1 Convex sets

凸优化学习笔记(1)——凸集

凸集

⋆ 凸集(convex sets)

如果在集合C中的任意两点满足：

$$\theta x_1 + (1-\theta)x_2 \in C$$

其中$0 \leq \theta \leq 1$，则集合C为凸集。

⋆ 重要例子

1) 超平面和半空间(hyperplanes and halfspaces)

超平面定义为$\{x|a^Tx = b\}$，半空间被定义为$\{x|a^Tx \leq b\}$。从直观上看，超平面在空间中为一块板子，划分的两边则分别为半空间。图1

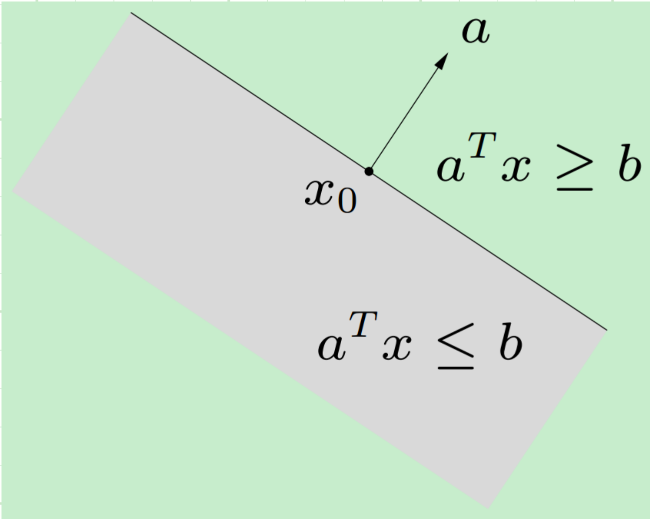

Figure 1: Hyperplanes and Halfplanes

2) 球和椭球

球的形式为

$$\{x| \|x - x_c\|_2 \leq r\} = \{x|(x-x_c)^T(x-x_c) \leq r^2\}$$

椭球的形式为

$$\{x|(x-x_c)^TP^{-1}(x-x_c) \leq 1\}$$

其中P是对称的正定矩阵。图2

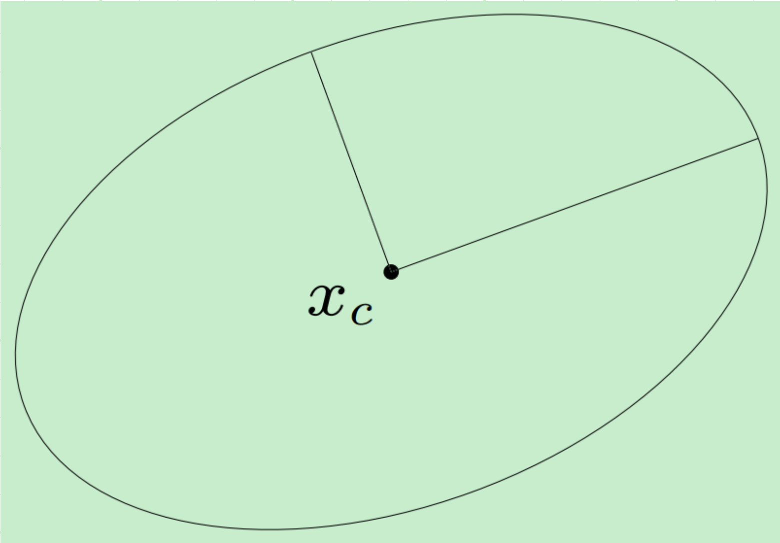

Figure 2: Ellipsoids

3) 范数球和范数锥

范数球为：

$$\{x \mid \|x - x_c\| \leq r\}$$

范数锥为：

$$\{(x,t) \mid \|x\| \leq t\}$$

图3

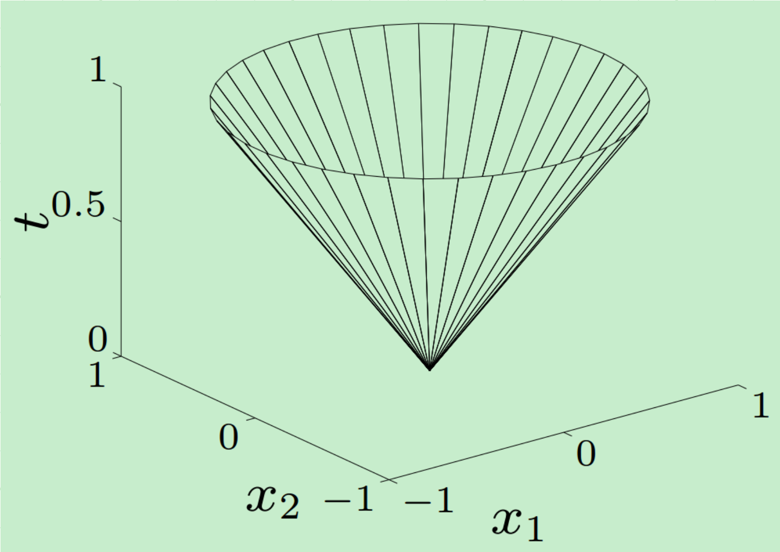

Figure 3: Norm cones

4) 多面体

$$\{x \mid a_j^T x \leq b_j, j = 1, \ldots, m, c_j^T x \leq d_j, j = 1, \ldots, p\}$$

图4

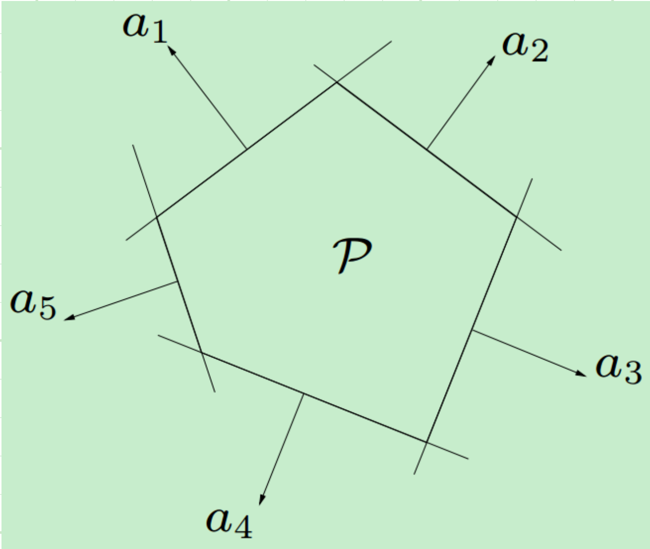

Figure 4: Polyhedra

5) 半正定锥

满足如下条件的集合$\mathbf{S}_+^n$是凸集：$\theta_1, \theta_2 \leq 0$并且$A, B \in \mathbf{S}_+^n$，则$\theta_1 A + \theta_2 B \in \mathbf{S}_+^n$。其中$\mathbf{S}_+^n$是半正定矩阵。图5

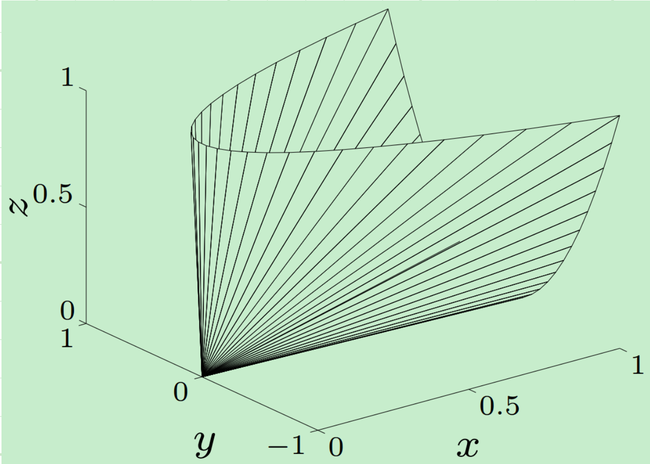

Figure 5: Positive semidefinite cone

⋆ 保凸运算

1) 交集

如果A与B均为凸集，则A与B的交集也为凸集。

2) 仿射函数

仿射函数即线性函数加常数。如果x为凸集，则$f(x) = Ax + b$为凸集。仿射函数的逆函数也保凸。

3) 线性分式以及透视函数

透视函数即$P(z,t)=z/t$，这里z是$n-1$维向量，t是最后一维分量。例如$P(x_1,x_2,x_3)=\{x_1/x_3,x_2/x_3\}$，$P$的定义域是正定对称矩阵。从几何上看，透视函数类似小孔成像，是从高维到低维的映射。图6

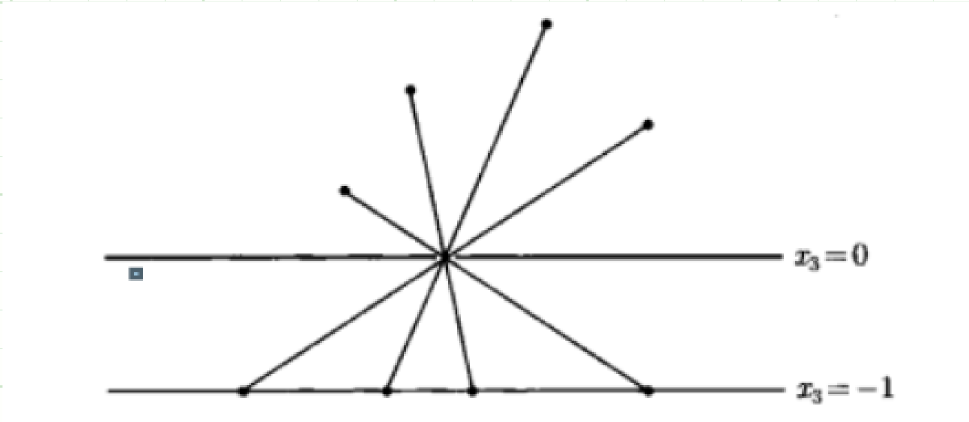

Figure 6: Perspective function

线性分式即$f(x)=(Ax+b)/(c^Tx+d)$其定义域为$\{x|c^T+d>0\}$。其逆函数也保凸。线性分式可看做在原集合内做拉伸，故而保凸。图7

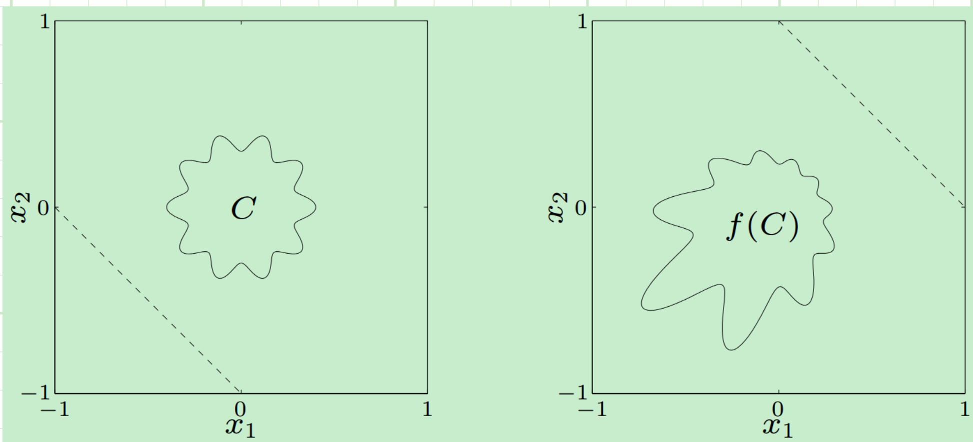

Figure 7: Linear-fractional function

⋆ 广义不等式

广义不等式即定义了拥有多个分量的变量之间的比较：

$$x \geq_k y \iff x-y \in k$$

$$x >_k y \iff x-y \in int(k)$$

$int(k)$即k的内部的眯。注意k必须为凸的、闭的、有非空内部且不包含直线。

⋆ minimum and minimal

Minimum即能和集合内所有点进行比较，且最小。Minimal即在集合内能比较的所有点中最小。图8

左图为minimum，右图为minimal。

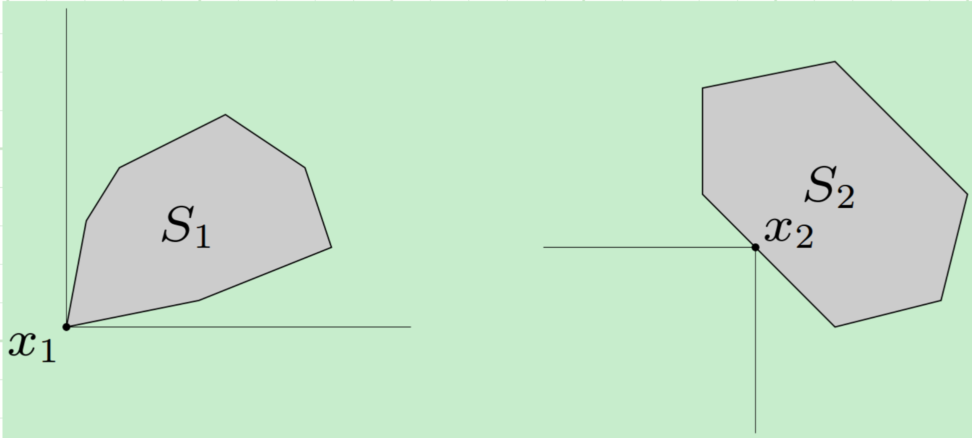

Figure 8: Minimum and minimal

⋆ 分割面与支撑面

分割面即能将两个集合分开的超平面，有严格不严格之分，严格即两个集合没有交点。两个凸集一定存在一个分割面。

支撑面即集合边缘有个点使得$a^T x \leq a^T x_0$成立。其中x是集合内的点，$a \neq 0$。

2 Convex functions

凸优化学习笔记(2)——凸函数

凸函数

⋆ 基本性质及例子

满足如下条件的从n维映射到1维的函数称凸函数：

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

其中$0 \leq \theta \leq 1$。凸函数的一维导数有如下性质：

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

可以注意到其实等号右边是f在x上的一阶泰勒展开式。另外，凸函数的二阶导数大于或等于0。凸函数举例如下：

1) 指数函数e^{ax}

2) 幂函数x^a，其中x大于或等于0，当a不属于0-1范围内时为凸函数，否则为凹函数

3) 绝对值的幂函数$|x|^a$，其中$a > 1$

4) 对数函数$\log x$

5) 负熵$x \log x$

6) 范数

7) 最大值函数max(x)

8) 二次-线性分式函数$f(x,y)=x^2/y$，y大于0

9) 指数和的对数$f(x)=\log(e^{x_1}+e^{x_2}+,\dots,e^{x_n})$，这其实是一个soft-max

10) 几何平均$f(x)=(\prod_{i=1}^n x_i)^{1/n}$

11) 行列式的对数$\log|A|$，其中A矩阵正定

另外还有两个概念。下水平集：即函数值小于某个阈值对应的定义域。上镜图：即$\{(x,t)|t\geq f(x)\}$，这是一片函数之上的区域。同时，凸函数满足jensen不等式。

⋆ 保凸运算

1) 非负加权和，即把多个函数乘上一个非负的权重加起来。

2) 复合仿射

$$g(x)=f(Ax+b)$$

如果f是凸的，则g是凸的。如果f是凹的，g也是凹的。即对自变量仿射后保凸。

3) 逐点最大，即多个函数中相同自变量取其中最大者，该运算保凸。

4) 复合

设$f(x)=h(g(x))$，则有如下情况：
如果h是凸函数且非减，g是凸函数，则f是凸函数。
如果h是凸函数且非增，g是凹函数，则f是凸函数。
如果h是凹函数且非减，g是凹函数，则f是凹函数。
如果h是凹函数且非增，g是凸函数，则f是凹函数。

f的拓展函数与上相同。

5) 下界

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

如果f在(x,y)中是凸的，则g也是凸的。

6) 透视函数

$$g(x,t)=tf(x/t)$$

其中t大于0。如果f是凸的，则g是凸的，如果f 是凹的，g也是凹的。

⋆ 共轭函数

共轭函数表达式是：

$$f^*(y)=\sup_x(y^Tx-f(x))$$

其中sup是求上界。其含义是，给定y时平面y^Tx到$f(x)$ 的最大距离。下图9的虚线即平面y^Tx，y^Tx到$f(x)$的最大距离出现在$f(x)$斜率为y的切线上。任何函数的共轭函数都是凸的，因为从表达中可以看到$f^*(y)$是y的仿射。

共轭函数的求法是令$xy-f(x)$，带入$f(x)$的表达式中并对x 求导，使求导结果等于0(即求$f(x)$与xy距离的最值)，把x 的结果回代进$f^*(y)=xy-f(x)$即可。

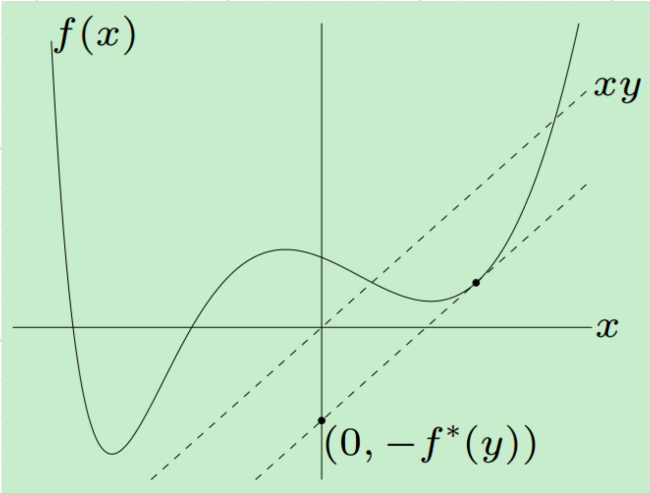

Figure 9: conjugate function

⋆ 拟凸函数

如果一个函数的下水平集(sublevel sets)是凸集，则该函数是拟凸函数。例如：

1) 对数函数$\log x$

2) 上取整函数$ceil(x) = \inf\{z \in Z | z \geq x\}$

3) 向量的长度

4) $f(x_1, x_2) = x_1 x_2$，由于该矩阵的Hessian矩阵不是正定，也不是负定，但是其下水平集是凸集。

5) 线性分式$f(x) = (a^T x + b)/(c^T x + d)$

对于一个函数来说，如果满足其非减，或非增，或存在一个占函数小于该点非增，大于该点非减，则该函数为拟凸的。图10

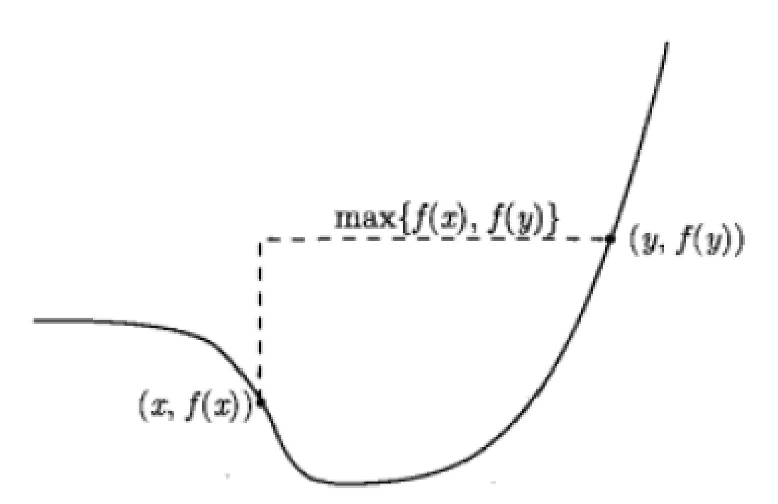

Figure 10: quasiconvex function

⋆ 对数凸函数

如果函数的值域大于0，并且$\log f(x)$是凸的，则该函数成对数凸函数。对数凹函数同理。例如：

1) 仿射函数

2) 幂函数$f(x) = x^a$，其中x大于0且a小于等于0。若a大于等于0则为对数凹函数

3) 指数函数

4) 高斯分布的累积分布函数

5) Gamma函数

6) 矩阵的行列式

7) 矩阵的行列式除以迹

需要注意，两个对数凸函数相乘依然是对数凸函数，相加却未必。两个对数凸函数的卷积运算是对数凸函数。对数凸函数的积分运算是对数凸函数。

3 Convex optimization problems

凸优化学习笔记(3)——优化问题

凸优化问题

⋆ 优化问题的基本形成

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq b_i, i = 1, \ldots, m \\
 & h_i(x) = 0, i = 1, \ldots, p
\end{array}
$$

需要注意的是除了显式的约束外，每个函数还有隐式的定义域约束。整个问题的定义域是所有函数的定义域的交集。对于每个这样的问题，其最优解定义为：

$$p^* = \inf\{f_0(x) | f_i(x) \leq 0, i = 1, \ldots, m, h_i(x) = 0, i = 1, \ldots, p\}$$

局部最优解定义为半径为**R**的定义域范围内的最小值。如果这是个最小问题，则$f_0(x)$称为损失函数，如果是最大问题，则称为效用函数。

⋆ 凸优化

凸优化问题定义为如下形式：

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m \\
 & a^T x = 0, i = 1, \ldots, p
\end{array}
$$

其中，$f_0, \ldots, f_m$都是凸函数。非正式地说，对于一个一般的优化问题，1)目标函数是凸函数；2)不等式约束是凸函数；3) 等式约束是仿射函数，则这个问题是凸优化问题。同时，从以上的定义可以注意到，凸优化问题的可行域一定是凸集。如果目标函数是拟凸函数，则这个问题是拟凸问题。

如果点x满足如下等式，则该点是最优解：

$$\nabla f(x)^T (y - x) \geq 0$$

体现在几何上，即图11

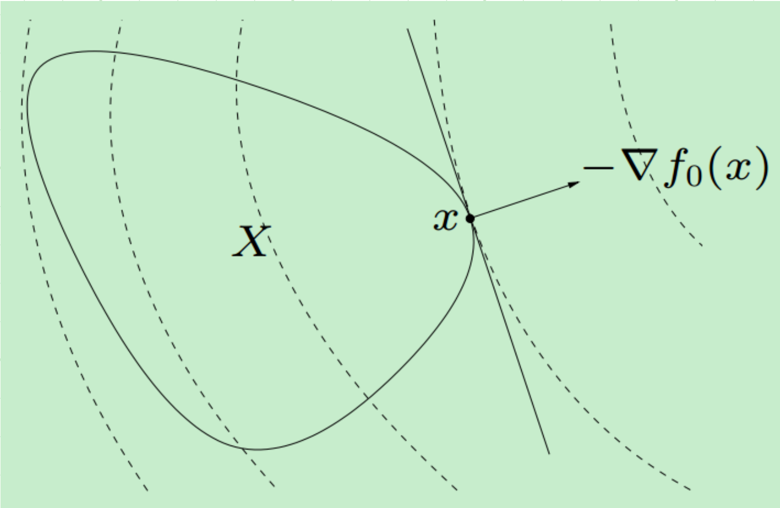

Figure 11: Convex optimization problems

有时间为了简化理论分析，可以将问题转化为线性的目标函数：

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m \\ & a^T x = 0, i = 1, \ldots, p \\ & f_0(x) - t \leq 0 \end{array}$$

同时，可以通过如下方式求解拟凸问题：

$$f_0(x) \leq \Longleftrightarrow \phi_t(x) \leq 0$$

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m \\ & a^T x = 0, i = 1, \ldots, p \\ & \phi_t(x) \leq 0 \end{array}$$

该方法通过在可行域上不停二分查找，找到一个恰好有可行域的t，并行解聘的x即为次优解。

⋆ 线性规划问题(LP)

问题可描述为：

如下问题可以转换为线性规划，图12：

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

1) 营养搭配问题，即每个食物有其价格和营养含量，目标是组合这些食物，在花费最少的情况下满足每一种营养需求。

2) 多边形的切比雪夫中心，即寻找多边形内半径最大圆的中心点。

3) 多个仿射函数最大值。

4) 分片线性极小化。

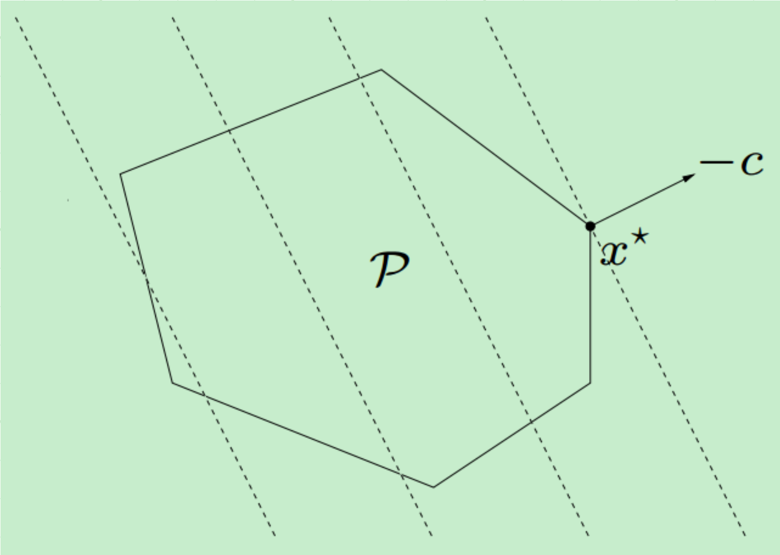

Figure 12: Linear program

★ 二次规划(QP)

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

其中P为正定矩阵。如果其不等式约束为二次约束，则该问题为二次约束的二次规划(QCQP)：

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i, i = 1, \ldots, m \\ & Ax = b \end{array}$$

二次规划有如下问题：

1) 最小二乘及回归

2) 求两个平面之间的距离

3) 求方差下界

4) x带随机损失的线性规划，即把随机损失的平方作为最小化项加入目标函数中

二阶锥规划(QCQP)：

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b\|_2 \leq c_i^T x + d \\ & Fx = g \end{array}$$

该规划可用于常数未知的线性规划，解法1是设置协方差矩阵P表示参数随机的程序，然后把约束条件设为容忍其极大损失；解法2是设置方差服从正态分布，通过0.95或者0.99等置信度确定约束范围。图13

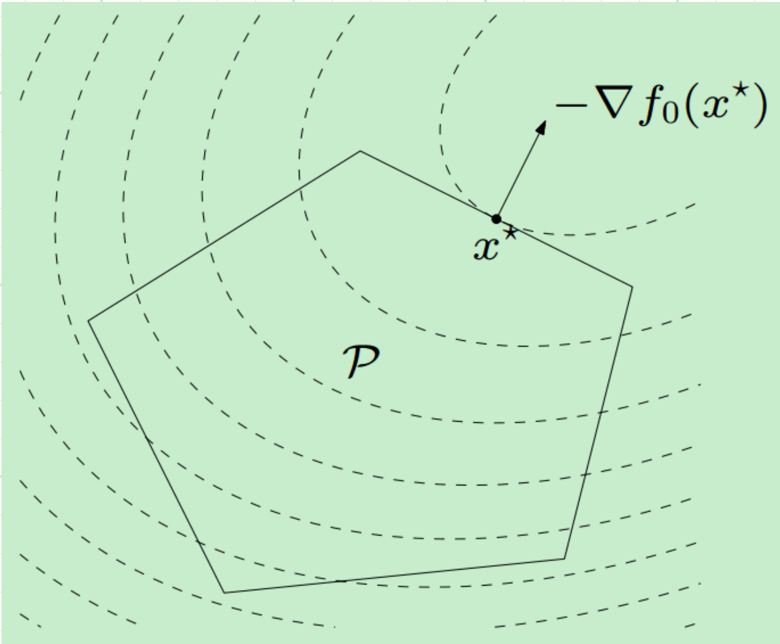

Figure 13: Quadratic program

⋆ 几何规划

单项式函数定义为：

$$f(x) = c x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}$$

多项式函数则定义为多个单项式函数的和：

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \ldots x_n^{a_{nk}}$$

其中x为正。定义几何规划为：

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 1, i = 1, \ldots, m \\ & h_i(x) = 1, i = 1, \ldots, p \end{array}$$

其中，f_i为多项函数，h_i为单项式函数。几何规划不是凸函数，但可以转换成凸函数，首先令$y_i = \log x_i$，然后在f_i外套log函数，则该问题的指数均转换为仿射函数。

⋆ 广义不等式约束

把约束扩展到广义不等式上，即认为f_i的映射结果是一个向量，因此f_i所在的不等式是一个广义不等式。这里记录一个半定规划：

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n \preceq G \\ & Ax = b \end{array}$$

可以看出，该问题包含SOCP，而SOCP包含QCQP，QCQP包含LP。