约束最优化

Atri

日期：May 26, 2023

1 等式约束

1.1 一般形式

等式约束问题的一般形式:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) = 0, i = 1, \cdots, m, \text{or}\, i \in \mathcal{E} \end{aligned}$$

1.2 可行域

可行域: 满足所有约束条件的点所组成的集合.

$$D = \{\boldsymbol{x} \in \mathbb{R}^n | h_i(\boldsymbol{x}) = 0, i \in \mathcal{E}\}.$$

1.3 极值点

局部极小值点: 设 $\boldsymbol{x}^* \in D$, 对于任何满足 $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon, \epsilon \to 0$ 的 $\boldsymbol{x} \in D$, 都有

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*),$$

则称 $\boldsymbol{x}^*$ 为局部极小值点.
全局极小值点: 顾名思义.
若不等号严格成立, 则为严格局部 or 全局极小值点.

1.4 一阶条件

令 $\boldsymbol{\lambda} = (\lambda_1; \cdots; \lambda_m)$, 构造拉格朗日函数

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{h}(\boldsymbol{x}).$$

给出一阶 KKT or 拉格朗日条件:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(\boldsymbol{x}^*) = 0.$$

$$h_i(\boldsymbol{x}^*) = 0, i \in \mathcal{E}.$$

这给出了极小值点所应满足的必要条件, 其前提是向量组 $(\nabla h_1(\boldsymbol{x}), \cdots, \nabla h_m(\boldsymbol{x}))$ 在点 $\boldsymbol{x}^*$ 线性独立, 对应于 $h(\boldsymbol{x})$ 的 Jacobi 矩阵 $\nabla \boldsymbol{h}(\boldsymbol{x})$ 列满秩.

1.5 二阶条件

Lagrange 函数对 $\boldsymbol{x}$ 的 Hessian 矩阵:

$$\nabla_{\boldsymbol{x}}^2 L(\boldsymbol{x},\boldsymbol{\lambda}) = \nabla^2 f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(\boldsymbol{x}).$$

定理 1.1. *设 $\boldsymbol{x}^*$ 是 f 在约束条件 $\boldsymbol{h}(\boldsymbol{x}) = 0$ 的局部极小值点, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m, m \le n, f, \boldsymbol{h}$ 是二阶连续可微函数. 如果 $\boldsymbol{x}^*$ 处 $\boldsymbol{h}$ 的 Jacobi 矩阵是非退化的, 则存在 $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ 使得*

- *一阶条件:*

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) = 0.$$

- *二阶必要条件: 对任意切向量 $\forall \boldsymbol{y} \in T(\boldsymbol{x}^*)$, 都有*

$$\boldsymbol{y}^\top \nabla_{\boldsymbol{x}}^2 L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\boldsymbol{y} \ge 0.$$

二阶充分条件: 若上述不等式严格成立, 则 $\boldsymbol{x}^*$ 是 f 在约束条件下的严格局部极小值点.

1.6 求解步骤

使用 KKT 条件计算最优性条件:

- 首先, 写出 Lagrange 函数, 分别计算其对 $\boldsymbol{x}$ 与 $\boldsymbol{\lambda}$ 的梯度以及对 $\boldsymbol{x}$ 的 Hessian 矩阵.
- 由一阶条件计算确定最优点 $(\boldsymbol{x},\boldsymbol{\lambda})$ 所满足的基本条件.

在此基础上:

- 首先确定最优点 $(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)$ 处, 约束 $\boldsymbol{h}(\boldsymbol{x}) = 0$ 的切空间:

$$T(\boldsymbol{x}^*) = \left\{ \boldsymbol{y} \in \mathbb{R}^n | \nabla_{\boldsymbol{x}} \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{y} = 0 \right\}.$$

- 判断, 对于 $\forall \boldsymbol{y} \in T(\boldsymbol{x}^*)$, 是否有 $\boldsymbol{y}^\top \nabla_{\boldsymbol{x}}^2 \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} > 0$.

2 不等式约束

2.1 一般形式

约束问题的一般形式:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = 0,\ i \in \mathcal{E} \\ & c_j(\boldsymbol{x}) \le 0,\ j \in \mathcal{I}. \end{aligned}$$

定义域为: $D = \operatorname{dom}(f) \cap_{i\in\mathcal{E}} \operatorname{dom}(c_i) \cap_{i\in\mathcal{I}} \operatorname{dom}(c_j)$.

- $c_i(\boldsymbol{x}), i \in \mathcal{E}$ 每个等式约束刻画了一个 $n-1$ 维曲面.
- $c_j(\boldsymbol{x}), j \in \mathcal{I}$ 不等式约束刻画了一个 n 维空间的区域.

2.2 有效约束

定义 2.1. 对任何 $\boldsymbol{x} \in \mathbb{R}^n$, 点 $\boldsymbol{x}$ 处的有效约束集为:

$$\mathcal{A}(\boldsymbol{x}) = \mathcal{E} \cup \mathcal{I}(\boldsymbol{x}), \quad \mathcal{I}(\boldsymbol{x}) := \{ j | h_j(\boldsymbol{x}) = 0, j \in \mathcal{I} \}.$$

对于定义域内某点 $\boldsymbol{x}$ 满足 $h_j(\boldsymbol{x}) \neq 0$ 的约束为该点的非有效约束. 即至少在这个点附近区域内没有限制的效果.

2.3 可行集与切锥

可行方向: 设 $\boldsymbol{x}$ 是约束优化问题的一个可行点, 对于 $\boldsymbol{d} \in \mathbb{R}^n, \boldsymbol{d} \neq 0$, 如果存在 $\alpha > 0$ 使得

$$\boldsymbol{x} + \beta \boldsymbol{d} \in D,\ \beta \in [0, \alpha]$$

则称 $\boldsymbol{d}$ 为 $\boldsymbol{x}$ 处的可行方向.
$\boldsymbol{x}$ 处所有可行方向的集合记为 $FD(\boldsymbol{x})$.
切锥: 在**最优化计算方法**这本书中, 上述可行方向也称为切向量 (这很反常), 更严谨的定义如下:

定义 2.2. 给定可行域 $\mathcal{X}$ 及其内一点 $\boldsymbol{x}$, 若存在可行序列 $\{\boldsymbol{z}_k\}_{k=1}^{\infty} \subset \mathcal{X}$ 逼近 $\boldsymbol{x}$ (即 $\lim_{k\to\infty} \boldsymbol{z}_k = \boldsymbol{x}$) 以及正标量序列 $\{t_k\}_{k=1}^{\infty}$, $t_k \to 0$ 满足

$$\lim_{k\to\infty} \frac{\boldsymbol{z}_k - \boldsymbol{x}}{t_k} = \boldsymbol{d},$$

则称向量 $\boldsymbol{d}$ 为 $\mathcal{X}$ 在点 $\boldsymbol{x}$ 处的一个切向量. 点 $\boldsymbol{x}$ 处的所有切向量的集合称为**切锥**, 用 $T_{\mathcal{X}}(\boldsymbol{x})$ 表示.

定义层面上, $T_{\mathcal{X}}(\boldsymbol{x}) = FD(\boldsymbol{x})$.
根据可行方向与极小值的定义, 可以得出不等式约束问题的几何最优性条件:

定理 2.1. *假设可行点 $\boldsymbol{x}^*$ 是不等式约束最优化问题的一个局部极小点. 如果 $f(\boldsymbol{x})$ 和 $c_i(\boldsymbol{x}), i \in \mathcal{I} \cup \mathcal{E}$ 在点 $\boldsymbol{x}^*$ 处是可微的, 那么*

$$\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0, \quad \boldsymbol{d} \in T_{\mathcal{X}}(\boldsymbol{x})$$

等价于

$$T_{\mathcal{X}}(\boldsymbol{x}) \cap \left\{\boldsymbol{d} | \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) < 0\right\} = \varnothing.$$

2.4 线性化可行方向锥

对于可行点 $\boldsymbol{x} \in \mathcal{X}$, 其线性化可行方向锥为

$$\mathcal{F}(\boldsymbol{x}) = \left\{\boldsymbol{d} | \boldsymbol{d}^\top \nabla c_i(\boldsymbol{x}) = 0, i \in \mathcal{E},\ \boldsymbol{d}^\top \nabla c_j(\boldsymbol{x}) \leq 0, j \in \mathcal{A}(\boldsymbol{x}) \cap \mathcal{I}\right\}.$$

直观理解:

- 线性化的含义是: 通过对约束函数值的一阶线性信息得到可行方向的另一个定义.
- 对于等式约束, 方向 $\boldsymbol{d}$ 不改变等式约束函数值.
- 对于不等式约束, 方向 $\boldsymbol{d}$ 可能使不等式约束函数值得到优化.
- 对非有效约束, 不管它.

很不幸的是, 线性化可行方向锥并不完全等价于切锥 (或者可行方向集), 事实上, 有如下结论:

$$T_{\mathcal{X}}(\boldsymbol{x}) \subseteq \mathcal{F}(\boldsymbol{x}).$$

2.5 约束品性

定义 2.3. 给定可行点 $\boldsymbol{x}$ 及相应的积极集 $\mathcal{A}(\boldsymbol{x})$. 如果积极集对应的约束函数的梯度 $\nabla c_i(\boldsymbol{x}), i \in \mathcal{A}(\boldsymbol{x})$ 是线性无关的, 则**线性无关约束品性** (LICQ) 在点 $\boldsymbol{x}$ 处成立.

然而, 课上给出了一个等价条件: $\forall i \in \mathcal{A}(\boldsymbol{x})$, 都有 $\nabla c_i(\boldsymbol{x}) \neq 0$. 显然我不知道为什么可以这样, 个人感觉不应该是等价条件.
顺带一提, 满足 LICQ 的点称为正则点, 在 EqCons 的讨论中也有涉及到.
对于所有约束函数都是线性函数的时候, 我们可以认为它们的梯度在任何点都是一致的, 如果一个点是正则点, 那么所有的点都是正则点, 因此默认这种情况下 $T_{\mathcal{X}}(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x})$. 当然, 最优化计算方法中没有讨论约束本身可以线性相关的情况, 这种情况应该很少, 默认不会出现.

2.6 一阶条件 (KKT)

定理 2.2. *假设 $\boldsymbol{x}^*$ 是不等式约束最优化问题的一个局部极小点, 如果*

$$T_{\mathcal{X}}(\boldsymbol{x}^*) = \mathcal{F}(\boldsymbol{x}^*)$$

成立, 那么存在拉格朗日乘子 $\boldsymbol{\lambda}^*$ 使得如下条件成立:

$$\begin{aligned}
&\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \nabla f(\boldsymbol{x}^*) + \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i^* \nabla c_i(\boldsymbol{x}^*) = 0, \\
&c_i(\boldsymbol{x}^*) = 0, \forall i \in \mathcal{E}, \\
&c_i(\boldsymbol{x}^*) \leq 0, \forall i \in \mathcal{I}, \\
&\lambda_i^* \geq 0, \forall i \in \mathcal{I}, \\
&\lambda_i^* c_i(\boldsymbol{x}^*) = 0, \forall i \in \mathcal{I}.
\end{aligned}$$

可以使用 Farkas 引理推出.
值得注意的是, 如果前置条件 $T_{\mathcal{X}}(\boldsymbol{x}^*) = \mathcal{F}(\boldsymbol{x}^*)$ 不满足, 那么 $\boldsymbol{x}^*$ 不一定是 KKT 点. 同样地, 因为 KKT 条件只是必要的, 我们还需引出二阶条件来判断局部最优性.

2.7 二阶条件

定义 2.4. 设 $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ 满足 KKT 条件, 定义临界锥为

$$\mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \left\{ \boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*) | \boldsymbol{d}^\top \nabla c_i(\boldsymbol{x}^*) = 0, \forall i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} \text{ and } \lambda_i^* > 0 \right\}$$

事实上, 由于临界锥是线性化可行方向锥的子集, 它对等式约束也满足上述条件, 因此有:

$$\boldsymbol{d} \in \mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \Rightarrow \lambda_i^* \nabla c_i(\boldsymbol{x}^*)^\top \boldsymbol{d} = 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}.$$

定理 2.3. *假设 $\boldsymbol{x}^*$ 是约束优化问题的一个局部极小值点, 且线性独立约束条件成立, $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ 满足 KKT 条件, 则*

$$\boldsymbol{d}^\top \nabla^2_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{d} \geq 0, \quad \forall \boldsymbol{d} \in \mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*), \ \boldsymbol{d} \neq 0.$$

定理 2.4. *假设在可行点 $\boldsymbol{x}^*$ 处, $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ 满足 KKT 条件, 如果*

$$\boldsymbol{d}^\top \nabla^2_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{d} > 0, \quad \forall \boldsymbol{d} \in \mathcal{C}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*), \ \boldsymbol{d} \neq 0.$$

那么 $\boldsymbol{x}^*$ 是约束优化问题的一个严格局部极小值.

与等式约束的二阶条件一样, 充分条件和必要条件之间的区别是不等式是否严格成立.

2.8 求解步骤

- 列出一阶 KKT 条件.
- 计算 KKT 点 (常见做法是分类: $\lambda_i = 0$ 或者 $c_i(\boldsymbol{x}^*) = 0$).
- 判断是否是正则点 (是否满足 LICQ, 可以首先判断是否满足约束条件的梯度为 0).
- 对每个 KKT 点, 求出 Lagrange 函数对 $\boldsymbol{x}$ 的 Hessian 矩阵以及 $\boldsymbol{x}^*$ 对应的切锥, 验证是否满足二阶充分条件.

可以发现这些步骤是包含等式约束题目的求解步骤的.

2.9 拓展

ICPC 竞赛中遇到的一道 C++ 矩阵求逆板子题 (python 不允许使用 numpy 库): https://ac.nowcoder.com/acm/contest/55992/D 其核心就是不等式约束的 KKT 条件.