最优性理论

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

致谢：本教案由谢中林协助准备

提纲

最优化问题解的存在性

考虑优化问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & x \in \mathcal{X}, \end{aligned}$$

其中 $\mathcal{X} \subseteq \mathbb{R}^n$ 为可行域.

- 首先要考虑的是最优解的存在性, 然后考虑如何求出其最优解.
- 在数学分析课程中, 我们学习过Weierstrass 定理, 即定义在紧集上的连续函数一定存在最大(最小) 值点.
- 而在许多实际问题中, 定义域可能不是紧的, 目标函数也不一定连续, 因此需要将此定理推广来保证最优化问题解的存在性.

定理 (推广的Weierstrass定理)

若函数$f:\mathcal{X}\to(-\infty,+\infty]$ 适当且闭, 且以下条件中任意一个成立:

1. $\mathrm{dom} f \stackrel{\text{def}}{=\!=} \{x\in\mathcal{X} : f(x)<+\infty\}$ 是有界的;
2. 存在一个常数$\bar{\gamma}$ 使得下水平集
$$C_{\bar{\gamma}} \stackrel{\text{def}}{=\!=} \{x\in\mathcal{X} : f(x)\le\bar{\gamma}\}$$
是非空且有界的;
3. f 是强制的, 即对于任一满足$\|x^k\|\to+\infty$ 的点列$\{x^k\}\subset\mathcal{X}$, 都有
$$\lim_{k\to\infty} f(x^k)=+\infty,$$

则函数f 的最小值点集$\{x\in\mathcal{X} \mid f(x)\le f(y),\ \forall y\in\mathcal{X}\}$ 非空且紧.

条件(2) 下的证明

假设条件(2) 成立, 且$t \xlongequal{\text{def}} \inf_{x\in\mathcal{X}} f(x) = -\infty$.

- 由下确界的定义, 存在点列$\{x^k\}_{k=1}^{\infty} \subset C_{\bar{\gamma}}$, 使得 $\lim\limits_{k\to\infty} f(x^k) = -\infty$.
- 由$C_{\bar{\gamma}}$ 有界知点列$\{x^k\}$ 存在聚点x^*.
- 由f 是闭函数知$\mathbf{epi} f$ 为闭集, 因此$(x^*, t) \in \mathbf{epi} f$. 根据上方图的定义知$f(x^*) \le t = -\infty$, 这与$f$ 适当矛盾, 故$t \xlongequal{\text{def}} \inf_{x\in\mathcal{X}} f(x)$ 有限.
- 由f 为闭函数知$C_{\bar{\gamma}}$ 为闭集, 由假设知$C_{\bar{\gamma}}$ 有界, 故为紧集.
- 由$f(x^*) = t$ 知最小值点集非空, 且为紧集$C_{\bar{\gamma}}$ 的子集, 而紧集的子集也是紧集, 故最小值点集为非空紧集.

条件(1)(3)下的证明

假设条件(1)成立, 则$\mathrm{dom} f$ 是有界的.

- 由f 适当知存在$x_0 \in \mathcal{X}$ 使得$f(x_0) < +\infty$.
- 记$\bar{\gamma} = f(x_0)$, 此时f 的$\bar{\gamma}$ 下水平集$C_{\bar{\gamma}}$ 非空有界, 条件(2)成立.

假设条件(3)成立, 我们证明条件(2)成立.

- 用反证法, 假设存在一个无界的下水平集$C_{\bar{\gamma}}$, 那么可以取点列$\{x^k\} \subset C_{\bar{\gamma}}$ 使得$\lim\limits_{k\to\infty} \|x^k\| = +\infty$, $\lim\limits_{k\to\infty} f(x^k) = +\infty$, 这与$f(x^k) \leq \bar{\gamma}$ 矛盾, 故此时f 的任意下水平集都有界, 条件(2)成立.

推广的Weierstrass定理:注记

- 推广的Weierstrass定理的三个条件在本质上都是保证$f(x)$的最小值不能在无穷远处取到.
- 因此我们可以仅在一个有界的下水平集中考虑$f(x)$的最小值.
- 定理仅要求$f(x)$为适当且闭的函数, 并不需要$f(x)$的连续性, 因此比数学分析中的Weierstrass定理应用范围更广.
- 当定义域不是有界闭集时, 对于强制函数$f(x) = x^2$, $x \in \mathcal{X} = \mathbb{R}$, 其全局最优解一定存在.
- 对于适当且闭的函数$f(x) = e^{-x}$, $x \in \mathcal{X} = \mathbb{R}$, 它不满足三个条件中任意一个, 因此我们不能断言其全局极小值点存在.事实上, 其全局极小值点不存在.

解的存在唯一性

- 推广的Weierstrass定理给出了最优解的存在性条件, 但其对应的解可能不止一个.
- 当最优解是唯一存在时, 我们可以通过比较不同算法收敛至该解的速度来判断算法好坏.
- 但是如果问题有多个最优值点, 不同的算法收敛到的最优值点可能不同, 那么这些算法收敛速度的比较就失去了参考价值.
- 因此, 最优化问题解的唯一性在理论分析和算法比较中扮演着重要角色.

解的存在唯一性: 拟强凸函数

定义 (拟强凸函数)

给定凸集$\mathcal{X}$ 和函数$f: \mathcal{X} \to (-\infty, +\infty]$.若任取$x \neq y$ 和$\lambda \in (0,1)$, 有

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\},$$

那么我们称函数f 是**强拟凸**的.

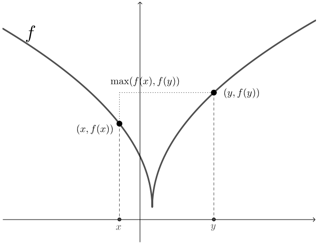

Figure: 一个强拟凸函数

拟强凸函数

- 强拟凸函数不一定是凸函数，但其任意一个下水平集都是凸集，并可以包含一部分性质较好的非凸函数.
- 任意强凸函数均为强拟凸的，但凸函数并不一定是强拟凸的.
- 任何定义在闭有界凸集上的强凸函数(如$f(x) = x^2$)，其最优解都是唯一存在的.
- 对于一般的凸函数，其最优解可能不唯一，比如$f(x) = \max\{x, 0\}$，任意$x \leq 0$ 都是$f(x)$ 的最优解.

唯一性定理

定理 (唯一性定理)

设$\mathcal{X}$ 是$\mathbb{R}^n$ 的一个非空，紧且凸的子集，如果$f:\mathcal{X}\to(-\infty,+\infty]$ 是适当，闭且强拟凸函数，那么存在唯一的x^* 满足

$$f(x^*) < f(x), \quad \forall x \in \mathcal{X}\backslash\{x^*\}.$$

证明: 由Weierstrass定理知f 至少存在一个全局极小点x^*.若x^*,y^* 皆为全局极小点，则有:

$$f(\lambda x^* + (1-\lambda)y^*) < \max\{f(x^*), f(y^*)\} = f(x^*), \quad \forall \lambda \in (0,1).$$

这与x^* 的全局极小性矛盾.

提纲

无约束可微问题的最优性理论:引言

无约束可微优化问题通常表示为如下形式：

$$\min_{x\in\mathbb{R}^n} \quad f(x),$$

其中f 是连续可微函数.

- 给定一个点$\bar{x}$, 我们想要知道这个点是否是函数f 的一个局部极小解或者全局极小解.
- 如果从定义出发, 需要对其邻域内的所有点进行判断, 这不可行.
- 因此, 需要一个更简单的方式来验证一个点是否为极小值点.我们称其为最优性条件, 它主要包含一阶最优性条件和二阶最优性条件.

一阶必要条件:下降方向

定义 (下降方向)

对于可微函数f 和点$x \in \mathbb{R}^n$, 如果存在向量d 满足

$$\nabla f(x)^{\mathrm{T}} d < 0,$$

那么称d 为f 在点x 处的一个下降方向.

- 一阶最优性条件是利用梯度(一阶)信息来判断给定点的最优性.
- 由下降方向的定义, 容易验证: 如果f 在点x 处存在一个下降方向d, 那么对于任意的$T > 0$, 存在$t \in (0, T]$, 使得

$$f(x + td) < f(x).$$

因此, 在局部最优点处不能有下降方向.

一阶必要条件

定理 (一阶必要条件)

假设f 在全空间$\mathbb{R}^n$ 可微.如果x^* 是一个局部极小点, 那么

$$\nabla f(x^*) = 0.$$

证明:任取$v \in \mathbb{R}^n$, 考虑f 在点$x = x^*$ 处的泰勒展开

$$f(x^* + tv) = f(x^*) + tv^{\mathrm{T}}\nabla f(x^*) + o(t),$$

整理得

$$\frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}}\nabla f(x^*) + o(1).$$

根据x^* 的最优性, 在上式中分别对t 取点0 处的左, 右极限可知

$$\lim_{t\to 0^+} \frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}}\nabla f(x^*) \geq 0,$$
$$\lim_{t\to 0^-} \frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}}\nabla f(x^*) \leq 0,$$

即对任意的v 有$v^{\mathrm{T}}\nabla f(x^*) = 0$, 由$v$ 的任意性知$\nabla f(x^*) = 0$.

二阶最优性条件

- 在没有额外假设时，如果一阶必要条件满足，我们仍然不能确定当前点是否是一个局部极小点.
- 假设f 在点x 的一个开邻域内是二阶连续可微的.类似于一阶必要条件的推导，可以借助当前点处的二阶泰勒展开来逼近该函数在该点附近的取值情况，从而来判断最优性.
- 在点x 附近我们考虑泰勒展开

$$f(x+d)=f(x)+\nabla f(x)^{\mathrm{T}}d+\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x)d+o(\|d\|^2).$$

- 当一阶必要条件满足时，$\nabla f(x)=0$，那么上面的展开式简化为

$$f(x+d)=f(x)+\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x)d+o(\|d\|^2).$$

由此，我们可以导出二阶最优性条件.

二阶最优性条件

定理 (二阶最优性条件)

必要条件:若x^* 是f 的一个局部极小点, 则$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0$.

充分条件:若$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$, 则$x^*$ 是f 的一个局部极小点.

- **必要性**:若$\nabla^2 f(x^*)$ 有负的特征值$\lambda_- < 0$, 设$\nabla^2 f(x^*)d = \lambda_- d$, 则

$$\frac{f(x^* + d) - f(x^*)}{\|d\|^2} = \frac{1}{2}\frac{d^{\mathrm{T}}}{\|d\|}\nabla^2 f(x^*)\frac{d}{\|d\|} + o(1) = \frac{1}{2}\lambda_- + o(1).$$

  当$\|d\|$ 充分小时, $f(x^* + d) < f(x^*)$, 这和点x^* 的最优性矛盾.

- **充分性**:由$\nabla f(x^*) = 0$ 时的二阶展开,

$$\frac{f(x^* + d) - f(x^*)}{\|d\|^2} = \frac{d^{\mathrm{T}}\nabla^2 f(x^*)d + o(\|d\|^2)}{\|d\|^2} \geq \frac{1}{2}\lambda_{\min} + o(1).$$

  当$\|d\|$ 充分小时有$f(x^* + d) \geq f(x^*)$, 即二阶充分条件成立.

二阶最优性条件:注记

- 设点$\bar{x}$ 满足一阶最优性条件(即$\nabla f(\bar{x}) = 0$), 且该点处的海瑟矩阵$\nabla^2 f(\bar{x})$ 不是半正定的, 那么$\bar{x}$ 不是一个局部极小点.
- 进一步地, 如果海瑟矩阵$\nabla^2 f(\bar{x})$ 既有正特征值又有负特征值, 我们称稳定点$\bar{x}$ 为一个鞍点.
- 事实上, 记d_1, d_2 为其正负特征值对应的特征向量, 那么对于任意充分小的$t > 0$, 我们都有$f(\bar{x} + td_1) > f(\bar{x})$ 且$f(\bar{x} + td_2) < f(\bar{x})$.
- 注意, 二阶最优性条件给出的仍然是关于局部最优性的判断.对于给定点的全局最优性判断, 我们还需要借助实际问题的性质, 比如目标函数是凸的、非线性最小二乘问题中目标函数值为0 等.

- 线性最小二乘:

$$\min_{x\in\mathbb{R}^n} \quad f(x) \xlongequal{\text{def}} \frac{1}{2}\|b-Ax\|_2^2,$$

其中$A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$.由f 可微且凸知

$$x^* \text{ 为一个全局最优解} \Leftrightarrow \nabla f(x^*) = A^{\mathrm{T}}(Ax^*-b) = 0.$$

- 非线性最小二乘(实数情形的相位恢复):

$$\min_{x\in\mathbb{R}^n} \quad f(x) \xlongequal{\text{def}} \sum_{i=1}^m r_i^2(x),$$

其中$r_i(x) = (a_i^{\mathrm{T}}x)^2 - b_i^2, i = 1, 2, \cdots, m$.

实数情形的相位恢复

$$\nabla f(x) = 2\sum_{i=1}^{m} r_i(x)\nabla r_i(x) = 4\sum_{i=1}^{m}((a_i^{\mathrm{T}}x)^2 - b_i^2)(a_i^{\mathrm{T}}x)a_i,$$

$$\nabla^2 f(x) = \sum_{i=1}^{m}(12(a_i^{\mathrm{T}}x)^2 - 4b_i^2)a_i a_i^{\mathrm{T}}.$$

- 如果x^* 为局部最优解, 那么其满足一、二阶必要条件

$$\sum_{i=1}^{m}((a_i^{\mathrm{T}}x^*)^2 - b_i)(a_i^{\mathrm{T}}x^*)a_i = 0, \quad \sum_{i=1}^{m}(12(a_i^{\mathrm{T}}x^*)^2 - 4b_i^2)a_i a_i^{\mathrm{T}} \succeq 0.$$

- 如果一个点$x^\#$ 满足二阶充分条件

$$\sum_{i=1}^{m}((a_i^{\mathrm{T}}x^\#)^2 - b_i)(a_i^{\mathrm{T}}x^\#)a_i = 0, \quad \sum_{i=1}^{m}(12(a_i^{\mathrm{T}}x^\#)^2 - 4b_i^2)a_i a_i^{\mathrm{T}} \succ 0,$$

那么$x^\#$ 为局部最优解.

提纲

切锥

定义 (切锥)

给定可行域$\mathcal{X}$ 及$x \in \mathcal{X}$, 若存在序列$\{z_k\}_{k=1}^{\infty} \subset \mathcal{X}$, $\lim_{k\to\infty} z_k = x$ 以及正标量序列$\{t_k\}_{k=1}^{\infty}$, $t_k \to 0$ 满足$\lim_{k\to\infty} \frac{z_k - x}{t_k} = d$ 则称向量d 为$\mathcal{X}$ 在点x 的一个切向量. 所有点x 处的切向量构成的集合称为切锥, 用$T_{\mathcal{X}}(x)$ 表示.

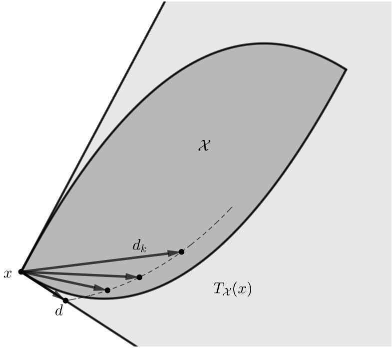

Figure: 不等式约束

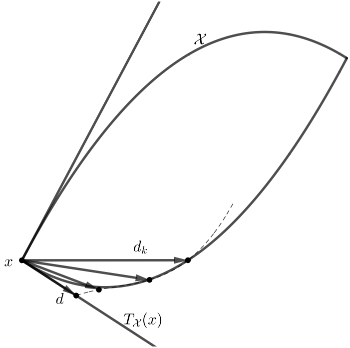

Figure: 等式约束

几何最优性条件

一般优化问题:

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

定理 (几何最优性条件)

假设可行点x^* 是上述问题的一个局部极小点.如果$f(x)$和$c_i(x)$, $i\in\mathcal{I}\cup\mathcal{E}$ 在点x^* 处是可微的, 那么

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

并且它等价于

$$T_{\mathcal{X}}(x^*)\cap\{d \mid \nabla f(x^*)^{\mathrm{T}}d < 0\} = \emptyset.$$

几何最优性条件证明

证明：$d^{\mathrm{T}}\nabla f(x^*) \geq 0, \quad \forall d \in T_{\mathcal{X}}(x^*)$

- 假设存在$d \in T_{\mathcal{X}}(x^*)$ 使得$d^{\mathrm{T}}\nabla f(x^*) < 0$。令$\{z_k\}, \{t_k\}$是$d$满足切锥定义的点列，因此有

$$\begin{aligned} f(z_k) &= f(x^*) + (z_k - x^*)^{\mathrm{T}}\nabla f(x^*) + o(\|z_k - x^*\|) \\ &= f(x^*) + t_k d^{\mathrm{T}}\nabla f(x^*) + o(t_k). \end{aligned}$$

其中第二行由切锥定义得到$z_k = x^* + t_k d + o(t_k)$。

- 由于$d^{\mathrm{T}}\nabla f(x^*) < 0$，余项最终由一阶项主导，即当$k$ 充分大时：

$$f(z_k) < f(x^*) + t_k d^{\mathrm{T}}\nabla f(x^*).$$

- 因此对于x^*任意开临域，可以选取充分大的k使得z_k属于该临域，且函数值比$f(x^*)$小。与x^*是局部极小点矛盾。

线性化可行锥

定义 (线性化可行锥)

对于可行点$x \in \mathcal{X}$, 定义该点的积极集$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \ : \ c_i(x) = 0\}$,
点x 处的**线性化可行方向锥**定义为

$$\mathcal{F}(x) = \left\{ d \ \middle| \ \begin{array}{l} d^{\mathrm{T}} \nabla c_i(x) = 0, \ \forall\, i \in \mathcal{E}, \\ d^{\mathrm{T}} \nabla c_i(x) \leq 0, \ \forall\, i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}$$

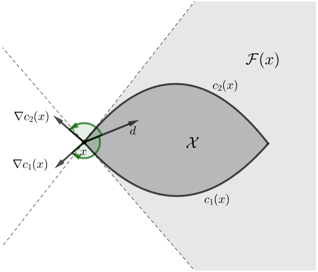

Figure: $\mathbb{R}^2$ 上的约束集合和线性化可行方向锥

线性化可行锥包含切锥

定理 (线性化可行锥包含切锥)

设$c_i(x), i \in \mathcal{E} \cup \mathcal{I}$ 一阶连续可微, 则对任意可行点x 有$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$.

证明:不妨设积极集$\mathcal{A}(x) = \mathcal{E} \cup \mathcal{I}$, 设$d \in T_{\mathcal{X}}(x)$, 由定义

$$\lim_{k\to\infty} t_k = 0, \quad \lim_{k\to\infty} \frac{z_k - x}{t_k} = d \Leftrightarrow z_k = x + t_k d + e_k$$

其中残量e_k 满足$\|e_k\| = o(t_k)$.对$i \in \mathcal{E}$, 根据泰勒展开,

$$\begin{aligned} 0 &= \frac{1}{t_k} c_i(z_k) \\ &= \frac{1}{t_k}\left(c_i(x) + \nabla c_i(x)^{\mathrm{T}}(t_k d + e_k) + o(t_k)\right) \\ &= \nabla c_i(x)^{\mathrm{T}} d + \frac{\nabla c_i(x)^{\mathrm{T}} e_k}{t_k} + o(1). \end{aligned}$$

注意到$\frac{\|e_k\|}{t_k} \to 0$, 令$k \to \infty$ 即可得到

$$\nabla c_i(x)^{\mathrm{T}} d = 0, \quad i \in \mathcal{E}.$$

线性化可行锥包含切锥

同理, 对$i \in \mathcal{I}$, 根据泰勒展开,

$$\begin{aligned}0 &\geq \frac{1}{t_k} c_i(z_k) \\ &= \frac{1}{t_k}\left(c_i(x) + \nabla c_i(x)^{\mathrm{T}}(t_k d + e_k) + o(t_k)\right) \\ &= \nabla c_i(x)^{\mathrm{T}} d + \frac{\nabla c_i(x)^{\mathrm{T}} e_k}{t_k} + o(1).\end{aligned}$$

注意到$c_i(x) = 0,\ i \in \mathcal{I}$, 因此我们有

$$\nabla c_i(x)^{\mathrm{T}} d \leq 0, \quad i \in \mathcal{I}$$

结合以上两点, 最终可得到$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$

反例:切锥未必包含线性化可行锥

$$\begin{aligned} \min_{x\in\mathbb{R}} \quad & f(x) = x \\ \text{s.t.} \quad & c(x) = -x + 3 \leq 0 \end{aligned}$$

- 则$T_{\mathcal{X}}(3) = \{d \mid d \geq 0\}, \mathcal{F}(3) = \{d : d \geq 0\}$, 于是$T_{\mathcal{X}}(3) = \mathcal{F}(3)$
- 将问题的约束变形为

$$c(x) = (-x+3)^3 \leq 0$$

因为可行域不变, 故点$x^* = 3$ 处, 切锥$T_{\mathcal{X}}(x^*) = \{d : d \geq 0\}$ 不变

- 由$c'(x^*) = -3(x^* - 3)^2 = 0$ 知线性化可行锥$\mathcal{F}(x^*) = \{d \mid d \in \mathbb{R}\}$
- 此时, $\mathcal{F}(x^*) \supset T_{\mathcal{X}}(x^*)$ (严格包含)

约束品性的引入

- 线性化可行方向锥$\mathcal{F}(x)$ 受可行域$\mathcal{X}$ 代数表示方式的影响
- 切锥$T_{\mathcal{X}}(x)$ 仅由可行域$\mathcal{X}$ 决定
- 线性可行化方向锥容易计算, 但不能反映可行域$\mathcal{X}$ 的本质特征
- 切锥能反映可行域$\mathcal{X}$ 的本质特征, 但不容易计算
- 引入约束品性来沟通两者, 确保最优点x^* 处$T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$, 从而可以用$\mathcal{F}(x)$ 取代$T_{\mathcal{X}}(x)$

线性无关约束品性

定义 (线性无关约束品性)

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

定理

给定任意可行点$x \in \mathcal{X}$, 若在该点LICQ 成立, 则有$T_{\mathcal{X}}(x) = \mathcal{F}(x)$.

- **证明**:不失一般性, 我们假设积极集$\mathcal{A}(x) = \mathcal{E} \cup \mathcal{I}$.记矩阵

$$A(x) = [\nabla c_i(x)]_{i \in \mathcal{I} \cup \mathcal{E}}^{\mathrm{T}}.$$

- 假设集合$\mathcal{A}(x)$ 的元素个数为m, 那么矩阵$A(x) \in \mathbb{R}^{m \times n}$ 并且$\mathrm{rank}(A) = m$.
- 令矩阵$Z \in \mathbb{R}^{n \times (n-m)}$ 为$A(x)$ 的零空间的基矩阵, 则Z 满足

$$\mathrm{rank}(Z) = n - m, \quad A(x)Z = 0.$$

- 令$d \in \mathcal{F}(x)$ 为任意线性化可行方向，给定$\lim_{k\to\infty}\ t_k = 0$ 的正标量$\{t_k\}_{k=1}^{\infty}$，定义映射$R:\mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$:

$$R(z,t) = \begin{bmatrix} c(z) - tA(x)d \\ Z^{\mathrm{T}}(z - x - td) \end{bmatrix},$$

其中$c(z)$ 为向量值函数，其第i 个分量为$c_i(z)$.

- 由$\mathcal{A}(x) = \mathcal{E} \cup \mathcal{I}$ 知

$$R(x,0) = \begin{bmatrix} c(x) \\ Z^{\mathrm{T}}(x - x) \end{bmatrix} = 0, \quad \frac{\partial R(x,0)}{\partial z} = \begin{bmatrix} A(x) \\ Z^{\mathrm{T}} \end{bmatrix}.$$

- 根据Z 的构造，雅可比矩阵$\frac{\partial R(x,0)}{\partial z}$ 是非奇异的.因此，由隐函数定理，对任意充分小的t_k，都存在唯一的z_k，使得$R(z_k,t_k) = 0$.
- 由于$R(z_k,t_k) = 0$，故$c_i(z_k) = t_k \nabla c_i(x)^{\mathrm{T}} d,\ i \in \mathcal{I} \cup \mathcal{E}$.根据线性化可行方向$d$ 的定义，$c_i(z_k) \geq 0,\ i \in \mathcal{I},\ c_i(z_k) = 0,\ i \in \mathcal{E}$，即$z_k$ 为可行点.

线性无关约束品性

- 由泰勒展开得

$$
\begin{aligned}
0 = R(z_k, t_k) &= \begin{bmatrix} c(z_k) - t_k A(x) d \\ Z^{\mathrm{T}}(z_k - x - t_k d) \end{bmatrix} \\
&= \begin{bmatrix} A(x)(z_k - x) + e_k - t_k A(x) d \\ Z^{\mathrm{T}}(z_k - x - t_k d) \end{bmatrix} \\
&= \begin{bmatrix} A(x) \\ Z^{\mathrm{T}} \end{bmatrix} (z_k - x - t_k d) + \begin{bmatrix} e_k \\ 0 \end{bmatrix}.
\end{aligned}
$$

其中残量e_k 满足$\|e_k\| = o(t_k)$.

- 两边同时作用$\begin{bmatrix} A(x)^{\mathrm{T}} & Z \end{bmatrix}^{-\mathrm{T}}$ 并除以t_k, 则有

$$
\frac{z_k - x}{t_k} = d + \begin{bmatrix} A(x) \\ Z^{\mathrm{T}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{e_k}{t_k} \\ 0 \end{bmatrix} \Rightarrow \lim_{k\to\infty} \frac{z_k - x}{t_k} = d,
$$

即$d \in T_{\mathcal{X}}(x)$.故$\mathcal{F}(x) \subseteq T_{\mathcal{X}}(x)$.又$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$, 则两集合相同.

Mangasarian–Fromovitz约束品性(MFCQ)

定义 (Mangasarian–Fromovitz约束品性)

给定可行点x 及相应的积极集$\mathcal{A}(x)$.如果存在一个向量$w \in \mathbb{R}^n$, 使得

$$\begin{aligned} \nabla c_i(x)^{\mathrm{T}} w < 0, &\quad \forall i \in \mathcal{A}(x) \cap \mathcal{I}, \\ \nabla c_i(x)^{\mathrm{T}} w = 0, &\quad \forall i \in \mathcal{E}, \end{aligned}$$

并且等式约束对应的梯度集$\{\nabla c_i(x),\ i \in \mathcal{E}\}$ 是线性无关的, 则称点x处*MFCQ*成立.

- Mangasarian–Fromovitz约束品性是LICQ的的一个常用推广, 简称为MFCQ.
- LICQ可以推出MFCQ, 但是反过来不成立.在MFCQ 成立的情况下, 我们也可以证明$T_{\mathcal{X}}(x) = \mathcal{F}(x)$.

线性约束品性

- 另外一个用来保证$T_{\mathcal{X}}(x)=\mathcal{F}(x)$的约束品性是线性约束品性.

定义 (线性约束品性)

若所有的约束函数$c_i(x)$, $i\in\mathcal{I}\cup\mathcal{E}$都是线性的, 则称线性约束品性成立.

- 当线性约束品性成立时, 也有$T_{\mathcal{X}}(x)=\mathcal{F}(x)$.
- 因此对只含线性约束的优化问题, 例如线性规划、二次规划, 很自然地有$T_{\mathcal{X}}(x)=\mathcal{F}(x),\forall\, x$.我们无需再关注约束函数的梯度是否线性无关.
- 一般来说, 线性约束品性和LICQ之间没有互相包含的关系.

Karush-Kuhn-Tucker(KKT)条件:引入

- 回顾几何最优性条件:

$$x^* \text{局部极小} \Leftrightarrow T_{\mathcal{X}}(x^*) \cap \{d \mid \nabla f(x^*)^{\mathrm{T}} d < 0\} = \varnothing.$$

- $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ 时(约束品性成立), 上述条件变为

$$\left\{ d \;\middle|\; \begin{array}{l} d^{\mathrm{T}} \nabla f(x^*) < 0, \\ d^{\mathrm{T}} \nabla c_i(x^*) = 0, \ i \in \mathcal{E}, \\ d^{\mathrm{T}} \nabla c_i(x^*) \leq 0, \ i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{array} \right\} = \varnothing.$$

- 上式依然难以验证, 但可使用Farkas引理进行化简.

Farkas引理

定理 (Farkas引理)

设p和q为两个非负整数，给定$\mathbb{R}^n$中的向量$\{a_i\}_{i=1}^p, \{b_i\}_{i=1}^q$和$c$. 则满足:

$$
\begin{aligned}
d^{\mathrm{T}} a_i &= 0,\ i = 1, 2, \cdots, p, \\
d^{\mathrm{T}} b_i &\geq 0,\ i = 1, 2, \cdots, q, \\
d^{\mathrm{T}} c &< 0
\end{aligned}
$$

的d不存在当且仅当存在$\{\lambda_i\}_{i=1}^p$和$\mu_i \geq 0,\ i = 1, 2, \cdots, q$, 使得

$$c = \sum_{i=1}^p \lambda_i a_i + \sum_{i=1}^q \mu_i b_i. \quad (*)$$

证明: "$\Longleftarrow$" 假设存在λ_i和μ_i使得(*)成立。令$d^{\mathrm{T}} a_i = 0$，$d^{\mathrm{T}} b_i \geq 0$，则

$$d^{\mathrm{T}} c = \sum_{i=1}^p \lambda_i d^{\mathrm{T}} a_i + \sum_{i=1}^q \mu_i d^{\mathrm{T}} b_i \geq 0.$$

Farkas引理

证明: “$\Longrightarrow$” 构造以下线性规划及其对偶问题：

$$(P)\begin{cases}\min\limits_{d} & d^{\mathrm{T}}c\\ \text{s.t.} & d^{\mathrm{T}}a_i=0,\ i=1,2,\cdots,p,\\ & d^{\mathrm{T}}b_i\ge 0,\ i=1,2,\cdots,q.\end{cases}$$

$$(D)\begin{cases}\max\limits_{\lambda_i,\mu_i} & 0\\ \text{s.t.} & c=\sum\limits_{i=1}^{p}\lambda_i a_i+\sum\limits_{i=1}^{q}\mu_i b_i,\quad \mu_i\ge 0.\end{cases}$$

假设(*)不成立。则(D)不可行。由LP对偶理论，(P)无界或者不可行。由于$d=0$是(P)的一个可行解，因此(P)无界，也就是存在d使得$d^{\mathrm{T}}c<0$。矛盾。

从Farkas引理到KKT条件

- 由Farkas引理，取$a_i = \nabla c_i(x^*), i \in \mathcal{E}, b_i = \nabla c_i(x^*), i \in \mathcal{A}(x^*) \cap \mathcal{I}$ 以及$c = -\nabla f(x^*)$，则$T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ 时几何最优性条件等价于：

$$-\nabla f(x^*) = \sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i(x^*) + \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{I}} \lambda_i^* \nabla c_i(x^*),$$

其中$\lambda_i^* \in \mathbb{R}, i \in \mathcal{E}$, $\lambda_i^* \geq 0, i \in \mathcal{A}(x^*) \cap \mathcal{I}$.

- 如果补充定义$\lambda_i^* = 0$, $i \in \mathcal{I} \backslash \mathcal{A}(x^*)$，那么

$$-\nabla f(x^*) = \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i^* \nabla c_i(x^*),$$

这恰好对应于拉格朗日函数关于x 的一阶最优性条件.

- **互补松弛条件**:对于任意的$i \in \mathcal{I}$，我们注意到

$$\lambda_i^* c_i(x^*) = 0.$$

这说明$i \in \mathcal{A}(x^*) \cap \mathcal{I}$ 时乘子$\lambda_i^* = 0$ 或$c_i(x^*) = 0$ 至少出现一种，当两种情况恰好只有一种满足时，我们也称严格互补松弛条件.

KKT条件:总结

假设x^* 是一般优化问题

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

的一个局部最优点.如果

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

成立, 那么存在拉格朗日乘子λ_i^* 使得如下条件成立:

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

KKT条件:注记

- 称满足KKT条件的变量对(x^*, λ^*) 为KKT对.
- 称x^* 为KKT点.
- 如果局部最优点x^* 处$T_{\mathcal{X}}(x^*) \neq \mathcal{F}(x^*)$, 那么$x^*$ 不一定是KKT点.
- KKT条件只是必要的, KKT点不一定是局部最优点.

二阶最优性条件: 引入

依旧考虑一般优化问题:

$$\min_{x\in\mathbb{R}^n} f(x), \quad \text{s.t.} \;\; c_i(x) \le 0,\; i\in\mathcal{I};\; c_i(x)=0,\; i\in\mathcal{E}.$$

若x^* 是满足KKT条件的点, 假设$T_{\mathcal{X}}(x^*)=\mathcal{F}(x^*)$, 则$\forall d\in\mathcal{F}(x^*)$,

$$d^{\mathrm{T}}\nabla f(x^*) = -\sum_{i\in\mathcal{E}}\underbrace{\lambda_i^* d^{\mathrm{T}}\nabla c_i(x^*)}_{=0} - \sum_{i\in\mathcal{A}(x^*)\cap\mathcal{I}}\underbrace{\lambda_i^* d^{\mathrm{T}}\nabla c_i(x^*)}_{\le 0} \ge 0,$$

此时一阶条件无法判断x^* 是否是最优值点.

- 若$d^{\mathrm{T}}\nabla f(x^*)=0$, 则需要利用二阶信息来进一步判断在其可行邻域内的目标函数值.
- 拉格朗日函数在这些方向上的曲率即可用来判断x^* 的最优性.
- 首先引入临界锥来精确刻画这些方向.

临界锥

定义 (临界锥)

设(x^*, λ^*) 是满足*KKT*条件的*KKT*对，定义临界锥为

$$\mathcal{C}(x^*, \lambda^*) = \{d \in \mathcal{F}(x^*) \mid \nabla c_i(x^*)^{\mathrm{T}} d = 0,\ \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ 且 } \lambda_i^* > 0\},$$

其中$\mathcal{F}(x^*)$ 为点x^* 处的线性化可行方向锥.

- 临界锥是线性化可行方向锥$\mathcal{F}(x^*)$ 的子集.
- 沿着临界锥中的方向进行优化，所有等式约束和$\lambda_i^* > 0$ 对应的不等式约束(此时这些不等式均取等)都会尽量保持不变.
- 当$d \in \mathcal{C}(x^*, \lambda^*)$ 时，$\forall i \in \mathcal{E} \cup \mathcal{I}$ 有$\lambda_i^* \nabla c_i(x^*)^{\mathrm{T}} d = 0$，故
$$d^{\mathrm{T}} \nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* d^{\mathrm{T}} \nabla c_i(x^*) = 0.$$
- 临界锥定义了依据一阶导数不能判断是否为下降或上升方向的线性化可行方向，必须使用高阶导数信息加以判断.

二阶最优性条件

考虑一般优化问题:

$$\min_{x\in\mathbb{R}^n} f(x), \qquad \text{s.t.} \;\; c_i(x) \le 0,\; i\in\mathcal{I};\; c_i(x)=0,\; i\in\mathcal{E}.$$

定理 (二阶最优性条件)

必要性:假设x^* 是问题的一个局部最优解, 并且$T_{\mathcal{X}}(x^*)=\mathcal{F}(x^*)$ 成立.令λ^* 为相应的拉格朗日乘子, 即(x^*,λ^*) 满足*KKT*条件, 那么

$$d^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda^*)d \ge 0, \quad \forall d\in\mathcal{C}(x^*,\lambda^*).$$

充分性:假设在可行点x^* 处, 存在一个拉格朗日乘子λ^*, 使得(x^*,λ^*) 满足*KKT*条件.如果

$$d^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda^*)d > 0, \quad \forall d\in\mathcal{C}(x^*,\lambda^*),\; d\neq 0,$$

那么x^* 为问题的一个严格局部极小解.

回顾无约束优化问题的二阶最优性条件:

- 问题: $\min_{x\in\mathbb{R}^n} f(x)$.
- 必要条件:若x^* 是f 的一个局部极小点, 则$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0$.
- 充分条件:若$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$, 则$x^*$ 是f 的一个局部极小点.

约束优化问题的二阶最优性条件也要求某种"正定性", 但只需要考虑临界锥$\mathcal{C}(x^*, \lambda^*)$ 中的向量而无需考虑全空间的向量.

有些教材中将其称为"投影半正定性".

$$\min \quad x_1^2 + x_2^2, \qquad \text{s.t.} \quad \frac{x_1^2}{4} + x_2^2 - 1 = 0,$$

其拉格朗日函数为

$$L(x, \lambda) = x_1^2 + x_2^2 + \lambda(\frac{x_1^2}{4} + x_2^2 - 1).$$

该问题可行域在任意一点$x = (x_1, x_2)^{\mathrm{T}}$ 处的线性化可行方向锥为

$$\mathcal{F}(x) = \{(d_1, d_2) \mid \frac{x_1}{4} d_1 + x_2 d_2 = 0\}.$$

因为只有一个等式约束且其对应函数的梯度非零, 故有LICQ成立, 于是$\mathcal{F}(x) = T_{\mathcal{X}}(x)$.若$(x, \lambda)$ 为KKT对, 由于无不等式约束,
故$\mathcal{C}(x, \lambda) = \mathcal{F}(x)$.
可以计算出其4 个KKT对

$$(x^{\mathrm{T}}, \lambda) = (2, 0, -4), \quad (-2, 0, -4), \quad (0, 1, -1) \quad \text{和} \quad (0, -1, -1).$$

约束优化的最优性理论: 例子

考虑第一个KKT对$y = (2, 0, -4)^{\mathrm{T}}$, 计算可得

$$\nabla_{xx}^2 L(y) = \begin{bmatrix} 0 & 0 \\ 0 & -6 \end{bmatrix}, \quad \mathcal{C}(y) = \{(d_1, d_2) \mid d_1 = 0\}.$$

取$d = (0, 1)$, 则

$$d^{\mathrm{T}} \nabla_{xx}^2 L(y) d = -6 < 0,$$

因此y 不是局部最优点.类似地, 对第三个KKT对$z = (0, 1, -1)$,

$$\nabla_{xx}^2 L(z) = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathcal{C}(z) = \{(d_1, d_2) \mid d_2 = 0\}.$$

对于任意的$d = (d_1, 0)$ 且$d_1 \neq 0$,

$$d^{\mathrm{T}} \nabla_{xx}^2 L(z) d = \frac{3}{2} d_1^2 > 0.$$

因此, z 为一个严格局部最优点.

最大割问题的半定规划松弛

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X_{ii} = 1, \ i = 1, 2, \cdots, n \\ & X \succeq 0. \end{aligned}$$

- 拉格朗日函数:

$$L(X, \mu, \Lambda) = \langle C, X \rangle + \sum_{i=1}^{n} \mu_i (X_{ii} - 1) - \mathrm{Tr}(X\Lambda), \quad \Lambda \in \mathcal{S}_+^n, \ \mu \in \mathbb{R}^n.$$

- Slater条件成立, 最优性条件:

$$\begin{cases} C + \mathrm{Diag}(\mu^*) - \Lambda^* = 0, \\ X_{ii}^* = 1, \ i = 1, \cdots, n, \\ X^* \succeq 0, \\ \Lambda^* \succeq 0, \\ \mathrm{Tr}(X^* \Lambda^*) = 0. \end{cases}$$

由于X^* 与Λ^* 的半正定性, $\mathrm{Tr}(X^*\Lambda^*) = 0$ 等价于$X^*\Lambda^* = 0$.

$$\begin{aligned}\max_{Y\in\mathbb{R}^{n\times p}} \quad & \mathrm{Tr}(CYY^{\mathrm{T}}),\\ \text{s.t.} \quad & \mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}.\end{aligned}$$

- 利用$X = YY^{\mathrm{T}}$, $Y \in \mathbb{R}^{n\times p}$ 来减小规模, 此时$YY^{\mathrm{T}} \succeq 0$ 自然满足.
- 可行点$Y = [y_1, y_2, \cdots, y_n]^{\mathrm{T}}$ 处约束为$c_i(Y) = \|y_i\|^2 - 1 = 0, \forall i$.

$$\nabla c_i(Y) = 2[0, \cdots, 0, y_i, 0, \cdots 0]^{\mathrm{T}}.$$

  因为$y_i \neq 0$, 故$\{c_i(Y)\}_{i=1}^n$ 是线性无关的, 即LICQ 成立.
- 拉格朗日函数: $L(Y,\lambda) = \mathrm{Tr}(CYY^{\mathrm{T}}) - \sum_{i=1}^n \lambda_i c_i(Y)$.
- KKT条件:

$$\begin{cases} 2CY - 2[\lambda_1 y_1, \lambda_2 y_2, \cdots, \lambda_n y_n]^{\mathrm{T}} = 0, \\ \mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}. \end{cases}$$

- 令 $\Lambda = \mathrm{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, KKT条件第一式可以转换为

$$(C - \Lambda)Y = 0.$$

- 因为该问题只有等式约束, 故临界锥就是切锥, 即

$$\mathcal{C}(Y, \lambda) = \{D \in \mathbb{R}^{n \times p} \mid \mathrm{diag}(YD^{\mathrm{T}}) = 0\}.$$

- 拉格朗日函数的海瑟矩阵算子形式为

$$\nabla^2_{YY} L(X, \lambda)[D] = 2(C - \Lambda)D.$$

- **必要性**:假设Y 为一个局部最优解, 则存在λ_i, 使得

$$\begin{cases} \langle (C - \Lambda)D, D \rangle \geq 0, \quad \forall \mathrm{diag}(YD^{\mathrm{T}}) = 0, \\ \mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}, \\ (C - \Lambda)Y = 0. \end{cases}$$

光滑非凸优化: 最大割问题半定规划松弛的非凸分解

- **充分性**:假设在一点Y处, 存在λ_i, $i = 1, 2, \cdots, n$, 使得

$$\begin{cases} \langle (C - \Lambda)D, D\rangle > 0, & \forall \mathrm{diag}(YD^{\mathrm{T}}) = 0, \\ \mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}, \\ (C - \Lambda)Y = 0, \end{cases}$$

  那么Y为问题的一个严格局部最优解.

- 利用关系式$(C - \Lambda)Y = 0$和约束$\mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}$可以显式求得$\lambda = \mathrm{diag}(CYY^{\mathrm{T}})$.
- 在这个例子中根据约束的特殊结构, 我们能显式给出乘子λ的表达式.这个性质在一般约束优化问题中是没有的.

提纲

1. 最优化问题解的存在性
2. 无约束可微问题的最优性理论
3. 一般约束优化问题的最优性理论
4. 总结

总结: 无约束优化问题及其最优性条件

问题	一阶条件	二阶条件
可微问题	$\nabla f(x^*) = 0$ (必要)	$\nabla^2 f(x^*) \succeq 0$ (必要) $\nabla^2 f(x^*) \succ 0$ (充分)
凸问题	$0 \in \partial f(x^*)$ (充要)	—
复合优化问题	$-\nabla f(x^*) \in \partial h(x^*)$ (必要)	—
非凸非光滑	$0 \in \partial f(x^*)$ (必要)	—

总结: 约束优化问题的最优性条件和相应约束品性

问题	一阶条件	二阶条件	约束品性
一般问题	KKT 条件(必要)	$d^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda^*)d \geq 0, \quad \forall d \in \mathcal{C}(x^*,\lambda^*)$ (必要) $d^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda^*)d > 0, \quad \forall d \in \mathcal{C}(x^*,\lambda^*),\ d \neq 0$, (充分)[1]	LICQ[2]
凸问题	KKT 条件(充要)	—	Slater

1. 一般约束优化问题的二阶充分条件不需要LICQ 作为前提.
2. 或其他可推出$T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ 的约束品性.