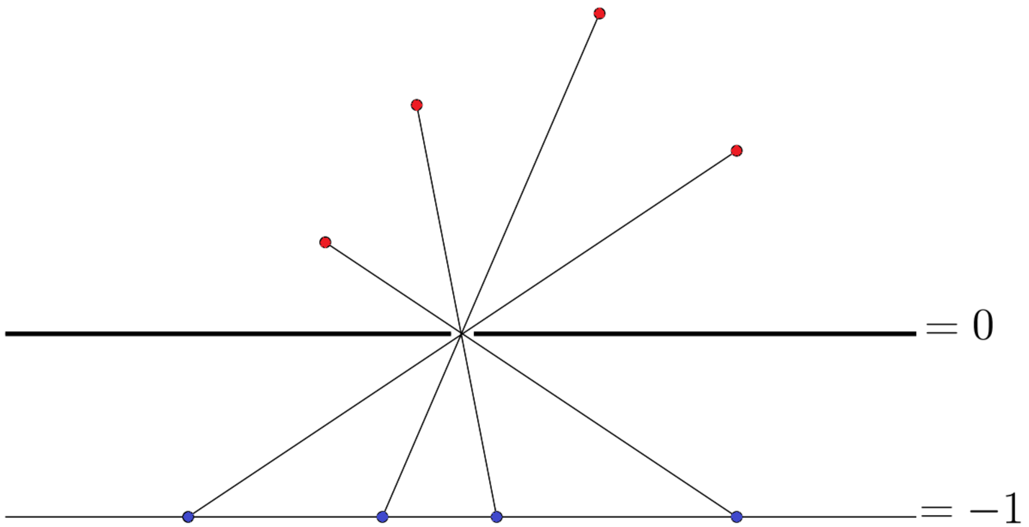
- 仿射变换 $f(x) = Ax + b,\ A \in \mathbf{R}^{m\times n}, b \in \mathbf{R}^m$
  - 伸缩、平移、投影
- 若f是仿射变换， $f : \mathbf{R}^n \to \mathbf{R}^m$ $f(S) = \{f(x) \mid x \in S\}$
  - 若S为凸集，则f(S)为凸集；
  - 若f(S)为凸集，则S为凸集。

透视变换

- 透视函数对向量进行伸缩(规范化)，使得最后一维的分量为1并舍弃之。

$$P : \mathbf{R}^{n+1} \to \mathbf{R}^n,\ P(z,t) = z/t$$

- 小孔成像



投射函数(线性分式函数)

- 定义f为线性分式函数：

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

- 线性分式函数是仿射函数和透视函数的复合。

- 若c = 0, d > 0，则 f 即为普通的仿射函数。

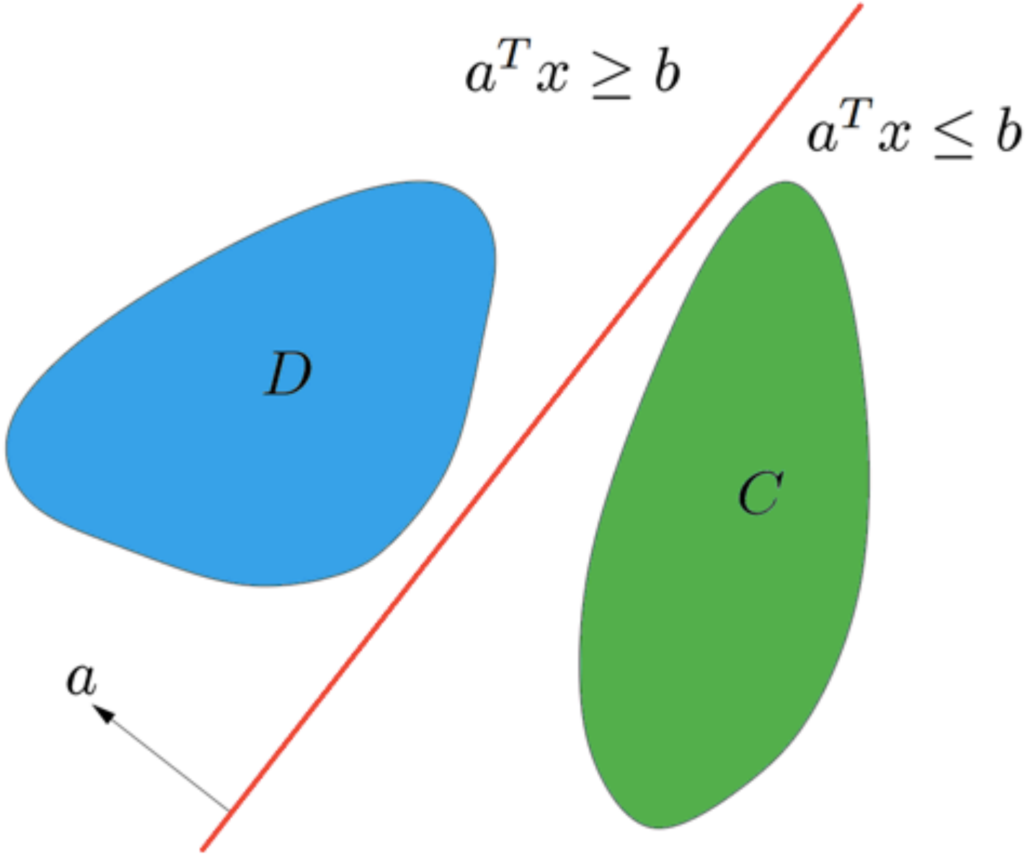
分割超平面

- 设C和D为两不相交的凸集，则存在超平面P，P可以将C和D分离。

$$\forall x \in C, a^T x \le b \text{且} \forall x \in D, a^T x \ge b$$

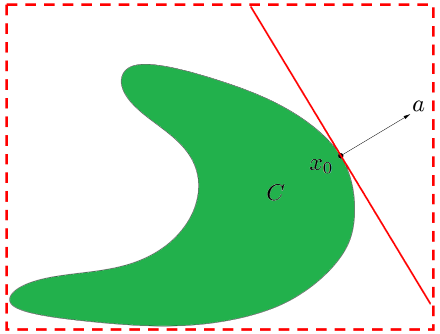
- 注意上式中可以取等号：
  - 所以：逆命题：“若两个凸集C和D的分割超平面存在，C和D不相交”为假命题。
  - 加强条件：若两个凸集至少有一个是开集，那么当且仅当存在分割超平面，它们不相交。

分割超平面



支撑超平面

- 设集合C，x0为C边界上的点。若存在a≠0，满足对任意x∈C，都有$a^T x \le a^T x_0$成立，
- 则称超平面$\{x|a^T x = a^T x_0\}$为集合C在点x0处的支撑超平面。



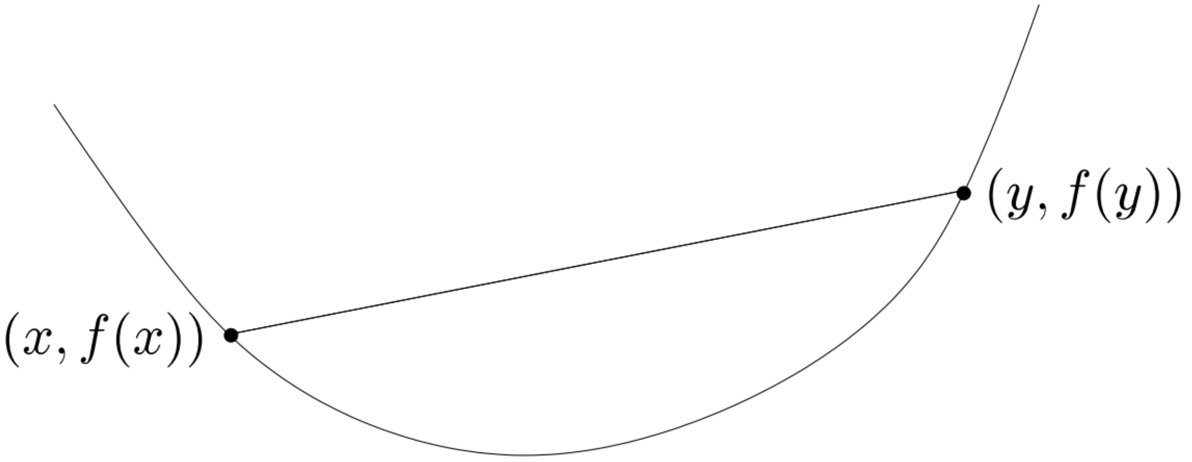
- 凸集边界上任意一点，均存在支撑超平面。
- 反之，若一个闭的非中空（内部点不为空）集合，在边界上的任意一点存在支撑超平面，
- 则称该集合为凸集。

凸函数

- 若函数f的定义域domf为凸集，且满足

$$\forall x, y \in \operatorname{dom} f, 0 \le \theta \le 1，\text{有}$$

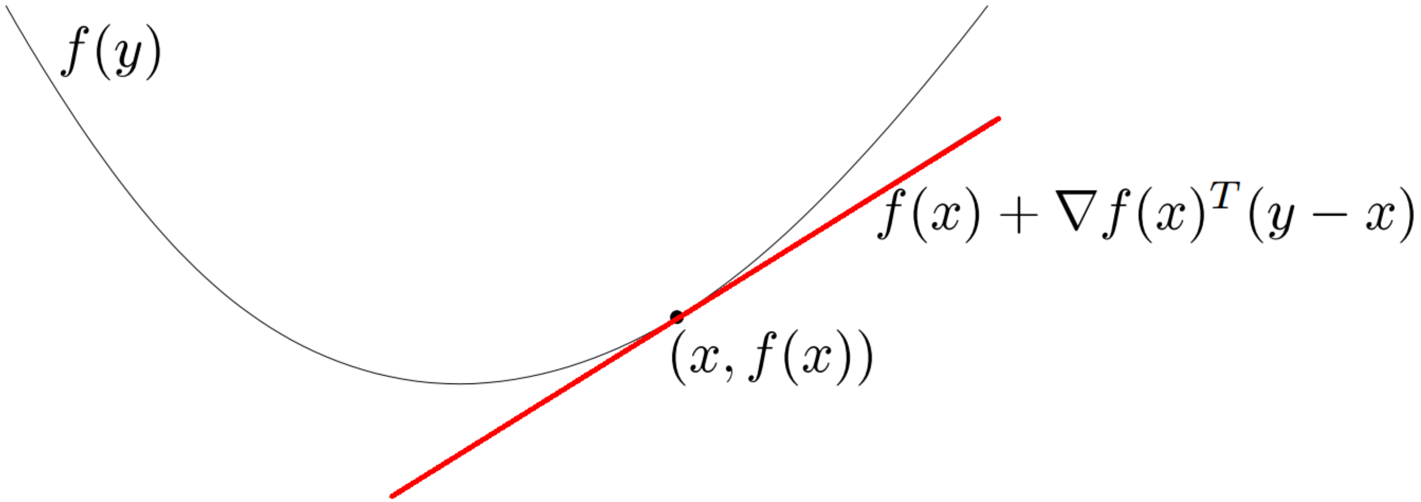
$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$



一阶条件

- 若f一阶可微(即其梯度在dom f内处处存在)，则函数f为凸函数当前仅当f的定义域dom f为凸集，且

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



一阶条件

- 对于凸函数，其一阶Taylor近似本质上是该函数的全局下估计。
- 该不等式说明从一个凸函数的**局部信息**(即它在某点的函数值和导数)，我们可以得到一些**全局信息**(如它的全局下估计)。这是凸函数最重要的性质。

二阶条件

- 若函数f二阶可微(即对于开集dom f内的任意一点，它的Hessian矩阵或者二阶导数存在)，则函数f为凸函数当且仅当dom f 为凸集，且

$$\nabla^2 f(x) \succ= 0$$

- 若f是一元函数，上式表示二阶导数大于等于0
- 若f是多元函数，上式表示其Hessian矩阵是半正定的。

常见凸函数

指数函数 e^{ax}

幂函数 $x^a, x \in R_+, a \geq 1 \text{ or } a \leq 0$

负对数函数 $-\log x$

负熵函数 $x \log x$

范数函数 $\|x\|_p$

下水平集

- 对于$f: R^n \to R$的$\alpha -$下水平集定义为：
- C_α={x $\in dom\ f | f(x) \le \alpha\}$

- 对于任意α值，凸函数的下水平集仍然时凸集。
- 证明：如果$x, y \in C_a$,则有$f(x) \le \alpha, f(y) \le \alpha$,因此对于任意的$0 \le \theta \le 1$，

- $\theta x + (1-\theta)y \in C_a$。

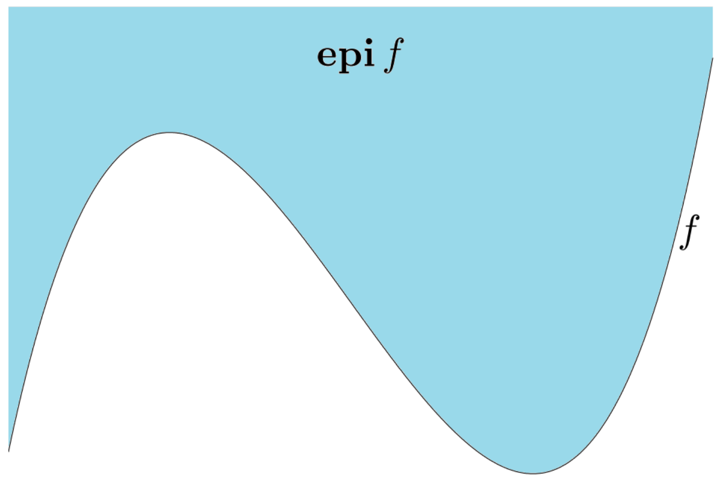
- 反过来则不一定正，某个函数的所有下水平集都是凸集，但这个函数可能不是凸函数。例如，$f(x) = -e^x$在R上不是凸函数(实际上，是严格凹函数)，但是其所有下水平集均为凸集。

上镜图

- 函数f的图像定义为： $\{(x, f(x))\,|\,x \in \textbf{dom}\, f\}$

- 函数f的上镜图(epigraph)定义为：

$$\textbf{epi}\, f = \{(x,t) \mid x \in \textbf{dom}\, f,\ f(x) \le t\}$$



凸函数与凸集的联系

- 一个函数是凸函数，当且仅当其上镜图是凸集。
- 一个函数是凹函数，当且仅当其亚图(hypograph)是凸集。

$$\mathbf{hypo}\, f = \{(x, t) \mid t \leq f(x)\}$$

Jensen不等式

- 基本Jensen不等式：

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

- 扩展：若f是凸函数 $\theta_1,\ldots,\theta_k\geq 0,\theta_1+\cdots+\theta_k=1$
- 则 $f(\theta_1x_1+\cdots+\theta_kx_k)\leq\theta_1f(x_1)+\cdots+\theta_kf(x_k)$

Jensen不等式证明

- 证明：$\sum_{i=1}^{n}\theta_i f(x_i) \geq f(\sum_{i=1}^{n}\theta_i x_i)$, 其中, $\sum_{i=1}^{n}\theta_i$=1
- a) 当n=1,2时，Jensen不等式显然成立。
- b)假设n=k时，Jensen不等式成立，即$\sum_{i=1}^{k}\theta_i f(x_i) \geq f(\sum_{i=1}^{k}\theta_i x_i)$, 其中, $\sum_{i=1}^{k}\theta_i$=1
- 则，当n = k+1时：
- $\sum_{i=1}^{k+1}\theta_i f(x_i)$ = $\sum_{i=1}^{k}\theta_i f(x_i) + \theta_{k+1} f(x_{k+1})$
  - $= \theta_{k+1} f(x_{k+1}) + Z_k \sum_{i=1}^{k}\frac{\theta_i}{Z_k} f(x_i)$, $\qquad Z_k = \sum_{i=1}^{k}\theta_i$
  - $\geq \theta_{k+1} f(x_{k+1}) + Z_k f\left(\sum_{i=1}^{k}\frac{\theta_i}{Z_k} x_i\right)$
  - $\geq f(\theta_{k+1}x_{k+1} + Z_k \sum_{i=1}^{k}\frac{\theta_i}{Z_k} x_i)$
  - $= f(\theta_{k+1}x_{k+1} + \sum_{i=1}^{k}\theta_i x_i)$
  - $= f(\sum_{i=1}^{k+1}\theta_i x_i)$,
- 说明，n=k+1时，Jensen不等式成立
- 综合a,b,可知，Jensen不等式成立

凸性+Jensen不等式：不等式理论的基础

- 考虑算术-几何平均不等式：

$$\frac{a+b}{2} \geq \sqrt{ab}, \quad a>0, b>0$$

- 证明：函数-log x 是凸函数，利用Jensen不等式，令$\theta = \frac{1}{2}$可得：
  - $-\log(\frac{a+b}{2}) \leq \frac{-log a - log b}{2}$
- 等式两边取指数即可。

保凸运算

- 凸函数的非负加权和 $$f(x) = \omega_1 f_1(x) + \ldots + \omega_n f_n(x)$$
- 凸函数与仿射函数的复合 $$g(x) = f(Ax + b)$$
- 凸函数逐点最大值、逐点上确界

$$f(x) = \max(f_1(x), \ldots, f_n(x))$$

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

共轭函数

- 设函数$f:R^n \to R$, 定义函数$f^*:R^n \to R$为:
- $f^*(y) = sup_{x\epsilon domf}\left(y^T x - f(x)\right)$
- 此函数称为函数f的共轭函数。
- 使上述上确界有限，即差值$y^T x - f(x)$在dom f有上界的所有y构成了共轭函数的定义域。
- 显然，f^*是凸函数(无论f是否是凸函数)，因为它是一系列y的凸函数(实质上是仿射函数)的逐点上确界。

共轭函数的性质

- (1)Fenchel不等式：从共轭函数的定义可以得到，对于任意x,y，如下不等式成立：
- $f(x)+f^*(y)\geq x^T y$

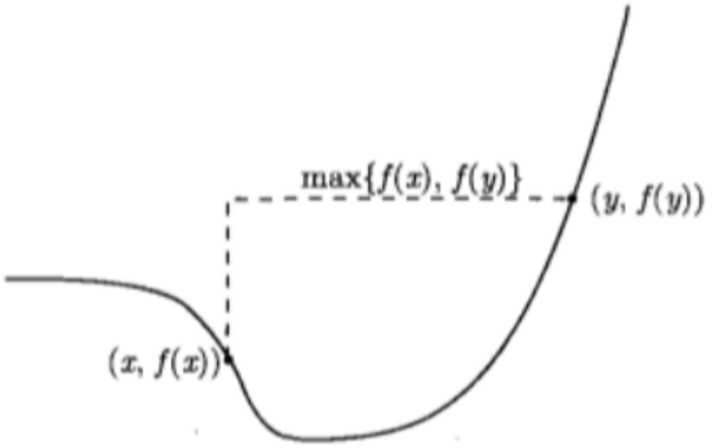
- (2)共轭的共轭：凸函数f共轭的共轭还是f 。

- (3)可微函数：可微函数f的共轭函数也称为f的Legendre变换。
- 设函数f是凸函数且可微，dom $f\in R^n$,使得$y^T x-f(x)$取最大的x^*满足$y=\nabla f(x^*)$,
- 反之，如果x^*满足$y=\nabla f(x^*)$，$y^T x-f(x)$在x^*取最大值。因此，对于$y=\nabla f(x^*)$，有：
- $f^*(y)=x^{*T}\nabla f(x^*)-f(x^*)$

拟凸函数

- 函数$f: R^n \to R$称为拟凸函数，如果其定义域及所有下水平集$S_\alpha = \{x_i \in dom\ f\ |f(x) \le \alpha\}$都是凸集。
- 凸函数具有凸的下水平集，所以也是拟凸函数

- 函数f是拟凸函数的充要条件： $dom\ f$ 是凸集，且对于任意的$x, y \in dom\ f$ 及$0 \le \theta \le 1$,有 ：
- $f(\theta x + (1-\theta)y) \le \max\{f(x), f(y)\}$
- 即线段中任意一点的函数值不超过其端点函数值中最大的那一个



优化问题的基本形式

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathbf{R}^n \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \\
& h_j(x) = 0, \quad j = 1, \dots, p
\end{aligned}
$$

优化变量　$x \in R^n$

不等式约束　$f_i(x) \le 0$

等式约束　$h_j(x) = 0.$

无约束优化　$m = p = 0$

凸优化问题的基本形式

$$\begin{aligned} \text{minimize} \quad & f_0(x), \quad x \in \mathbf{R}^n \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \dots, m \\ & h_j(x) = 0, \quad j = 1, \dots, p \end{aligned}$$

- $f_i(x)(0\le i \le m)$为凸函数，$h_j(x)(1\le i \le p)$为仿射函数。
- 凸优化问题的重要性质：
  - **可行域为凸集**
  - **局部最优解即为全局最优解**

局部最优解与全局最优解

- 证明：

- 设x是凸优化问题的局部最优解，即x是可行的，并且对于某些R，有：
- $f_0(x) = \inf\{f_0(z) \mid z \text{ 可行}, ||z - x||_2 \ \leq R \}$
- 现在，假设x不是全局最优解，即存在一个可行的y使得$f_0(y) < \ f_0(x)$，显然
- $||y - x||_2 \ > \mathrm{R}$。

- 考虑$z = (1 - \theta)x + \ \theta y$。根据可行集的凸性，z是可行的。并且，当$\theta$足够小时，z在x的邻域内。
- 根据目标函数的凸性，有：$f_0(z) \leq (1 - \theta)f_0(x) + \ \theta f_0(y) < \ f_0(x)$。出现矛盾。
- 所以，局部最优解x即是全局最优解。

凸优化问题最优解的微分条件

- 设凸优化问题的目标函数$f_0(x)$是可微的，对于所有的$x, y \in dom\ f_0$ 有：
- $f_0(y) \geq f_0(x) + \nabla f_0(x)^T(y - x)$
- 令X表示其可行集，那么x是最优解，当且仅当$x \in X$ 且
- $\nabla f_0(x)^T(y - x) \geq 0$，$\forall y \in X$

凸优化问题最优解的微分条件

- 证明：

- 首先，假设x ∈ X 满足$\nabla f_0(x)^T(y-x) \geq 0$，根据$f_0(y) \geq f_0(x) + \nabla f_0(x)^T(y-x)$，得到$f_0(y) \geq f_0(x)$，所以，x是最优解。

- 反之，若x是最优解但条件不成立，即存在$y \in X$，有$\nabla f_0(x)^T(y-x) < 0$。
- 考虑z(t) = ty + (1 – t)x，$t \in [0,1]$。可行集是凸集，因此z(t)可行。

- 因为，$\frac{d}{dt} f_0(z(t))\Big|_{t=0} = \nabla f_0(x)^T(y-x) < 0$
- 所以，对于小正数t，有$f_0(z(t)) < f_0(x)$。
- 出现矛盾，所以当x是最优解时条件一定成立。

拟凸优化问题

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathbf{R}^n \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_j(x) = 0, \quad j = 1, \ldots, p
\end{aligned}
$$

- 其中，不等式约束是凸的，而目标函数f_0是拟凸的(而不是凸优化问题中的凸目标函数)。
- 拟凸优化问题和凸优化问题本质上的不同在于拟凸优化的局部最优解不一定是全局最优解。

二分求解拟凸优化问题

- 可以用一族凸不等式来表示拟凸函数的下水平集，令$\phi_t(x)$为满足
- $f_0(x) \le t \Leftrightarrow \phi_t(x) \le 0$的一族凸函数，并且对于每一个$x$，$\phi_t(x)$都是$t$的非增函数。
- 用p^*表示拟凸优化问题的最优值，如果可行性问题：
- find x
- Subject to $\phi_t(x) \le 0$
  - $f_i(x) \le 0, \quad i = 1, \dots \dots$, m
  - $Ax = b$

- 是可行的，则有$p^* \le t$，不可行则有$p^* \ge t$。
- 因此，可以通过二分法来求解。

对偶问题

- 一般优化问题的Lagrange乘子法：

$$\begin{aligned} \text{minimize} \quad & f_0(x), \quad x \in \mathbf{R}^n \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p \end{aligned}$$

- Lagrange函数

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} v_j h_j(x)$$

- 对于固定的x，Lagrange函数L(x,λ,v)为关于λ和v的仿射函数

Lagrange对偶函数(dual function)

- Lagrange对偶函数

$$g(\lambda,\nu)=\inf_{x\in D} L(x,\lambda,\nu)=\inf_{x\in D}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\nu_i h_i(x)\right)$$

- 若没有下确界，定义： $g(\lambda,\nu)=-\infty$
- 根据定义，显然有：对∀λ≥0，∀v，若原优化问题有最优值p*，则

$$g(\lambda,\nu)\le p^*$$

- 进一步：Lagrange对偶函数为凹函数。

Lagrange对偶问题

- 对于任意一组(λ, ν)，其中$\lambda \geq 0$，Lagrange对偶函数给出了优化问题的最优值p^*的一个下界。
- 那么，一个自然的问题是：从 Lagrange 函数能够得到的最好下界是什么：

- 可以将这个问题表述为优化问题：
- $maximize\ g(\lambda, \nu)$
- $subject\ to\ \ \lambda \geq 0$

- 上述问题称为优化问题的Lagrange对偶问题，称该问题的最优解(λ^*, ν^*)为对偶最优解或者最优Lagrange乘子。

- Lagrange对偶问题是一个凸优化问题，这是因为极大化的目标函数是凹函数且约束集合是凸集。
- 因此，对偶问题的凸性与原问题是否是凸优化问题无关。

弱对偶性

- Lagrange对偶问题的最优值，用d^*表示，根据定义，这是通过Lagrange函数得到的原问题的最优值p^*的最好下界。特别的，我们有下面非常简单但非常重要的不等式：
- $d^* \leq p^*$
- 即使原问题不是凸问题，上述不等式亦成立，这个性质称为弱对偶性。
- 即使当d^*和p^*无限时，弱对偶性不等式依然成立。
- 定义差值$p^* - d^*$是原问题的最优对偶间隙。它给出了原问题最优值以及通过Lagrange对偶函数所能得到的最好下界之间的差值。最优对偶间隙总是非负的。
- 当原问题很难求解时，弱对偶不等式可以给出原问题最优值的一个下界，这是因为对偶问题总是凸问题，而且在很多情况下都能有效求解得到d^*。

强对偶性

- 如果等式$d^* = p^*$成立，即最优对偶间隙为0，那么强对偶性成立。
- 这说明从Lagrange对偶函数得到的最好下界是紧的。
- 对于一般情况，强对偶性不成立。但是如果原问题是凸问题，即可以表示为如下形式：

- $minimize \quad f_0(x)$
- Subject to $\quad f_i(x) \le 0, \quad i = 1, \dots\dots, \mathrm{m}$
- $\quad Ax = b$
- 其中函数f_0， $\dots\dots$， f_m是凸函数。

- 强对偶性通常(但不总是)成立，很多研究成果给出了强对偶性成立的条件(除了凸性条件外，这些条件称为约束准则。

Slater准则

- Slater条件：存在一点x∈ $relint\ D$ 使得下式成立：
- $f_i(x) < 0$，$i = 1, \ldots \ldots, m$　　Ax = b
- 满足上述条件的点称为严格可行。因为不等式约束严格成立。

- Slater定理：当Slater条件成立且原问题是凸问题时，强对偶性成立。

- 当不等式约束f_i中有一些是仿射函数时，Slater条件可以进一步改进：如果最前面的k个约束是仿射的，则下列弱化的条件成立时，强对偶性成立：存在一点x∈ $relint\ D$ 使得
- $f_i(x) \leq 0$，$i = 1, \ldots \ldots, k$　$f_i(x) < 0$，$i = k + 1, \ldots \ldots, m$　　Ax = b

- 换言之，仿射不等式不需要严格成立。

互补松弛性

- 若要对偶函数的最大值即为原问题的最小值，考察需要满足的条件：

$$
\begin{aligned}
f_0(x^\star) &= g(\lambda^\star, \nu^\star) \\
&= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \right) \\
&\le f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(x^\star) \\
&\le f_0(x^\star).
\end{aligned}
$$

互补松弛性

- 在上面的式子链中，两个不等式取等号。
- 那么，一个重要的结论是：$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$

- 因为求和项每一项都非正，因此有：$\lambda_i^* f_i(x^*) = 0$

- 上述条件称为互补松弛性。当强对偶性成立时，它对于任意原问题最优解x^*以及对偶问题最优解
- (λ^*, v^*)都成立。我们可以将互补松弛条件写为：

- $\lambda_i^* > 0 \Longrightarrow f_i(x^*) = 0$或者等价的：
- $f_i(x^*) < 0 \Longrightarrow \lambda_i^*=0$。

- 上式意味着在最优点处，除了第i个约束条件起作用的情况下，最优Lagrange乘子的第i项都为0.

Karush-Kuhn-Tucker (KKT)条件

- 假设函数$f_0, \dots\dots, f_m, h_1, \dots\dots, h_p$可微，但是不假设是凸函数。
- 令x^*和(λ^*, v^*)分别为原问题和对偶问题的最优解，对偶间隙为0，因为L(x, λ^*, v^*)关于x求极小在x^*处取得最小值，因此函数在x^*处的导数必须为0:
- $\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} v_i^* \nabla h_i(x^*) = 0.$
- 因此，有KKT条件：
- $f_i(x^*) \leq 0, \quad i = 1, \dots\dots, m$
- $h_i(x^*) = 0, \quad i = 1, \dots\dots, p$
- $\lambda_i^* \geq 0, \quad i = 1, \dots\dots, m$
- $\lambda_i^* f_i(x^*) = 0, \quad i = 1, \dots\dots, m$
- $\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} v_i^* \nabla h_i(x^*) = 0$

- 对于目标函数和约束函数可微的任意优化问题，如果强对偶性成立，那么任何一对原问题最优解和对偶问题最优解必须满足KKT条件。

鞍点解释

- 为表述方便，假设没有等式约束，只考虑不等式约束，结论可方便的扩展到等式约束。
- 注意到：

$$\sup_{\lambda\geq 0} L(x,\lambda)=\sup_{\lambda\geq 0}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)\right)=\begin{cases} f_0(x) & f_i(x)\leq 0, i=1,2\cdots,m \\ \infty & otherwise \end{cases}$$

- 假设x不可行，即存在某些i，使得$f_i(x) > 0$则选择λ_j=0，j!=i 以及$\lambda_i \to \infty$, 可以得出$sup_{\lambda\geq 0} L(x,\lambda) = \infty$。
- 反之，如果x可行，则有$f_i(x) \leq 0$，λ的最优选择为λ=0，此时$sup_{\lambda\geq 0} L(x,\lambda) = f_0(x)$

鞍点解释

- 这意味着我们可以将原问题的最优值写成如下形式：

$$\inf_{x} \sup_{\lambda \geq 0} L(x, \lambda)$$

- 而对偶问题是求对偶函数的最大值，即：

$$\sup_{\lambda \geq 0} \inf_{x} L(x, \lambda)$$

- 因此弱对偶性可以表示为如下不等式：

$$\sup_{\lambda \geq 0} \inf_{x} L(x, \lambda) \leq \inf_{x} \sup_{\lambda \geq 0} L(x, \lambda)$$

鞍点解释

- 事实上，上述不等式成立与L的性质无关，对于任意的f, 以及任意的$w \in R^n$，$z \in R^m$,
- 下式成立：
- $sup_z \ inf_w \ f(w,z) \le \ inf_w \ sup_z \quad f(w,z)$

- 这个一般性的不等式称为极大极小不等式，若等式成立，即：
- $sup_z \ inf_w \ f(w,z) = \ inf_w \ sup_z \quad f(w,z)$

- 则称f满足强极大极小性质或鞍点性质。

- 如果x^*和λ^*分别是原问题和对偶问题的最优点，且强对偶性成立，则它们是拉格朗日函数的一个鞍点。反过来同样成立：如果(x,λ)是拉格朗日函数的一个鞍点，那么x,λ分别是原问题和对偶问题的最优点，且最优对偶间隙为0。

证明： $\max_x \min_y f(x,y) \le \min_y \max_x f(x,y)$

- 对于任意的(x,y)∈dom f

$$f(x,y) \le \max_x f(x,y)$$

$$\Rightarrow \min_y f(x,y) \le \min_y \max_x f(x,y)$$

$$\Rightarrow \max_x \min_y f(x,y) \le \min_y \max_x f(x,y)$$

参考文献

- Convex Optimization, Stephen Boyd,Lieven Vandenberghe, Cambridge University Press, 2004

- 中译本：王书宁译，凸优化，清华大学出版社，2013