运筹学与优化方法

晁国清
计算机科学与技术学院

课程回顾

- 凸函数的定义

  f是定义在非空凸集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

- 常见凸函数

  指数函数、幂函数$x^a, a \geq 1 \ or \ a \leq 0 \ \ over \ R_+$, 仿射函数、二次函数、最小平方损失，范数、谱范数、核范数

- 凸函数的性质

  一阶特性、二阶特性

- 保持凸性的操作

  非负线性组合、凸函数与仿射函数的复合、逐点最大、部分最小，一般的复合

- 凸集与凸函数的关系

  水平集、上镜图

凸优化问题

- 凸优化问题
- 常见凸优化问题的类型

凸优化问题

最优化问题形式：$\begin{cases} \min\limits_{x \in D} f(x) \\ s.t.\ g_i(x) \leq 0, i = 1, \cdots, m \\ \quad h_j(x) = 0, j = 1, \cdots, l \end{cases}$

- 当f(x)是凸函数。可行域$S = \{x \in X | g_i(x) \leq 0, i = 1, \cdots, m, h_i(x) = 0, i = 1, \cdots, l\}$是凸集合，则上述优化问题为凸优化问题
- 当$f(x), g_i(x), i = 1, \cdots, m$是凸函数，$h_i(x), i = 1, \cdots, l$是线性函数，上述优化问题也称为凸优化问题

凸优化问题

- 局部最优解$\bar{x}$：$f(\bar{x}) \leq f(x), \forall x \in S \cap N_\epsilon(\bar{x})$
- 全局最优解x^*：$f(x^*) \leq f(x), \forall x \in S$
- 对于凸优化问题，局部最优解就是全局最优解？

反证法：

设$\bar{x}$是局部最优解，但不是全局最优解，即$\exists x^* \in S$使

"$f(x^*) < f(\bar{x})$“

对于$\lambda \in (0,1)$，$f\big(\bar{x} + \lambda(x^* - \bar{x})\big) = f(\lambda x^* + (1-\lambda)\bar{x})$

$$\leq \lambda f(x^*) + (1-\lambda) f(\bar{x})$$

$$< \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) = f(\bar{x})$$

当$\lambda \to 0$，　上面不等式与$\bar{x}$是局部最优解矛盾

凸优化问题的最优性条件几何解释

- $x^* \in S$ 是最优解$\Leftrightarrow \nabla f(x^*)^T(x-x^*) \geq 0, \forall x \in S$
- 几何解释

$x^* \in S$ 是最优解$\Leftrightarrow \nabla f(x^*)^T(x-x^*) \geq 0, \forall x \in S$

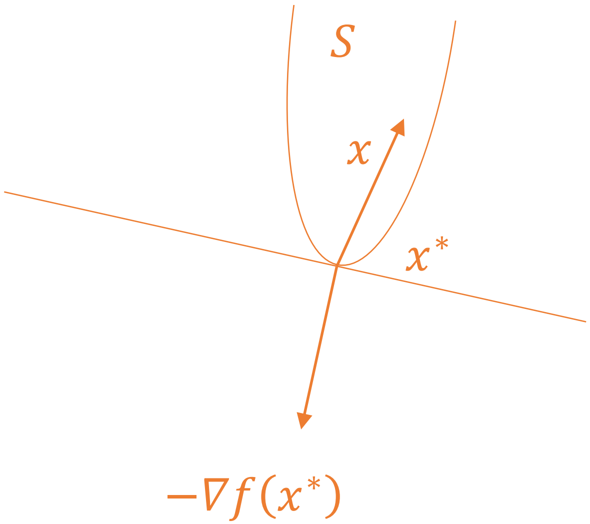
$\Leftrightarrow -\nabla f(x^*)^T x^* \geq -\nabla f(x^*)^T x, \forall x \in S$

令$\alpha = -\nabla f(x^*)$, 则有$\alpha^T x^* \geq \alpha^T x, \forall x \in S$

如果$\alpha \neq 0$ 即 $-\nabla f(x^*)^T \neq 0$

则$\mathrm{H} = \{x | -\nabla f(x^*)^T x = -\nabla f(x^*)^T x^*\}$就是集合$S$在$x^*$处的支撑超平面

如果$\alpha = 0$即 $-\nabla f(x^*)^T = 0$, 就无法确立支撑超平面



凸优化问题的最优性条件证明

➢ $x^* \in S$是最优解$\Leftrightarrow \nabla f(x^*)^T(x-x^*) \geq 0, \forall x \in S$

证明:

充分性: $\forall x \in S,\ f(x) \geq f(x^*) + \nabla f(x^*)^T(x-x^*) \geq f(x^*)$

必要性: 反证法

有$\exists \bar{x} \in S$使$\nabla f(x^*)^T(\bar{x}-x^*) < 0$

对于$\lambda \in [0,1]$, $f\left(x^* + \lambda(\bar{x}-x^*)\right) - f(x^*) = \lambda \nabla f(x^*)^T(\bar{x}-x^*) + \circ\,(\lambda\|\bar{x}-x^*\|)$

两边同除以λ, 然后令$\lambda \to 0$, 得到

$$\frac{f\left(x^* + \lambda(\bar{x}-x^*)\right) - f(x^*)}{\lambda} < 0$$

与$x^* \in S$ 是最优解矛盾

几种特殊凸优化问题的最优性条件

- 无约束凸优化问题$\min f(x), x \in R^n$的最优性条件：

$$x^* 是最优解 \Leftrightarrow \nabla f(x^*) = 0$$

证明：x^*是最优解 $\Leftrightarrow \nabla f(x^*)^T(x - x^*) \geq 0, \forall x \in R^n \Leftrightarrow \nabla f(x^*) = 0$

- 等式约束凸优化问题$\min\{f(x) | Ax = b\}$ 的最优性条件：

$$x^* 是最优解 \Leftrightarrow \nabla f(x^*) + A^T\mu = 0, \exists \mu$$

证明：x^*是最优解 $\Leftrightarrow \nabla f(x^*)^T(x - x^*) \geq 0, \forall x$满足$Ax = b, Ax^* = b$

$$\Leftrightarrow \nabla f(x^*)^T d = 0, \forall d \in N(A)$$

$$\Leftrightarrow \nabla f(x^*) \in N(A)^\perp = range(A^T)$$

$$\Leftrightarrow \nabla f(x^*) + A^T\mu = 0$$

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

几种特殊凸优化问题的最优性条件

- 非负约束凸优化问题$\min\{f(x)|x\geq 0\}$的最优性条件：

$$x^*\text{是最优解} \Leftrightarrow \nabla f(x^*)_i x_i^* = 0, x^* \geq 0, \nabla f(x^*) \geq 0$$

证明：
$$
\begin{aligned}
x^*\text{是最优解} &\Leftrightarrow \nabla f(x^*)^T(x-x^*) \geq 0, \forall x \geq 0, x^* \geq 0 \\
&\Leftrightarrow \nabla f(x^*)^T x \geq \nabla f(x^*)^T x^*, \forall x \geq 0, x^* \geq 0 \\
&\Leftrightarrow \nabla f(x^*) \geq 0, x^* \geq 0, \nabla f(x^*)^T x^* = 0 \\
&\Leftrightarrow \nabla f(x^*) \geq 0, x^* \geq 0, \nabla f(x^*)_i{}^T x_i^* = 0, i = 1, \cdots, n
\end{aligned}
$$

常见的凸优化问题

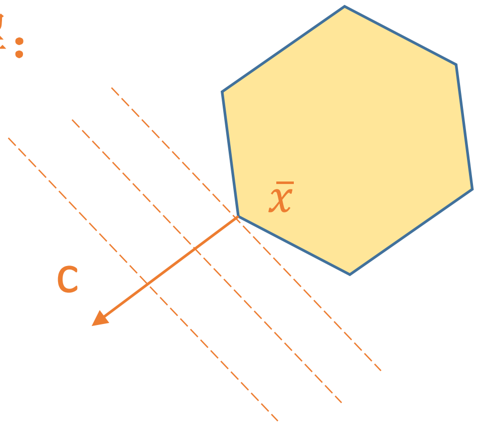
- 线性规划(Linear programming, LP)
- 二次规划(Quadratic programming, QP)
- 带二次约束的二次规划(Quadratically constrained quadratic program, QCQP)
- 二阶锥规划(Second-order cone program, SOCP)
- 半定规划(Semidefinite program, SDP)
- 几何规划(Geometric programming, GP)
- …

线性规划(Linear programming, LP)

- 线性规划的标准形式

$$\min c^T x$$
$$s.t.\ Ax = b$$
$$x \geq 0$$

- 单纯形方法求解：



如果线性规划问题存在最优解，则最优解一定在极值点(顶点)

找到极值点$\bar{x}$，判断是否最优，如果最优，结束。否则从当前极值点$\bar{x}$出发寻找一个更优的极值点，继续判断，…，一直到找到最优解或发现无解

凸二次规划(Quadratic programming, QP)

➢基本形式：

$$
\begin{aligned}
\min \frac{1}{2} x^T Q x + \mathrm{c}^T x \\
s.t.\ Ax = b \\
x \geq 0
\end{aligned}
$$

其中，$Q \succcurlyeq 0$

➢例子：

均值-方差模型

$$
\begin{cases}
\min \sum_{i,j=1}^{n} x_i x_j \sigma_{ij} \\
s.t.\ \sum_{i=1}^{n} x_i \mathrm{r}_i \geq \mu \\
\sum_{i=1}^{n} x_i = 1 \\
x_i \geq 0, i = 1, \cdots, n
\end{cases}
$$

最小二乘模型

$\min \|Ax - b\|_2^2$

➢ 求解方法：有效集法

带二次约束的二次规划(Quadratically constrained quadratic program, QCQP)

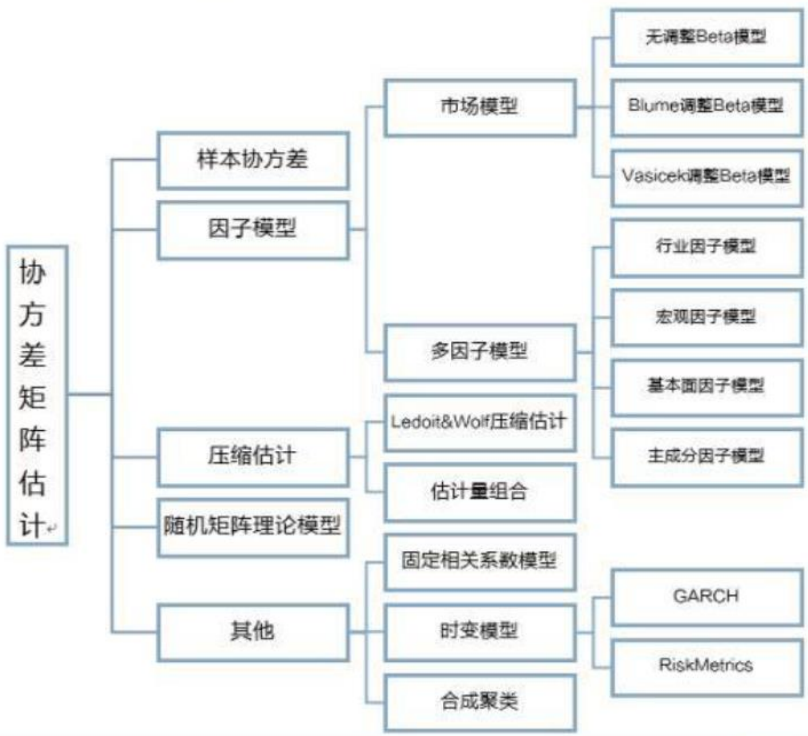
➢基本形式：

$$\min \ \frac{1}{2}x^T Q_0 x + c_0^T x$$
$$s.t. \ \frac{1}{2}x^T Q_i x + c_i^T x + b_i \leq 0, i = 1, \cdots, k$$
$$Ax = d$$

其中，$Q_i \succcurlyeq 0, i = 0,1,\cdots,k.$

➢例子：带多种风险约束的均值方差模型

图 1：协方差矩阵估计方法梳理



二阶锥规划(Second-order cone program, SOCP)

- 基本形式：

$$\begin{aligned} &\min \quad f^T x \\ &s.t. \quad \|A_i x + b_i\|_2 \le c_i^T x + d_i, i = 1, \cdots, m \\ &\qquad\quad Fx \le g \end{aligned}$$

$f \in R^n, A_i \in R^{n_i \times n}, b_i \in R^{n_i}, c_i \in R^n d_i \in R, F \in R^{p \times n}, g \in R^p, x \in R^n$

- 该问题之所以称为二阶锥规划是因为约束中$(Ax + b, c^T x + d)$为R^{m+1}空间中的二阶锥

- 当m=0，退化为LP

二阶锥规划例子

➢例子：一类特殊的概率约束问题$\min\{c^T x | P(\xi^T x \le a) \ge 95\%\}$，其中$\xi \sim N(\mu, Q)$

$$\xi^T x \sim N(\mu^T x, x^T Q x)$$

$$P(\xi^T x \le a) \ge 95\% \Leftrightarrow P\left(\frac{\xi^T x - \mu^T x}{\sqrt{x^T Q x}} \le \frac{a - \mu^T x}{\sqrt{x^T Q x}}\right) \ge 95\%$$

令$\bar{\xi} = \frac{\xi^T x - \mu^T x}{\sqrt{x^T Q x}}$，则有$\bar{\xi} \sim N(0, I)$

通过查阅标准正态分布的分位数表格可以查到对应于95%的一个值z_c

有$P\left(\bar{\xi} \le z_c\right) = 95\%$

概率约束$\Leftrightarrow \frac{a - \mu^T x}{\sqrt{x^T Q x}} \ge z_c$

$$\Leftrightarrow \sqrt{x^T Q x} \le \frac{1}{z_c}(a - \mu^T x)$$

$$\Leftrightarrow \|Lx\|_2 \le \frac{1}{z_c}(a - \mu^T x)$$

cholesky分解得到$Q = L^{\perp} L$，则有

$$x^T Q x = x^T L^{\perp} L x = \|Lx\|_2^2$$

半定规划(Semidefinite program, SDP)

- 半定规划的标准形式：

$$\begin{aligned} \min \quad & tr(CX) \\ s.t. \quad & tr(Q_i X) = b_i, i = 1, \cdots, m \\ & X \succcurlyeq 0 \end{aligned}$$

- 矩阵不等式形式：

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & x_1 Q_1 + \cdots + x_n Q_n \preccurlyeq Q_0 \\ & Ax = b \end{aligned}$$

线性矩阵不等式(Linear matrix inequality, LMI)

几何规划(Geometric program, GP)

- 几何规划的基本形式：

$$
\begin{aligned}
\min_x \quad & f(x) \\
s.t. \quad & g_i(x) \le 1, i = 1, \cdots, m \\
& h_j(x) = 1, j = 1, \cdots, r
\end{aligned}
$$

其中$f(x), g_i(x), i = 1, \cdots, m$ 都是正项多项式，$h_j(x), j = 1, \cdots, r$都是单项式

单项式：$\mathrm{f}: \mathrm{R}_{++}^{\mathrm{n}} \to R$, $f(x) = \gamma x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$, $\gamma > 0, a_1, \cdots, a_n \in R.$

正项多项式是单项式的求和：$f(x) = \sum_{k=1}^{p} \gamma_k x_1^{a_{k1}} x_2^{a_{k2}} \cdots x_n^{a_{kn}}$

该几何规划问题是非凸的

几何规划(Geometric program, GP)

- 对于单项式, $f(x)=\gamma x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$，令$y_i=\log x_i$， 则有

$$\gamma(e^{y_1})^{a_1}\ (e^{y_2})^{a_2}\ldots(e^{y_n})^{a_n}=e^{a^Ty+b}$$

$b=\log\gamma$

正项多项式就可以写作$\sum_{k=1}^{p}e^{a_k^Ty+b_k}$

通过上述变量转换，并且取log, 前面的几何规划问题就等价于

$$\min_{\mathrm{x}}\quad \log\left(\sum_{k=1}^{p_0}e^{a_0^Ty+b_{0k}}\right)$$

$$s.t.\ \ \log\left(\sum_{k=1}^{p_i}e^{a_{ik}^Ty+b_k}\leq 0\right), i=1,\cdots,m$$

$$c_j^Ty+d_j=0, j=1,\cdots,r$$

此优化问题提为凸优化问题

Log-sum-exp 凸函数

- Log-sum-exp函数：$g(x)=\log(\sum_{i=1}^{k}e^{a_i^Tx+b_i})$，对于固定的$a_i,b_i,i=1,\cdots,k$.
- 该函数也叫“soft max”，因为它光滑的近似$\max\limits_{i=1,\cdots,k}(a_i^Tx+b_i)$

对于凸性的证明：首先利用仿射组合规则，即$f(x)=\log(\sum_{i=1}^{n}e^{x_i})$
然后利用二阶特性证明$f(x)$的凸性

$$\nabla_i f(x)=\frac{e^{x_i}}{\sum_{i=1}^{n}e^{x_i}}$$

$$\nabla_{ij}^2 f(x)=\frac{e^{x_i}}{\sum_{i=1}^{n}e^{x_i}}1\{i=j\}-\frac{e^{x_i}e^{x_j}}{\left(\sum_{\ell=1}^{n}e^{x_\ell}\right)^2}$$

$\nabla^2 f(x)=diag(z)-zz^T, where\ z_i=\frac{e^{x_i}}{\sum_{\ell=1}^{n}e^{x\_\ell}}$. 该矩阵是对角占优矩阵，因此是半正定的

为什么log-sum-exp函数是softmax函数?

➢通过泰勒展开式对$f(x)=\log\left(\sum_{k=1}^{p}\exp x_i\right)$进行一阶近似

$f(x)\approx f(x_0)+f'(x)(x-x_0).$

对于$f(x)=\log(x+a)$来说，令$\mathrm{x}_0=x-a$，有

$f(x)=\log(x+a)\approx f(x-a)+f'(x-a)(x-x_0)=\log x+\frac{1}{x}\mathrm{a}$

因此$\log(x+a)\approx\log x+\frac{a}{x}$

相应地，$\log\left(\sum_{k=1}^{p}\exp(x_i)\right)\approx\log\left(\exp x_j\right)+\frac{\sum_{i\neq j}\exp x_i}{\exp x_j}$

不妨取$x_j=\max_i x_i$，则有$\log\left(\sum_{k=1}^{p}\exp(x_i)\right)>\max_i x_i$ （1）

由于$\sum_i x_i\leq n\times\ \max_i x_i$, 且$\sum_i x_i\geq x_i$

$\max\{x_1,\cdots,x_n\}=\log(\exp(\max_i x_i))$

$\leq\log(\exp x_1+\cdots+\exp x_n)=\ \log\left(\sum_{k=1}^{p}\exp(x_i)\right)$

$\leq\log(n\times\exp(\max_i x_i))$

$\leq\max\{x_1,\cdots,x_n\}+\log n$ （2）

$$\max_i x_i<\log\left(\sum_{k=1}^{p}\exp(x_i)\right)\leq\max\{x_1,\cdots,x_n\}+\log n$$

参考文献

- 运筹学与最优化方法， Page 32
- Convex Optimization, Stephen Boyd and Lieven Vandenberghe, Cambridge University Press, 2016. Chapters 4 and 3
- 网课：最优化理论与方法
- 网课： Convex Optimization