6. 凸问题和 KKT 条件

朱天宇

1 凸问题建模和性质

凸优化问题一般要解决

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ s.t. \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, ...m \\ & a_j^\top \mathbf{x} = 0, \quad j = 1, ...p \end{aligned} \tag{1}$$

这其中，

1. $\mathbf{x} \in \mathbb{R}^n$, 是优化变量
2. $f_0 : \mathbb{R}^n \to \mathbb{R}$，被称为目标函数，或者是 cost function
3. $f_i(\mathbf{x}) \leq 0$，不等式约束
4. $f_i : \mathbb{R}^n \to \mathbb{R}$，被称为不等式约束函数
5. $a_j^\top \mathbf{x} = 0$，线性等式约束

凸问题的条件 在1中，如果 $f_0, f_1, ..., f_m$ 都是凸函数，那么这就是一个凸优化问题。需要注意的是，等式约束**必须是**线性约束。

可行域、可行点 可行域就是所有约束的可行域的交集；如果可行域 $\mathcal{D}$ 非空，则称1为可行的 (feasible)。在实际问题之前，需要先判定可行域是否非空。

最优值 $v^\star$ 最优解 $\mathbf{x}^\star$ 在可行域中，使得 $f_0(\mathbf{x})$ 的下确界所对应的值。如果问题是不可行的，则 $v^\star = \infty$。最优值是一个下确界，不能保证最优解存在。$f_0(\mathbf{x}^\star) = v^\star$。局部最优的定义和之前的也是相似的。凸优化问题的任意局部最优解也是全局最优解。

1.1 一阶最优性条件

如果目标 f_0 是可微的，那么点 $\mathbf{x}$ 是最优解，当且仅当 对于任意定义域的点 $\mathbf{y}$, 满足

$$\nabla f_0(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \geq 0$$

。

对于无约束的情况，这个条件退化为

$$\nabla f_0(\mathbf{x}) = 0$$

对于**仅包括等式约束**的情况，问题可以写成：

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ s.t. \quad & \mathbf{A}\mathbf{x} = b \end{aligned}$$

此时最优性条件可以写成

$$\nabla f_0(\mathbf{x})^\top \mathbf{u} \geq 0, \forall \mathbf{u} \in \mathcal{N}(\mathbf{A})$$

这里 $\mathcal{N}(\mathbf{A})$ 表示矩阵的零空间，也就是使得 $\mathbf{A}\mathbf{x} = 0$ 的解空间。仅包含等式约束的凸优化问题的解与等式约束的零空间正交。

相对于一般的优化问题，一阶最优性条件仅仅是充分条件，而对于凸优化，一阶最优性条件是充要条件。[1]

[1] https://www.zhihu.com/question/24641575/answer/507237591

2 具体的凸优化问题

2.1 线性规划 (LP)

$$
\begin{aligned}
\min \quad & \mathsf{q}^\top \mathsf{x} + \mathsf{r} \\
s.t. \quad & \mathsf{G}\mathsf{x} \leq \mathsf{h} \\
& \mathsf{A}\mathsf{x} = \mathsf{b}
\end{aligned}
\tag{2}
$$

线性规划问题的目标函数和约束都是线性的。通常在目标函数的 r 可以忽略。其求解的基本思想给其他凸优化问题很多启发。

2.2 二次规划 (QP)

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathsf{x}^\top \mathsf{P}\mathsf{x} + \mathsf{q}^\top \mathsf{x} + \mathsf{r} \\
s.t. \quad & \mathsf{G}\mathsf{x} \leq \mathsf{h} \\
& \mathsf{A}\mathsf{x} = \mathsf{b}
\end{aligned}
\tag{3}
$$

2.3 二次约束的二次规划 (QCQP)

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathsf{x}^\top \mathsf{P}_0\mathsf{x} + \mathsf{q}_0^\top \mathsf{x} + \mathsf{r}_0 \\
s.t. \quad & \frac{1}{2}\mathsf{x}^\top \mathsf{P}_i\mathsf{x} + \mathsf{q}_i^\top \mathsf{x} + \mathsf{r}_i \leq 0, i = 1, ..., m \\
& \mathsf{A}\mathsf{x} = \mathsf{b}
\end{aligned}
\tag{4}
$$

2.4 二阶锥规划 (SOCP)

$$
\begin{aligned}
\min \quad & f^\top \mathsf{x} \\
s.t. \quad & \|\mathsf{L}_i\mathsf{x} + \mathsf{g}_i\|_2 \leq \mathsf{c}_i^\top \mathsf{x} + \mathsf{d}_i, i = 1, ..., m \\
& \mathsf{A}\mathsf{x} = \mathsf{b}
\end{aligned}
\tag{5}
$$

“二阶锥” 体现在其约束上。二阶锥规划的目标函数是**线性**的。

在 $\mathsf{c}_i = 0$ 的时候，SOCP 是 QCQP 的一个特殊情形。但实际上，SOCP 有更好的模型化能力，也就是说更加“广义”，任意的 QCQP 都可以被归纳为一种 SOCP 的特殊情况。

2.5 QCQP 转为 SOCP

引入辅助变量 y, 使得 $\frac{1}{2}\mathsf{x}^\top \mathsf{P}_0\mathsf{x} + \mathsf{q}_0^\top \mathsf{x} + \mathsf{r}_0 \leq \mathsf{y}$，那么 QCQP 问题4就可以写成

$$
\begin{aligned}
\min_{\mathsf{x},\mathsf{y}} \quad & \mathsf{y} \\
s.t. \quad & \frac{1}{2}\mathsf{x}^\top \mathsf{P}_i\mathsf{x} + \mathsf{q}_i^\top \mathsf{x} + \mathsf{r}_i \leq 0, i = 1, ..., m \\
& \frac{1}{2}\mathsf{x}^\top \mathsf{P}_0\mathsf{x} + \mathsf{q}_0^\top \mathsf{x} - \mathsf{y} + \mathsf{r}_0 \leq 0 \\
& \mathsf{A}\mathsf{x} = \mathsf{b}
\end{aligned}
\tag{6}
$$

问题6仍然是一个 QCQP 问题。但是6的目标函数已经是一个线性约束了 (SOCP 的目标函数是线性函数)，接下来需要将约束变成 SOCP 的形式就可以了。

对于一个二次约束 $\frac{1}{2}\mathsf{x}^\top\mathsf{P}\mathsf{x}+\mathsf{q}^\top\mathsf{x}+\mathsf{r}\le 0$，由于 $\mathsf{P}\in S_+^n$，即是半正定的，故而可以找到一个根 L，使得 $\mathsf{L}\mathsf{L}=\mathsf{P}$。令

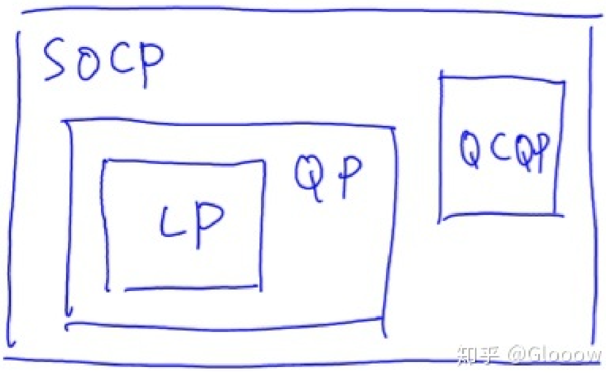
$$\widetilde{\mathsf{L}}=\begin{bmatrix}\mathsf{L}\\ \mathsf{q}^\top\end{bmatrix},\quad \widetilde{\mathsf{g}}=\begin{bmatrix}0\\ \vdots\\ 0\\ \mathsf{r}+\frac{1}{2}\end{bmatrix}\in\mathbb{R}^{n+1}$$

这里，$\widetilde{\mathsf{L}}$ 是一个 $(n+1)\times n$ 的矩阵，$\widetilde{\mathsf{g}}$ 是一个 $n+1$ 维。

从而有

$$\|\widetilde{\mathsf{L}}\mathsf{x}+\widetilde{\mathsf{g}}\|_2\le-(\mathsf{q}^\top\mathsf{x}+\mathsf{r}-\frac{1}{2})$$

从而转为了 SOCP 的形式。



3 凸问题的特殊性

3.1 拉格朗日函数 Lagrangian

回顾本章开头，凸优化问题的通式1

$$\begin{aligned}\min\quad & f_0(\mathsf{x})\\ s.t.\quad & f_i(\mathsf{x})\le 0,\quad i=1,...m\\ & h_j(\mathsf{x})=0,\quad j=1,...p\end{aligned}\tag{7}$$

记录符号 $\mathcal{D}=\cap_{i=0}^m\mathrm{dom}f_i\cap\cap_{j=1}^p\mathrm{dom}h_j$ 为问题的定义域。也就是约束的定义域的交集。

定义 拉格朗日函数 ：

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

函数 L 的定义域为 $\mathrm{dom}L=\mathcal{D}\times\mathbb{R}^m\times\mathbb{R}^p$。如果 x 被固定住，那么 L 关于 λ,ν 是线性的。λ,ν 被称为拉格朗日乘子，或者叫做**对偶变量**。

3.2 拉格朗日对偶函数

定义拉格朗日对偶函数 $g:\mathbb{R}^m\times\mathbb{R}^p\to\mathbb{R}$:

$$g(\lambda,\nu)=\inf_{\mathsf{x}\in\mathcal{D}}L(\mathsf{x},\lambda,\nu)=\inf_{\mathsf{x}\in\mathcal{D}}(f_0(\mathsf{x})+\sum_{i=1}^m\lambda_i f_i(\mathsf{x})+\sum_{j=1}^p\nu_j h_j(\mathsf{x}))\tag{8}$$

值得注意的是，对偶函数 g 是在 x 的整个定义域 $\mathcal{D}$ 中，对 L 取下确界获得的。$\mathcal{D}$ 是由约束函数的定义域的交集构成，并不要求一定要满足约束。也就是说，可能存在 $\mathsf{x}_0\in\mathcal{D}$ 使得某个 $f_i(\mathsf{x})>0$。在取完下确界 inf 之后，x 就被消掉了。

由于拉格朗日对偶函数是关于 λ,ν 线性的，故而其一定是凸的，无论原始问题7是不是凸的。

最优值的下界 记 $v^\star$ 是原始问题7的最优值 (不是最优解，是最优值)，那么对于任意的 $\lambda \geq 0, \nu$，必然有

$$g(\lambda, \nu) \leq v^\star$$

这样就为原问题的最优值确定了一个**下界**。尽管这是一个粗糙的下界，这个下界的取值可能是 $-\infty$。为了获得一个**最好的下界**，就需要解优化问题:

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ s.t. \quad & \lambda \geq 0 \end{aligned} \tag{9}$$

这个问题被称为 Lagrange dual problem，相应的，原问题7被称为原问题 (primal problem)。

整个逻辑思路为：原问题7$\Rightarrow$ 拉格朗日函数8$\Rightarrow$ 对偶问题9

原问题和对偶问题 对偶问题9总是一个凸问题，无论原问题是否是凸的；使用 $\lambda^\star, \nu^\star$ 表示对偶问题的最优解，对于对偶问题的最大值 $g^\star$，显然有

$$g^\star \leq v^\star$$

这条性质被称为 弱对偶性 。对偶问题的最大值小于原问题的最小值。**对偶问题和原问题之间存在一个 gap**。在如果 gap 等于 0，则对偶问题的最优值就等于原问题的最优值。

Slater's condition 满足 $g^\star = v^\star$，这被称为强对偶性。强对偶性是一个比较高的条件，对于一个凸的原问题，还需要对其附加额外条件，才能使其具备强对偶性。

Slater's condition[2]是一个比较容易验证的强对偶性条件。如果 $\exists \mathsf{x} \in \mathrm{relint}\mathcal{D}$ ，使得 $f_i(\mathsf{x}) < 0, i = 1, ..., m)$，那么就满足 Slater's 条件。这里 $\mathrm{relint}\mathcal{D} = \{\mathsf{x} \in \mathcal{D} | B(\mathsf{x}, r) \cap \mathrm{aff}\mathcal{D} \subseteq \mathcal{D},\ for\ some\ r > 0\}$，也就是说存在一个点 x，在其周围的球 $B(\mathsf{x}, r)$ 与 $\mathrm{aff}\mathcal{D}$ 相交的范围，是 $\mathcal{D}$ 的子集，且这个点满足所有不等式约束。

互补-充实条件 对于不等式约束 $\lambda_i^\star, f_i(\mathsf{x}^\star)$，满足互补充实条件，意味着

$$\lambda_i^\star f_i(\mathsf{x}^\star) = 0, i = 1, ..., m$$

也就是说，当 $\lambda > 0$ 的时候，要有 $f(x) = 0$；如果 $f(x) < 0$，就要求 $\lambda = 0$ 这里需要更加详细的说明。

4 KKT 条件

假设函数 $f_0, f_1, ..., f_m, h_1, ...h_p$ 都是可微分的。且 $\mathsf{x}^\star$ 是原问题的最优解，$(\lambda^\star, \nu^\star)$ 是对偶问题的最优解，且两个问题的 gap 为 0。

由于 x 可以最小化 $L(\mathsf{x}, \lambda^\star, \nu^\star)$，就有

$$\nabla f_0(\mathsf{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\mathsf{x}^\star) + \sum_{j=1}^{p} \nu_j^\star \nabla h_j(\mathsf{x}^\star) = 0$$

这条加上上述提及的各种约束，最终得到 KKT 系统 ：

$$\begin{cases} f_i(\mathsf{x}^\star) \leq 0, & i = 1...m \\ h_j(\mathsf{x}^\star) = 0, & j = 1...p \\ \lambda_i^\star \geq 0, & i = 1...m \\ \lambda_i^\star f_i(\mathsf{x}^\star) = 0, & i = 1...m \\ \nabla f_0(\mathsf{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\mathsf{x}^\star) + \sum_{j=1}^{p} \nu_j^\star \nabla h_j(\mathsf{x}^\star) = 0 & \end{cases} \tag{10}$$

这五个条件中，前两个是原本携带的等式不等式约束；三四是互补充实条件；五是 KT 条件。

后续的算法围绕着 KKT 系统出发。

[2]https://zhuanlan.zhihu.com/p/58064316