2. 凸集

仿射集：过集合$C \subseteq R^n$中任意两个不同点的直线仍在 C 中。直线、R^2是仿射集；线段、闭合图形不是仿射集。**仿射组合**：$\mathbf{1}^T\boldsymbol{\theta} = \mathbf{1}$，$\sum\theta_k x_k$是$x_1, \cdots x_k$的仿射组合。**仿射包**：构造尽可能小的仿射集合$\mathbf{aff}\boldsymbol{C} = \{\sum\theta_k x_k | \forall x_k \in \boldsymbol{C}, \mathbf{1}^T\boldsymbol{\theta} = \mathbf{1}\}$。
凸集：任意两点之间的线段仍在 C 中。$\forall x_1, x_2 \in C, \forall\theta \in [0,1], \theta x_1 + (1-\theta)x_2 \in C$。**凸组合**：$\mathbf{1}^T\boldsymbol{\theta} = \mathbf{1}, \theta_k \in [0,1]$，$\sum\theta_k x_k$是$x_1, \cdots x_k$的仿射组合。**凸包**：构造尽可能小的凸集合$\mathbf{Conv}\boldsymbol{C} = \{\sum\theta_k x_k | \forall x_k \in \boldsymbol{C}, \forall\theta_k \in [0,1], \mathbf{1}^T\boldsymbol{\theta} = \mathbf{1}\}$。
锥：$\forall x \in C, \theta \ge 0, \theta x \in C$。**凸锥**：$\forall x_1, x_2 \in C, \theta_1, \theta_2 \ge 0, \theta_1 x_1 + \theta_2 x_2 \in C$。**凸锥组合**：$\mathbf{1}^T\boldsymbol{\theta} = \mathbf{1}, \theta_k \ge 0$，$\sum\theta_k x_k$是$x_1, \cdots x_k$的凸锥组合。**凸锥包**：构造尽可能小的凸锥集合$\{\sum\theta_k x_k | \forall x_k \in \boldsymbol{C}, \forall\theta_k \ge 0\}$。
常见集合类型: $\emptyset, R^n, R^n$子空间是(a)仿射集、(b)凸集、(c)凸锥。任意直线 ab 过原点 c。任意线段 b 点 a 过原点 c。
超平面:$\{x | a^T x = b, x, a \in R^n, b \in R, a \ne 0\}$。超平面 ab 过原点 c 半空间 b 过原点 c。
范数:$\| x \| \ge 0; \| tx \| = t \| x \|; \| x + y \| \le \| x \| + \| y \|$。**范数球**:$\{x \,|\, \| x - x_c \| \le r\}$。**范数锥:** $\{(x,t) \,|\, \| x \| \le t\}$。
椭球:$\{x | (x - x_c)\boldsymbol{P}^{-1}(x - x_c) \le 1, \boldsymbol{P} \in \boldsymbol{S}_{++}^{\boldsymbol{n}}\}$球、椭球 b。
多面体:$\{x | a_i^T \boldsymbol{x} = b_i, c_j^T \boldsymbol{x} \le d_j, i = 1, \cdots, m, j = 1, \cdots, n\}$。**单纯形:**$\boldsymbol{R}^n$中选$v_0, \cdots, v_n$,n+1 个点,$v_1 - v_0, \cdots, v_k - v_0$线性无关单纯形为$\mathbf{Conv}\{v_0, \cdots, v_n\} = \{\sum\theta_n v_n | \theta_n \ge 0, \mathbf{1}^T\boldsymbol{\theta} = \mathbf{1}\}$。
对称矩阵集合$\boldsymbol{S}^{\boldsymbol{n}} = \{x \in \boldsymbol{R}^n \times \boldsymbol{R}^n | x = x^T\}$。**半正定矩阵集合**$\boldsymbol{S}_+^{\boldsymbol{n}} = \{x \in \boldsymbol{R}^n \times \boldsymbol{R}^n | x = x^T, x \succcurlyeq 0\}$。**正定矩阵集合**$\boldsymbol{S}_{++}^{\boldsymbol{n}} = \{x \in \boldsymbol{R}^n \times \boldsymbol{R}^n | x = x^T, x \succ 0\}$。$\boldsymbol{S}^{\boldsymbol{n}}c\boldsymbol{S}_+^{\boldsymbol{n}}$bc。
保凸运算（凸集经过某运算也为凸集）:交集、仿射函数($f(x) = Ax + b, f(x), b \in \boldsymbol{R}^m, A \in \boldsymbol{R}^{m \times n}$)、透视函数(对向量伸缩，使最后一维为 1 并舍弃，$P: \boldsymbol{R}^{n+1} \to \boldsymbol{R}^n, P(z,t) = \frac{z}{t}, z \in \boldsymbol{R}^n, t \in \boldsymbol{R}_{++}$)、线性分式(先仿射再透视，$f: \boldsymbol{R}^n \to \boldsymbol{R}^m, f(x) = \frac{Ax+b}{\boldsymbol{c}^T x + d}$)。
分离超平面定理:若$C \cap D = \emptyset$,则$\exists a \ne 0, b, \forall x \in C$有$a^T \le b$，$\forall x \in D$有$a^T \ge b$，$\{x | a^T x = b\}$称为分离超平面。
支撑超平面:$\boldsymbol{C} \subseteq R^n$,$x_0$是边界$\mathbf{bc}\,C$上一点，若$a \ne 0, \forall x \in C$有$a^T x \le a^T x_0$，则$\{x | a^T x = a^T x_0\}$称为支撑超平面。几何上说，支撑超平面切 C 于$x_0$，且半空间$\{x | a^T x = a^T x_0\}$包含 C。
支撑超平面定理:任意非空凸集在边界任何一点都有支撑超平面。
正常锥:锥$K \subseteq R^n$为正常锥，则 K 闭实尖（包含边界、内部无空、不包含直线）。**广义不等式**:$x \preccurlyeq_K y \Leftrightarrow y - x \in K$; $x \prec_K y \Leftrightarrow y - x = \mathbf{int}K$，性质和实数不等式相同。
最小元: $\forall y \in S$,均有$x \preccurlyeq_K y$，即$S \subseteq x + K$，也即$\forall\lambda \succ_{K^*} 0$,超平面$\{z | \lambda^T(z - x) = 0\}$是在$x$处对 S 的一个严格支撑超平面。
极小元: $y \in S, x \preccurlyeq_K y \Rightarrow y = x$，即$(x - K) \cap S = \{x\}$。如果$\lambda \succ_{K^*} 0$且 x 在$z \in S$上极小化$\lambda^T z$，那么 x 是极小的。
对偶锥:$K^* = \{y | y^T x \ge 0, x \in K\}$，几何上是与 K 内向量夹角不超过 90°的向量。对偶锥是闭凸锥；如果 K 是闭凸的，则$K^{**} = K$ 。$R_+^n, S_+^n, \{(x,t) | \| x \|_2 \le t\}$自对偶。
对偶广义不等式:若 K 是正常锥, 则K^*也是正常锥, 所以可以定义广义不等式的对偶$\preccurlyeq_{K^*}$。$x \preccurlyeq_K y, \text{iff } \forall\lambda \succcurlyeq_{K^*} 0, \text{then } \lambda^T x \le \lambda^T y$。$x \prec_K y, \text{iff } \forall\lambda \succcurlyeq_{K^*} 0, \lambda \ne 0, \text{then } \lambda^T x < \lambda^T y$。

3. 凸函数

凸函数:$\mathbf{dom}f$为凸函数，且$\forall x, y \in \mathbf{dom}f, \forall\theta \in [0,1]$,有$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$则$f: \boldsymbol{R}^n \to \boldsymbol{R}$是凸的。
扩展值延伸:定义凸函数定义域外值为∞,记作$\tilde{f}$。
可微凸函数一阶条件:f为凸函数⇔ $\mathbf{dom}f$为凸集，且$\forall x, y \in \mathbf{dom}f$有$f(y) \ge f(x) + \nabla f(x)^T(y - x)$。
二阶条件: f为凸函数⇔ $\forall x \in \mathbf{dom}f$,有$\nabla^2 f(x) \succcurlyeq 0$。
常见凸函数:二次函数($f(x) = \frac{1}{2}X^T P X + \delta^T X + r$, P 正定时凸)、仿射函数($\nabla^2 f(x) = 0_{n*n} \succcurlyeq 0$,既凸又凹)、指数函数、幂函数($f(x) = x^a, x \in R_{++}, \nabla^2 f(x) \ge 0 \Rightarrow a \ge 1 \; or \; a \le 0$时凸)、绝对值幂函数($f(x) = |x|^P, P \ge 1$时凸)、对数函数（凹）、负熵（$f(x) = xlogx$,凸）、范数（除零范数外凸）、极大值函数、解析逼近($f(x) = \log(\sum e^{x_k})$)、几何平均($f(x) = (\Pi x_k)^{1/n}, x \in R_{++}^n$，凹)。
判断凸函数:定义、$\boldsymbol{\nabla^2 f(x) \succcurlyeq 0}$、保凸运算:非负加权求和($\sum w_k f_k, w_k \ge 0,$)、复合仿射映射($f$凸⇒ $f(Ax + b)$凸)、逐点最大($f(x) = \max\{f_1(x), \dots, f_m(x)\}$)和逐点上确界、复合($f(x) = h(g(x))$,根据$f'' = h''(g)g'^2 + h'(g)g''$判断)、最小化(若$f$关于$(x,y)$时凸函数, C是非空凸集, 则$g(x) = \inf_{y \in C} f(x,y)$凸)、透视函数($f: R^n \to R$的透视函数$g: R^{n+1} \to R$定义为$g(x,t) = tf(\frac{x}{t})$)。
共轭函数:$f, f^*: \boldsymbol{R}^n \to \boldsymbol{R}, f^*(y) = \sup_{x \in \mathbf{dom}f}(y^T x - f(x))$，几何意义是线性函数$xy$和$f(x)$之间的最大差值。$f^*(y)$凸。若$f(x)$可微，则$f^*(y)$对应的$x$有$f'(x) = y$。如果函数$f$是闭凸的，则$f^{**} = f$。
拟凸/凹/线性函数:函数$f: \boldsymbol{R}^n \to \boldsymbol{R}$,若$\mathbf{dom}f$和所有下/上/上下水平集$S_\alpha = \{x \in \mathbf{dom}f | f(x) \le / \ge / = \alpha, \alpha \in R\}$，则称为拟凸/凹/线性函数。拟凸性要求每一个下水平集是一个区间（有可能包括无限区间）。凸函数⇒拟凸函数，反之不成立。
拟凸函数充要条件:$\mathbf{dom}f$为凸集，且$\forall x, y \in \mathbf{dom}f, \forall\theta \in [0,1]$,有$f(\theta x + (1-\theta)y) \le \max\{f(x), f(y)\}$，几何意义是线段任意一点函数值不超过端点函数值最大那个。

常见拟凸函数:向量的宽度(向量最后一个非零元素位置)、线性分式函数($f(x)=(a^x+b)/(c^Tx+d),\mathbf{dom}f=\{x|c^Tx+d>0\}$,拟线性)、向量零范数。

可微拟凸函数一阶条件: f为拟凸函数 $\Leftrightarrow\mathbf{dom}f$为凸集，且$\forall x,y\in\mathbf{dom}f$有$f(y)\le f(x)\Rightarrow\nabla f(x)^T(y-x)\le 0$。**二阶条件:** f为凸函数$\Leftrightarrow y^T\nabla f(x)=0\Rightarrow y^T\nabla^2 f(x)y\ge 0$。

对数凸凹函数:$f:\boldsymbol{R}^n\to\boldsymbol{R},\forall x\in\mathbf{dom}f$,有$f(x)>0$且$logf$凸/凹，则$f$是对数凸/凹的。或如此表达：$\mathbf{dom}f$为凸/凹集，$\forall x\in\mathbf{dom}f$,有$f(x)>0$，iff $\forall x,y\in\mathbf{dom}f,\forall\theta\in[0,1]$,有$f(\theta x+(1-\theta)y)\le/\ge f(x)^\theta f(y)^{1-\theta}$。**相关性质：**$\mathbf{dom}f$为凸集,有$\nabla^2 logf(x)=\frac{\nabla^2 f(x)}{f(x)}-\frac{\nabla f(x)\nabla f(x)^T}{f(x)^2}$,则$f$对数凸/凹,iff $\forall x\in\mathbf{dom}f$,有$f(x)\nabla^2 f(x)\succcurlyeq/\preccurlyeq\nabla f(x)\nabla f(x)^T$。乘积和正的伸缩性封闭。对数凸和封闭，对数凹的和一般不是对数凹。若$f:R^n\times R^m\to R$对数凹，则$g(x)=\int f(x,y)dy$在R^n上是x的对数凹。

广义不等式的凸性:$f:\boldsymbol{R}^n\to\boldsymbol{R}^m$，$\mathbf{dom}f$为凸集, $\forall x,y\in\mathbf{dom}f,\forall\theta\in[0,1]$且$f(\theta x+(1-\theta)y)\preccurlyeq_K\theta f(x)+(1-\theta)f(y)$，则$f$为K凸函数。

4. 凸优化问题

基本术语:标准形式问题: $\min f_0(x),s.t.f_i(x)\le 0,i=1,\cdots,m,h_i(x)=0,i=1,\cdots,p$;**优化变量:**$x\in R^n$;**目标函数:**$f_0:\boldsymbol{R}^n\to\boldsymbol{R}$; **不等式约束函数:**$f_i:\boldsymbol{R}^n\to\boldsymbol{R}$;**等式约束函数:**$h_i:\boldsymbol{R}^n\to\boldsymbol{R}$;**可行解集:**$x\in D$且满足约束;**最优值:**$p^*=\inf\{f_0(x)|f_i(x)\le 0,h_i(x)=0\}$,若无可行解，则$p^*=\infty$,若问题无下界，则$p^*=-\infty$;**最优解**:$x^*$可行且$f_0(x^*)=p^*$;**最优解集**:$X_{opt}=\{x|f_i(x)\le 0,h_i(x)=0,f_0(x)=p^*\}$;**$\epsilon$-次优解集:**$X_\epsilon=\{x|f_0(x)\le p^*+\epsilon\}$;**局部最优点:**$\exists R>0,\min f_0(z),s.t.f_i(x)\le 0,h_i(x)=0,\|z-x\|_2\le R$。**可行性优化问题:**find $x,s.t.f_i(x)\le 0,h_i(x)=0$。

等价问题:缩放($\min\tilde f(x)=\alpha_0 f_0(x),s.t.\tilde f_i(x)=\alpha_i f_i(x)\le 0,\tilde h_i(x)=\beta_i h_i(x)=0$)、变量变换($\phi:\boldsymbol{R}^n\to\boldsymbol{R}^n,D\subseteq\phi(\mathbf{dom}\phi)$ $\min\tilde f(x)=f_0(\phi(x)),s.t.\tilde f_i(x)=f_i(\phi(x))\le 0,\tilde h_i(x)=h_i(\phi(x))=0$)、松弛变量($f_i(x)\le 0\Rightarrow\exists s_i\ge 0,f_i(x)+s_i=0$,因此得到$\min f_0(x).s.t.\ s_i\ge 0,f_i(x)+s_i=0,h_i(x)=0$)、消除等式约束(参数化等式约束，$\phi:\boldsymbol{R}^k\to\boldsymbol{R}^n,\exists z\in\boldsymbol{R}^{\mathrm{k}},s.t.x=\phi(z)$则$\min\tilde f(z)=f_0(\phi(z)),s.t.\tilde f_i(z)=f_i(\phi(z))\le 0$)。

凸优化问题标准形式: $\min f_0(x)$，$s.t.f_i(x)\le 0,i=1,\cdots,m,a_i^T=b_i,i=1,\cdots,p$,其中$f_0,\cdots,f_m$为凸函数,等式约束仿射,可行集定义域为凸集。凸优化问题局部最优解也是全局最优解。

可微函数最优性准则:X是可行集，x是最优解$\Leftrightarrow x\in X$且$\nabla f_0(x)^T(y-x)\ge 0,\forall y\in X$，从几何上说若$\nabla f_0(x)\ne 0$，则$-\nabla f_0(x)$在$x$处定义了可行集的一个支撑超平面。

线性规划(LP):$\min c^Tx+d,s.t.Gx\preccurlyeq h,Ax=b$。可行解为多边形。**标准形式:**引入松弛变量$s\succcurlyeq 0$再$x=x^+-x^-,x^+,x^-\succcurlyeq 0$,从而得到$\min c^Tx^+-c^Tx^-+d,s.t.Gx^+-Gx^-+s=h,Ax^+-Ax^-=b,x^+,x^-,s\succcurlyeq 0$。

线性分式规划:多面体上极小化仿射函数之比，即$f_0(x)=c^Tx+d/(e^Tx+f),\mathbf{dom}f_0=\{x|e^Tx+f>0\}$,$\min f_0(x),s.t.Gx\preccurlyeq h,Ax=b$。为拟凸优化问题。令$y=x/(e^Tx+f),z=1/(e^Tx+f)$，则化为线性规划$\min c^Ty+dz,s.t.Gy\preccurlyeq hz,Ay=bz,e^Ty+fz=1,z\ge 0$。

二次规划:多面体上极小化凸二次函数，即$\min x^TPx/2+q^Tx+r$，$s.t.Gx\preccurlyeq h,Ax=b$。

二次约束二次规划(QCQP): $\min x^T P_0x/2+q_0^Tx+r_0$，$s.t.x^T P_ix/2+q_i^Tx+r_i\le 0,Ax=b$若$P_1,\dots,P_m\in\boldsymbol{S}_{++}^n$,可行区域为m个椭球和一个仿射集的交集。

广义不等式约束下凸优化问题:$f_0:\boldsymbol{R}^n\to\boldsymbol{R}$为凸函数，$f_i:\boldsymbol{R}^n\to\boldsymbol{R}^{k_i}K_i$为凸函数,$\min f_0(x),s.t.f_i(x)\preccurlyeq_{K_i}0,Ax=b$，具有标准凸优化问题性质。

半定规划(SDP):$F_i,G\in\boldsymbol{S}^k,\min c^Tx,s.t.\sum x_kF_k\preccurlyeq 0,Ax=b$。

二阶锥优化(SOCP):$A_i\in\boldsymbol{R}^{n_i\times n},F\in\boldsymbol{R}^{p\times n},\min f^Tx\ s.t.\|A_ix+b_i\|_2\le c_i^Tx+d_i,Fx=g$。若$n_i=0/c_i=0$则退化为LP/QCQP

5. 对偶

Lagrange 函数:对标准形式优化问题添加约束函数加权和得到$L(x,\lambda,\nu)=f_0(x)+\sum\lambda_if_i(x)+\sum\nu_ih_i(x)$。标准形式线性规划($\min c^Tx,s.t.Ax=b,x\succcurlyeq 0$))Lagrange 函数:$L(x,\lambda,\nu)=c^Tx-\lambda^Tx+\nu^T(Ax-b)=-b^T\nu+(c+A^T\nu-\lambda)^Tx$

Lagrange 对偶函数:$g:\boldsymbol{R}^m\times\boldsymbol{R}^p\to\boldsymbol{R}$为关于$x$取得的最小值，即对$\lambda\in\boldsymbol{R}^m,\nu\in\boldsymbol{R}^p$,有$g(\lambda,\nu)=\inf_{x\in D}L(x,\lambda,\nu)$。对偶函数为凹函数。对偶函数构成最优值$p^*$的下界，即$\forall\lambda\succcurlyeq 0,\nu$,有$g(\lambda,\nu)\le p^*$。标准形式线性规划对偶函数:$g(\lambda,\nu)=-b^T\nu+\inf(c+A^T\nu-\lambda)^Tx=-b^T\nu,\mathrm{if}\ A^T\nu-\lambda+c=0$

强/弱对偶性:原问题和对偶问题最优值分别是p^*和d^*,若$d^*\le p^*/d^*=p^*$,那么强/弱对偶性成立。p^*-d^*是对偶间隙。

Slater 条件:$\exists x\in\mathbf{relint}D,s.t.f_i(x)<0,Ax=b$.当 Slater 条件成立且原问题是凸问题时，强对偶性成立。

Lagrange 对偶问题:标准形式优化问题的 Lagrange 对偶问题是$\max g(\lambda,\nu),s.t.\lambda\succcurlyeq 0$。标准形式线性规划对偶问题:$\max -b^T\nu,s.t.A^T\nu-\lambda\succcurlyeq 0$

KKT 条件:令x^*和(λ^*,ν^*)分别是原问题和对偶问题的某对最优解。$L(x,\lambda^*,\nu^*)$关于x在x^*处导数为零，即$\nabla f_0(x^*)+\sum\lambda_i^*\nabla f_i(x^*)+\sum\nu_i^*\nabla h_i(x^*)=0$，再有$f_i(x^*)\le 0,h_i(x^*)=0,\lambda_i^*\ge 0,\lambda_i^*f_i(x^*)=0$，以上五式称为 KKT 条件。