带约束凸优化问题

- 前述问题都可以写为

$$
\begin{aligned}
\min_{x\in\mathcal{D}} \quad & f(x), \\
\text{s.t.} \quad & c_i(x) \leqslant 0, \quad i = 1, 2, \cdots, m, \\
& Ax = b,
\end{aligned}
$$

  其中$f(x)$ 为适当的凸函数, $\forall i, c_i(x)$ 是凸函数且$\mathrm{dom} c_i = \mathbb{R}^n$.

- $A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p$ 是已知的.
- 集合$\mathcal{D}$ 表示自变量x 的自然定义域, 即

$$\mathcal{D} = \mathrm{dom} f = \{x \mid f(x) < +\infty\}.$$

- 自变量x 还受约束的限制, 定义可行域

$$\mathcal{X} = \{x \in \mathcal{D} : c_i(x) \le 0, i = 1, 2, \cdots, m; Ax = b\}.$$

- 由于凸优化问题的可行域是凸集, 因此等式约束只能是线性约束.

Slater约束品性与强对偶原理

- 凸优化问题有很多好的性质.一个自然的问题是:我们能否像研究无约束问题那样找到该问题最优解的一阶充要条件?如果这样的条件存在, 它在什么样的约束品性下成立?
- 在通常情况下, 优化问题的对偶间隙大于0, 即强对偶原理不满足.
- 但对很多凸优化问题, 在特定约束品性下可以证明强对偶原理.
- 直观的一种约束品性是存在满足所有约束条件的严格可行解.

首先给出集合$\mathcal{D}$ 的相对内点集relint$\mathcal{D}$ 的定义.给定集合$\mathcal{D}$, 记其仿射包为

$$\textbf{affine}\,\mathcal{D} = \{x \mid x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k,\ x_1, x_2, \cdots, x_k \in \mathcal{D},\ \sum_{i=1}^{k} \theta_i = 1\}.$$

定义 (相对内点)

集合$\mathcal{D}$ 的相对内点集定义为

$$\mathrm{relint}\mathcal{D} = \{x \in \mathcal{D} \mid \exists\, r > 0,\ \text{使得}\ B(x, r) \cap \textbf{affine}\,\mathcal{D} \subseteq \mathcal{D}\}.$$

相对内点是内点的推广，若$\mathcal{D}$ 本身的“维数”较低, 则$\mathcal{D}$ 不可能有内点, 但如果在它的仿射包**affine** $\mathcal{D}$ 中考虑, 则$\mathcal{D}$ 可能有相对内点.

Slater约束品性

定义 (Slater约束品性)

若对凸优化问题

$$\min_{x\in\mathcal{D}} \quad f(x), \ \text{s.t. } c_i(x) \leqslant 0, \ i = 1, 2, \cdots, m, \quad Ax = b,$$

存在$x \in \mathrm{relint}\mathcal{D}$ 满足

$$c_i(x) < 0, \quad i = 1, 2, \cdots, m, \quad Ax = b,$$

则称对此问题Slater 约束品性满足.该约束品性也称为 Slater 条件.

- Slater约束品性实际上是要求自然定义域$\mathcal{D}$ 的相对内点中存在使得不等式约束严格成立的点, **affine** $\mathcal{D} = \mathbb{R}^n$ 时相对内点就是内点.
- 不等式约束是仿射函数时, Slater条件可以放宽.设前k 个不等式约束是仿射的, 此时Slater约束品性变为:存在$x \in \mathrm{relint}\mathcal{D}$, 满足

$$c_i(x) \le 0, \ i = 1, 2, \cdots, k; \quad c_i(x) < 0, \ i = k+1, k+2 \cdots, m; \quad Ax = b,$$

即对线性不等式约束无需要求其存在严格可行点.

定理

若凸优化问题满足*Slater*条件，则强对偶原理成立.

- 当$d^* > -\infty$ 时，对偶问题的最优解可以取到，即存在对偶可行解(λ^*, ν^*)，满足$g(\lambda^*, \nu^*) = d^* = p^*$.
- 假设集合$\mathcal{D}$ 内部非空(即$\mathrm{relint}\mathcal{D} = \mathbf{int}\,\mathcal{D}$)，A 行满秩(否则可以去掉多余的线性等式约束)以及原始问题最优函数值p^* 有限.
- 定义集合

$$
\begin{aligned}
\mathbb{A} = \{(u, v, t) \mid \exists x \in \mathcal{D},\ c_i(x) \le u_i,\ i = 1, 2, \cdots, m,\\
Ax - b = v, f(x) \le t\}.\\
\mathbb{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid s < p^*\}.
\end{aligned}
$$

- 可以证明集合$\mathbb{A}$ 和$\mathbb{B}$ 是不相交的.
- 假设$(u, v, t) \in \mathbb{A} \cap \mathbb{B}$.根据$(u, v, t) \in \mathbb{B}$，有$u = 0, v = 0$ 和$t < p^*$.
- 由$(u, v, t) \in \mathbb{A}$，可知$f(x) \le t < p^*$，这与p^* 是原始问题最优值矛盾.

Slater约束品性与强对偶原理

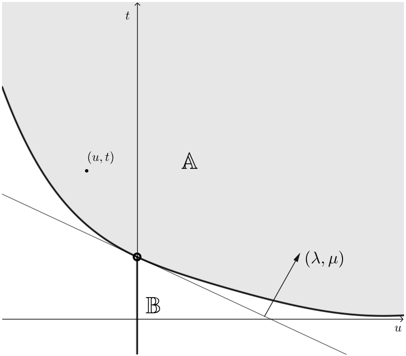

Figure: 集合$\mathbb{A}$ 和$\mathbb{B}$ 在u - t 方向投影的示意图
($\mathbb{A}$ 一般为有内点的凸集, $\mathbb{B}$ 是一条射线且不含点$(0,0,p^*)$)

因为$\mathbb{A}$ 和$\mathbb{B}$ 均为凸集, 由超平面分离定理, 存在$(\lambda,\nu,\mu)\neq 0$ 和α, 使得

$$\lambda^{\mathrm{T}}u+\nu^{\mathrm{T}}v+\mu t\geq\alpha,\quad\forall\,(u,v,t)\in\mathbb{A},$$

$$\lambda^{\mathrm{T}}u+\nu^{\mathrm{T}}v+\mu t\leq\alpha,\quad\forall\,(u,v,t)\in\mathbb{B}.$$

Slater约束品性与强对偶原理

- 我们断言$\lambda \geq 0$ 和$\mu \geq 0$ (否则可以取u_i 和t 为任意大的正实数以及$\nu = 0$, 这会导致$\lambda^{\mathrm{T}} u + \mu t$ 在集合$\mathbb{A}$ 上无下界).
- 同时, 由于$\mu t \leq \alpha$ 对于所有$t < p^*$ 成立, 可得$\mu p^* \leq \alpha$.
- 对任意$x \in \mathcal{D}$, 取$(u, v, t) = (c_i(x), Ax - b, f(x)) \in \mathbb{A}$, 可知

$$\sum_{i=1}^{m} \lambda_i c_i(x) + \nu^{\mathrm{T}}(Ax - b) + \mu f(x) \geq \alpha \geq \mu p^*.$$

- 假设$\mu > 0$, 则

$$L(x, \frac{\lambda}{\mu}, \frac{\nu}{\mu}) \geq p^*.$$

  进一步地, 我们有$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) \geq p^*$, 根据弱对偶性$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) \leq p^*$ 自然成立.因此, 必有$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) = p^*$ 成立.说明在此情况下强对偶性满足, 且对偶最优解可以达到.

Slater约束品性与强对偶原理

- 考虑$\mu = 0$ 的情况，可以从上面得到对于所有的$x \in \mathcal{D}$,

$$\sum_{i=1}^{m} \lambda_i c_i(x) + \nu^{\mathrm{T}}(Ax - b) \geq 0.$$

- 取满足Slater条件的点x_S, 有$\sum_{i=1}^{m} \lambda_i c_i(x_S) \geq 0$.
- 又$c_i(x_S) < 0$ 和$\lambda_i \geq 0$, 我们得到$\lambda = 0$, 上式化为

$$\nu^{\mathrm{T}}(Ax - b) \geq 0, \quad \forall\, x \in \mathcal{D}.$$

根据$(\lambda, \nu, \mu) \neq 0$ 可知$\nu \neq 0$, 结合A 行满秩可以得到$A^{\mathrm{T}}\nu \neq 0$.由于x_S 是可行解，我们有$\nu^{\mathrm{T}}(Ax_S - b) = 0$.

- 因为$x_S \in \mathbf{int}\,\mathcal{D}$, 则存在点$\tilde{x} = x_S + e \in \mathcal{D}$, 满足$\nu^{\mathrm{T}}(A\tilde{x} - b) < 0$.这与$\nu^{\mathrm{T}}(Ax - b) \geq 0$, $\forall\, x \in \mathcal{D}$ 矛盾.
- 综上所述，Slater条件能保证强对偶性.
- 在定理的证明中，Slater条件保证了$\mu \neq 0$.

一阶充要条件

- 对于一般的约束优化问题，当问题满足特定约束品性时，我们知道KKT条件是局部最优解处的必要条件.
- 而对于凸优化问题，当Slater条件满足时，KKT条件则变为局部最优解的充要条件(根据凸性，局部最优解也是全局最优解).

定理 (凸优化问题的一阶充要条件)

对于凸优化问题，用a_i 表示矩阵A^{T} 的第i 列，$\partial f, \partial c_i$ 表示次梯度，如果*Slater*条件成立，那么x^*, λ^* 分别是原始，对偶全局最优解当且仅当

$$
\begin{aligned}
\text{稳定性条件} \quad & 0 \in \partial f(x^*) + \sum_{i\in\mathcal{I}} \lambda_i^* \partial c_i(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* a_i, \\
\text{原始可行性条件} \quad & Ax^* = b, \ \forall i \in \mathcal{E}, \\
\text{原始可行性条件} \quad & c_i(x^*) \le 0, \ \forall i \in \mathcal{I}, \\
\text{对偶可行性条件} \quad & \lambda_i^* \ge 0, \ \forall i \in \mathcal{I}, \\
\text{互补松弛条件} \quad & \lambda_i^* c_i(x^*) = 0, \ \forall i \in \mathcal{I}.
\end{aligned}
$$

一阶充要条件:充分性

- 设存在$(\bar{x},\bar{\lambda})$ 满足KKT条件, 我们考虑凸优化问题的拉格朗日函数

$$L(x,\lambda)=f(x)+\sum_{i\in\mathcal{I}}\lambda_i c_i(x)+\sum_{i\in\mathcal{E}}\lambda_i(a_i^{\mathrm{T}}x-b_i).$$

- 当固定$\lambda=\bar{\lambda}$ 时, 注意到$\bar{\lambda}_i\geq 0, i\in\mathcal{I}$ 以及$\bar{\lambda}_i(a_i^{\mathrm{T}}x), i\in\mathcal{E}$ 是线性函数可知$L(x,\bar{\lambda})$ 是关于x 的凸函数.

- 由凸函数全局最优点的一阶充要性可知, 此时$\bar{x}$ 就是$L(x,\bar{\lambda})$ 的全局极小点.根据拉格朗日对偶函数的定义,

$$L(\bar{x},\bar{\lambda})=\inf_{x\in\mathcal{D}}L(x,\bar{\lambda})=g(\bar{\lambda}).$$

- 根据原始可行性条件$A\bar{x}=b$ 以及互补松弛条件$\bar{\lambda}_i c_i(\bar{x})=0, i\in\mathcal{I}$,

$$L(\bar{x},\bar{\lambda})=f(\bar{x})+0+0=f(\bar{x}).$$

- 根据弱对偶原理,

$$L(\bar{x},\bar{\lambda})=f(\bar{x})\geq p^*\geq d^*\geq g(\bar{\lambda})=L(\bar{x},\bar{\lambda})\Rightarrow p^*=d^*,$$

故$\bar{x},\bar{\lambda}$ 分别是原始问题和对偶问题的最优解.

关于充分性的评述

- 定理的充分性说明，若能直接求解出凸优化问题的KKT对，则其就是对应问题的最优解.
- 在充分性部分的证明中，我们没有使用Slater条件，这是因为在证明的一开始假设了KKT点是存在的.
- Slater条件的意义在于当问题最优解存在时，其相应KKT条件也会得到满足.
- 当Slater条件不满足时，即使原始问题存在全局极小值点，也可能不存在(x^*, λ^*) 满足KKT条件.

光滑凸优化实例:仿射空间的投影问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \frac{1}{2}\|x-y\|_2^2, \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

其中$A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ 以及$y \in \mathbb{R}^n$ 为给定的矩阵和向量且A 满秩.

- 拉格朗日函数: $L(x,\lambda) = \frac{1}{2}\|x-y\|^2 + \lambda^{\mathrm{T}}(Ax-b)$.
- Slater条件成立, x^* 为一个全局最优解当且仅当存在$\lambda^* \in \mathbb{R}^m$ 使得

$$\begin{cases} x^* - y + A^{\mathrm{T}}\lambda^* = 0, \\ Ax^* = b. \end{cases}$$

- 由上述KKT 条件第一式, 等号左右两边同时左乘A 可得

$$Ax^* - Ay + AA^{\mathrm{T}}\lambda = 0 \Rightarrow \lambda^* = (AA^{\mathrm{T}})^{-1}(Ay-b).$$

- 将λ^* 代回KKT 条件第一式可知

$$x^* = y - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(Ay-b).$$

因此点y 到集合$\{x \mid Ax = b\}$ 的投影为$y - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(Ay-b)$.