微积分 A(2) 习题课

Week07

王兆臻

wzz23@mails.tsinghua.edu.cn

2024 年 4 月 11 日

基本概念

局部极小值

以下所说的局部极小值，不一定在定义域内部，可能在边界上.

全局极小值

以下所说的全局极小值，即定义域上的最小值.

凸函数

函数 $f(\boldsymbol{X})$, $\boldsymbol{X} \in$ 非空开凸集 $D \subseteq \mathbb{R}^n$.

凸函数 (下凸函数)

$\forall \boldsymbol{X}, \boldsymbol{Y} \in D,\ \forall \lambda \in [0,1],\ f(\lambda \boldsymbol{X} + (1-\lambda)\boldsymbol{Y}) \leq \lambda f(\boldsymbol{X}) + (1-\lambda) f(\boldsymbol{Y})$.

一阶充要条件 (f 可微)

$\forall \boldsymbol{X}, \boldsymbol{Y} \in D,\ f(\boldsymbol{Y}) \geq f(\boldsymbol{X}) + \nabla f(\boldsymbol{X})^T (\boldsymbol{Y} - \boldsymbol{X})$.

二阶充要条件 ($f \in C^2$)

$\forall \boldsymbol{X} \in D$, Hesse 矩阵 $\nabla^2 f(\boldsymbol{X})$ 半正定.

凸规划

凸规划

称下述规划为凸规划:

$$\begin{aligned} \min\ & f(\boldsymbol{X}) \\ \text{s.t.}\ & g_i(\boldsymbol{X}) \geq 0,\ i = 1, 2, \ldots, m \\ & h_j(\boldsymbol{X}) = 0,\ j = 1, 2, \ldots, l \end{aligned}$$

其中 $f(\boldsymbol{X})$ 是凸函数, $g_i(\boldsymbol{X})$ 是凹函数, $h_j(\boldsymbol{X})$ 是线性函数.

性质: 可行域是凸集, 局部极小值即为全局极小值且构成凸集.
特例: 线性规划.

凸集相关定理

Motzkin Transposition Theorem [1]

A, B, C 为矩阵, A 非空. 下述两个命题择一成立:

- $A\boldsymbol{x} > \mathbf{0},\ B\boldsymbol{x} \geq \mathbf{0},\ C\boldsymbol{x} = \mathbf{0}$ 有解 $\boldsymbol{x}$.
- $A^{\mathrm{T}}\boldsymbol{y_1} + B^{\mathrm{T}}\boldsymbol{y_2} + C^{\mathrm{T}}\boldsymbol{y_3} = \mathbf{0},\ \boldsymbol{y_1} \gneqq \mathbf{0},\ \boldsymbol{y_2} \geq \mathbf{0}$ 有解 $\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}$.

[1] http://benisrael.net/MOTZKIN.pdf

无约束极值问题 $\min f(\boldsymbol{X}),\ \boldsymbol{X} \in \mathbb{R}^n$

一阶必要条件

$f(\boldsymbol{X})$ 在 $\boldsymbol{X}^*$ 可微，若其为局部极小值点，则 $\nabla f(\boldsymbol{X}^*) = \mathbf{0}$.

二阶必要条件

$f(\boldsymbol{X})$ 在 $\boldsymbol{X}^*$ 二阶连续可微，若其为局部极小值点，则 $\nabla f(\boldsymbol{X}^*) = \mathbf{0}$，且 Hesse 矩阵 $\nabla^2 f(\boldsymbol{X}^*)$ 半正定.

二阶充分条件

$f(\boldsymbol{X})$ 在 $\boldsymbol{X}^*$ (邻域) 二阶连续可微，若 $\nabla f(\boldsymbol{X}^*) = \mathbf{0}$，且 $\nabla^2 f(\boldsymbol{X})$ 在 $\boldsymbol{X}^*$ 正定 (在 $\boldsymbol{X}^*$ 邻域半正定)，则 $\boldsymbol{X}^*$ 为局部极小值点.

一阶充要条件 (凸规划)

$f(\boldsymbol{X})$ 是可微凸函数，则 $\nabla f(\boldsymbol{X}^*) = \mathbf{0} \iff \boldsymbol{X}^*$ 为全局极小值点.

总结

$f(\boldsymbol{X})$ 在 $\boldsymbol{X}^*$ 二阶连续可微, 且 $\nabla f(\boldsymbol{X}^*) = \mathbf{0}$, 则

- 若 $\nabla^2 f(\boldsymbol{X}^*)$ 正定, 则 $\boldsymbol{X}^*$ 为 (严格) 局部极小值点.
- 若 $\nabla^2 f(\boldsymbol{X}^*)$ 负定, 则 $\boldsymbol{X}^*$ 为 (严格) 局部极大值点.
- 若 $\nabla^2 f(\boldsymbol{X}^*)$ 不定, 则 $\boldsymbol{X}^*$ 为鞍点.
- 若 $\nabla^2 f(\boldsymbol{X}^*)$ 半正定 / 半负定, 则无法判断.

$f(\boldsymbol{X})$ 在 $\boldsymbol{X}^*$ 邻域二阶连续可微, 且 $\nabla f(\boldsymbol{X}^*) = \mathbf{0}$, 则

- 若 $\nabla^2 f(\boldsymbol{X})$ 在 $\boldsymbol{X}^*$ 邻域半正定, 则 $\boldsymbol{X}^*$ 为局部极小值点.
- 若 $\nabla^2 f(\boldsymbol{X})$ 在 $\boldsymbol{X}^*$ 邻域半负定, 则 $\boldsymbol{X}^*$ 为局部极大值点.

基本概念

考虑以下规划 $(f, h_j \in C^1)$:

$$\begin{aligned} &\min\ f(\boldsymbol{X}) \\ &\text{s.t.}\ \ h_j(\boldsymbol{X}) = 0,\ j = 1, 2, \ldots, l \end{aligned}$$

可行域 $S = \{\boldsymbol{X} \in \mathbb{R}^n \mid h_j(\boldsymbol{X}) = 0, j = 1, 2, \ldots, l\}$.

切平面 (切空间)

过曲面 S 上一点 $\boldsymbol{X}$ 的所有可微曲线的切向量组成的集合称为 S 在 $\boldsymbol{X}$ 处的切平面 (切空间) $T(\boldsymbol{X})$.

一般地, $T(\boldsymbol{X}) \subseteq \{\boldsymbol{d} \mid \nabla \boldsymbol{h}(\boldsymbol{X})^{\mathrm{T}} \boldsymbol{d} = 0\}$.
证明: 设可微曲线方程 $\boldsymbol{X}(t)$, $\boldsymbol{h}(\boldsymbol{X}(t)) = \boldsymbol{0}$ 两边对 t 求导.
添加正则条件后, 上述包含关系可以改进为等号.

切空间与法空间

切空间的性质

设 $\boldsymbol{X}_0$ 是曲面 $\{\boldsymbol{X} \mid \boldsymbol{h}(\boldsymbol{X}) = \boldsymbol{0}\}$ 的正则点 (即 $\nabla h_j(\boldsymbol{X}_0)$ $(j = 1, 2, \cdots, l)$ 线性无关), 则 $T(\boldsymbol{X}_0) = \{\boldsymbol{d} \mid \nabla \boldsymbol{h}(\boldsymbol{X}_0)^{\mathrm{T}}\boldsymbol{d} = \boldsymbol{0}\}$.
即 $\nabla h_j(\boldsymbol{X}_0)$ $(j = 1, 2, \cdots, l)$ 是 $\boldsymbol{X}$ 处法空间的一组基.

证明: $(\supseteq)$ 隐函数定理. 考虑方程 $\boldsymbol{h}(\boldsymbol{X}_0 + t\boldsymbol{d} + \nabla \boldsymbol{h}(\boldsymbol{X}_0)\boldsymbol{y}) = \boldsymbol{0}$, 在 $(t, \boldsymbol{y}) = (0, \boldsymbol{0})$ 附近确定了隐函数 $\boldsymbol{y} = \boldsymbol{y}(t)$. 令可微曲线 $\boldsymbol{X}(t) = \boldsymbol{X}_0 + t\boldsymbol{d} + \nabla \boldsymbol{h}(\boldsymbol{X}_0)\boldsymbol{y}(t)$, 可验证其在曲面上, 过 $\boldsymbol{X}_0$, 且 $\boldsymbol{X}'(0) = \boldsymbol{d}$, 从而 $\boldsymbol{d} \in T(\boldsymbol{X}_0)$.

局部极小值点的梯度

施加约束后，局部极小值点的梯度 $\nabla f(\boldsymbol{X}^*)$ 不一定为 $\mathbf{0}$，它需要满足什么性质？

一阶必要条件

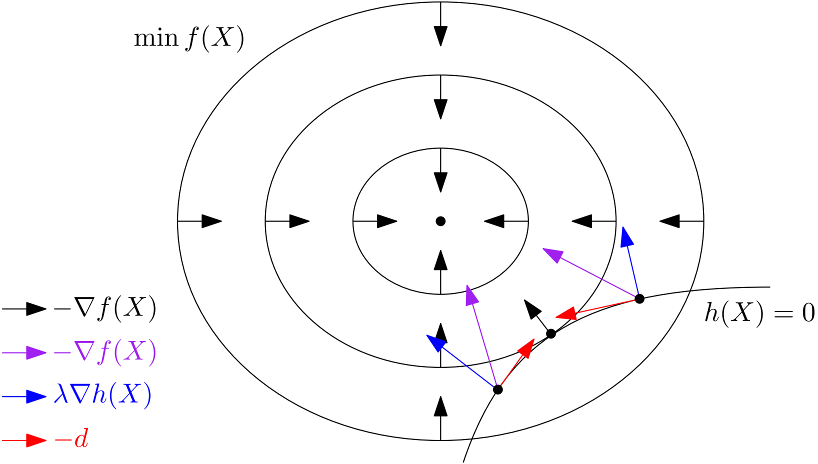
若 $\boldsymbol{X}^*$ 是局部极小值点、正则点，则 $\nabla f(\boldsymbol{X}^*)$ 无切空间分量，可用法空间的一组基 $\nabla h_j(\boldsymbol{X}^*)$ $(j=1,2,\cdots,l)$ 表示.

$$\nabla f(\boldsymbol{X}^*)=\sum_{j=1}^{l}\lambda_j\nabla h_j(\boldsymbol{X}^*)$$

否则，若存在切空间非零分量 $\boldsymbol{d}$，取曲面 $\{\boldsymbol{X}\mid \boldsymbol{h}(\boldsymbol{X})=\mathbf{0}\}$ 上过 $\boldsymbol{X}^*$ 的可微曲线 $\boldsymbol{X}(t)$ s.t. $\boldsymbol{X}(0)=\boldsymbol{X}^*$, $\boldsymbol{X}'(0)=\boldsymbol{d}$，则 $\left.\frac{\mathrm{d}f(\boldsymbol{X}(t))}{\mathrm{d}t}\right|_{t=0}=\nabla f(\boldsymbol{X}^*)^{\mathrm{T}}\boldsymbol{d}=\boldsymbol{d}^{\mathrm{T}}\boldsymbol{d}>0$, $f(\boldsymbol{X})$ 沿曲线 $\boldsymbol{X}(t)$ 不会在 $\boldsymbol{X}(0)=\boldsymbol{X}^*$ 处取极小值.

图解

几何意义：在局部极小值点，$\boldsymbol{h}(\boldsymbol{X}) = \mathbf{0}$ 与 $f(\boldsymbol{X})$ 等值线相切.



Lagrange 乘子法

设 Lagrange 函数

$$L(\boldsymbol{X}, \boldsymbol{\lambda}) = f(\boldsymbol{X}) - \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{X})$$

若 $\boldsymbol{X}^*$ 是局部极小值点，则是 Lagrange 函数的驻点:

$$\nabla_{\boldsymbol{X}} L(\boldsymbol{X}^*, \boldsymbol{\lambda}) = \nabla f(\boldsymbol{X}^*) - \sum_{j=1}^{l} \lambda_j \nabla h_j(\boldsymbol{X}^*) = \mathbf{0}$$

$$\nabla_{\lambda_j} L(\boldsymbol{X}^*, \boldsymbol{\lambda}) = -h_j(\boldsymbol{X}^*) = 0, \; j = 1, 2, \cdots, l$$

这是等式约束极值问题的一阶必要条件.
Lagrange 乘子法将等式约束极值问题转化为无约束极值问题.
如何判断 Lagrange 函数的驻点是否为条件极值点? 参见下一节.

基本概念

考虑以下规划 ($f, g_i, h_j \in C^1$):

$$
\begin{aligned}
&\min f(\boldsymbol{X}) \\
&\text{s.t. } g_i(\boldsymbol{X}) \geq 0,\ i = 1, 2, \ldots, m \\
&\qquad h_j(\boldsymbol{X}) = 0,\ j = 1, 2, \ldots, l
\end{aligned}
$$

可行域 $S = \{\boldsymbol{X} \in \mathbb{R}^n \mid g_i(\boldsymbol{X}) \geq 0, i = 1, 2, \ldots, m; h_j(\boldsymbol{X}) = 0, j = 1, 2, \ldots, l\}$.

起作用约束

设可行解 $\boldsymbol{X}_0 \in S$. 考虑某不等式约束 $g_i(\boldsymbol{X}) \geq 0$:

- 若 $g_i(\boldsymbol{X}_0) = 0$, 则称 $g_i(\boldsymbol{X}) \geq 0$ 是在 $\boldsymbol{X}_0$ 的起作用约束.
- 若 $g_i(\boldsymbol{X}_0) > 0$, 则称 $g_i(\boldsymbol{X}) \geq 0$ 是在 $\boldsymbol{X}_0$ 的不起作用约束.

在可行解处, 等式约束都是起作用约束.

基本概念

设可行解 $\boldsymbol{X}_0 \in S$.

下降方向

若 $\boldsymbol{d} \neq \mathbf{0}$ 满足 $\exists \delta > 0$ s.t. $\forall \lambda \in (0, \delta)$, $f(\boldsymbol{X}_0 + \lambda \boldsymbol{d}) < f(\boldsymbol{X}_0)$, 则称 $\boldsymbol{d}$ 为 $f(\boldsymbol{X})$ 在 $\boldsymbol{X}_0$ 的下降方向.

集合 $F = \{\boldsymbol{d} \mid \nabla f(\boldsymbol{X}_0)^{\mathrm{T}} \boldsymbol{d} < 0\}$ 中的方向均为下降方向.

可行方向

若 $\boldsymbol{d} \neq \mathbf{0}$ 满足 $\exists \delta > 0$ s.t. $\forall \lambda \in (0, \delta)$, $\boldsymbol{X}_0 + \lambda \boldsymbol{d} \in S$, 则称 $\boldsymbol{d}$ 为 $\boldsymbol{X}_0$ 的可行方向.

- 对不等式约束, 集合 $G = \{\boldsymbol{d} \mid \nabla g_i(\boldsymbol{X}_0)^{\mathrm{T}} \boldsymbol{d} > 0,\ i \in \mathcal{I}\}$ ($\mathcal{I}$ 为 X_0 的起作用约束下标) 中的方向均为可行方向.
- 对等式约束, 可能不存在可行方向. 考虑切空间 $H = \{\boldsymbol{d} \mid \nabla h_j(\boldsymbol{X}_0)^{\mathrm{T}} \boldsymbol{d} = 0\}$ 中的方向 (假设 $\boldsymbol{X}_0$ 是正则点).

局部极小值点的梯度

一阶必要条件 (几何表达: 无可行下降方向)

若 $\boldsymbol{X}^*$ 是局部极小值点、正则点, 则 $F \cap G \cap H = \varnothing$, 其中

$$F = \{\boldsymbol{d} \mid \nabla f(\boldsymbol{X}^*)^{\mathrm{T}} \boldsymbol{d} < 0\}$$
$$G = \{\boldsymbol{d} \mid \nabla g_i(\boldsymbol{X}^*)^{\mathrm{T}} \boldsymbol{d} > 0,\ i \in \mathcal{I}\}$$
$$H = \{\boldsymbol{d} \mid \nabla h_j(\boldsymbol{X}^*)^{\mathrm{T}} \boldsymbol{d} = 0,\ j = 1, 2, \cdots, l\}$$

证明: 略.

局部极小值点的梯度

一阶必要条件 (代数表达: Fritz John 条件)

若 $\boldsymbol{X}^*$ 是局部极小值点, 则存在不全为 0 的 w_0, w_i $(i \in \mathcal{I}), v_j$ s.t.

$$w_0 \nabla f(\boldsymbol{X}^*) - \sum_{i \in \mathcal{I}} w_i \nabla g_i(\boldsymbol{X}^*) - \sum_{j=1}^{l} v_j \nabla h_j(\boldsymbol{X}^*) = \mathbf{0}$$

$$w_0 \geq 0, \ w_i \geq 0 \ (i \in \mathcal{I})$$

证明: 若 $\boldsymbol{X}^*$ 不是正则点, 取 $w_0 = w_i = 0$ $(i \in \mathcal{I})$, 取不全为 0 的 v_j s.t. $\sum_{j=1}^{l} v_j \nabla h_j(\boldsymbol{X}^*) = \mathbf{0}$. 若 $\boldsymbol{X}^*$ 是正则点, 利用几何表达、Motzkin 定理.

问题: 我们不关心 $w_0 = 0$ 的情况.

局部极小值点的梯度

一阶必要条件 (代数表达: Karush-Kuhn-Tucker 条件)

若 $\boldsymbol{X}^*$ 是局部极小值点, $\nabla g_i(\boldsymbol{X}^*)$ $(i \in \mathcal{I})$, $\nabla h_j(\boldsymbol{X}^*)$ $(j = 1, 2, \cdots, l)$ 线性无关, 则 $\exists w_i$ $(i \in \mathcal{I})$, v_j $(j = 1, 2, \cdots, l)$ s.t.

$$\nabla f(\boldsymbol{X}^*) - \sum_{i \in \mathcal{I}} w_i \nabla g_i(\boldsymbol{X}^*) - \sum_{j=1}^{l} v_j \nabla h_j(\boldsymbol{X}^*) = \mathbf{0}, \quad w_i \geq 0 \ (i \in \mathcal{I})$$

或表达为: $\exists w_i$ $(i = 1, 2, \cdots, m)$, v_j $(j = 1, 2, \cdots, l)$ s.t.

$$\nabla f(\boldsymbol{X}^*) - \sum_{i=1}^{m} w_i \nabla g_i(\boldsymbol{X}^*) - \sum_{j=1}^{l} v_j \nabla h_j(\boldsymbol{X}^*) = \mathbf{0}$$

$$w_i g_i(\boldsymbol{X}^*) = 0 \ (\text{互补松弛条件}), \quad w_i \geq 0 \ (i = 1, 2, \cdots, m)$$

KKT 条件相关知识

互补松弛条件

$w_i g_i(\boldsymbol{X}^*) = 0$ 称为互补松弛条件. 由于 $w_i \geq 0$, $g_i(\boldsymbol{X}^*) \geq 0$, 易得

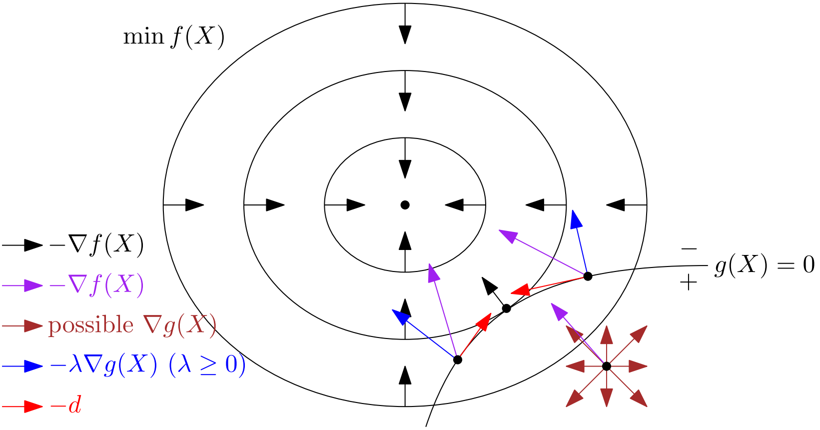
- 若 $g_i(\boldsymbol{X}^*) > 0$, 则该约束为不起作用约束, 删去它后 $\boldsymbol{X}^*$ 仍为局部极小值点, $\nabla f(\boldsymbol{X}^*)$ 不受影响, 即 $w_i = 0$.
- 若 $w_i > 0$, 则该约束为起作用约束, 对 $\nabla f(\boldsymbol{X}^*)$ 有影响, 即 $g_i(\boldsymbol{X}^*) = 0$.

KKT 乘子的经济学意义

$\boldsymbol{w}, \boldsymbol{v}$ 称为 KKT 乘子 (也可以称为 Lagrange 乘子).

- 约束: 资源.
- 约束对应的 KKT 乘子: 资源的影子价格 (边际价值).
- 不稀缺的资源无边际价值, 有边际价值的资源是稀缺的.

图解



$\min f(X)$
$-\nabla f(X)$
$-\nabla f(X)$
possible $\nabla g(X)$
$-\lambda\nabla g(X)$ $(\lambda \geq 0)$
$-d$
$-$
$+$
$g(X) = 0$

一阶必要条件 (KKT 条件)

总结: KKT 条件是 Lagrange 乘子法的推广. 设 Lagrange 函数

$$L(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{v}) = f(\boldsymbol{X}) - \sum_{i=1}^{m} w_i g_i(\boldsymbol{X}) - \sum_{j=1}^{l} v_j h_j(\boldsymbol{X})$$

则 KKT 条件为

$$\begin{aligned}
&\nabla_{\boldsymbol{X}} L(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{v}) = \mathbf{0} \\
&g_i(\boldsymbol{X}) \geq 0, \ i = 1, 2, \cdots, m \\
&h_j(\boldsymbol{X}) = 0, \ j = 1, 2, \cdots, l \\
&w_i g_i(\boldsymbol{X}) = 0, \ i = 1, 2, \cdots, m \\
&w_i \geq 0, \ i = 1, 2, \cdots, m
\end{aligned}$$

其它条件

一阶充分条件 (凸规划)

对于凸规划, 若 $\boldsymbol{X}^*$ 满足 KKT 条件, 则 $\boldsymbol{X}^*$ 为全局极小值点.

进一步假设 $f, g_i, h_j \in C^2$, $\nabla g_i(\boldsymbol{X}^*)$ $(i \in \mathcal{I})$, $\nabla h_j(\boldsymbol{X}^*)$ $(j = 1, 2, \cdots, l)$ 线性无关, 定义集合

$$\overline{T} = \left\{ \boldsymbol{d} \,\middle|\, \begin{array}{l} \nabla g_i(\boldsymbol{X}^*)^{\mathrm{T}} \boldsymbol{d} = 0,\ i \in \mathcal{I}, w_i^* > 0 \\ \nabla g_i(\boldsymbol{X}^*)^{\mathrm{T}} \boldsymbol{d} \geq 0,\ i \in \mathcal{I}, w_i^* = 0 \\ \nabla h_j(\boldsymbol{X}^*)^{\mathrm{T}} \boldsymbol{d} = 0,\ j = 1, 2, \cdots, l \end{array} \right\}$$

$$T = \overline{T} \setminus \{\mathbf{0}\}$$

注: $\overline{T} \neq \varnothing$, 但 T 可能为 $\varnothing$.

其它条件

二阶必要条件 (Lagrange 函数 Hesse 矩阵在 $\overline{T}$ 半正定)

若 $\boldsymbol{X}^*$ 为局部极小值点, KKT 乘子为 $\boldsymbol{w}^*, \boldsymbol{v}^*$, 则 $\forall \boldsymbol{d} \in \overline{T}$, $\boldsymbol{d}^{\mathrm{T}} \nabla^2_{\boldsymbol{X}} L(\boldsymbol{X}^*, \boldsymbol{w}^*, \boldsymbol{v}^*) \boldsymbol{d} \geq 0$.

二阶充分条件 (Lagrange 函数 Hesse 矩阵在 T 正定)

若 $\boldsymbol{X}^*$ 为可行解, $\exists \boldsymbol{w}^*, \boldsymbol{v}^*$ 使 KKT 条件成立,

- 若 $T = \varnothing$, 则 $\boldsymbol{X}^*$ 为 (严格) 局部极小值点.
- 若 $T \neq \varnothing$, $\forall \boldsymbol{d} \in T$, $\boldsymbol{d}^{\mathrm{T}} \nabla^2_{\boldsymbol{X}} L(\boldsymbol{X}^*, \boldsymbol{w}^*, \boldsymbol{v}^*) \boldsymbol{d} > 0$, 则 $\boldsymbol{X}^*$ 为 (严格) 局部极小值点.

若 $T \neq \varnothing$, Hesse 矩阵在切空间半正定, 则二阶条件无法判断. 此时可尝试将一些等式约束 $h_j = 0$ 代入 f, 具体问题具体分析.

极值问题总结

对于一般约束极值问题 (假设 C^2、正则性):

凸规划

1. 解一阶必要条件 (KKT 条件). 根据一阶充分条件, KKT 点即为最小值点 (全局极小值点).

一般规划 (求局部极小值点)

1. 解一阶必要条件 (KKT 条件), 计算切空间 $\overline{T}, T$.
2. 若 $T = \varnothing$, 则根据二阶充分条件, KKT 点即为 (严格) 局部极小值点. 若 $T \neq \varnothing$, 则验证二阶充分条件: 若成立, 则为 (严格) 局部极小值点.
3. 若二阶充分条件不成立, 则验证二阶必要条件: 若不成立, 则不为局部极小值点. 若成立, 则无法判断.

最值问题总结

最值问题：一般不用二阶条件判断极值点类型．首先判断存在性：

已知最值存在性

若已知存在性 (例如：f 连续，可行域是有界闭集)，则可求出最值点候选，再从中选出最值点．可用以下两种方法得到最值点候选：

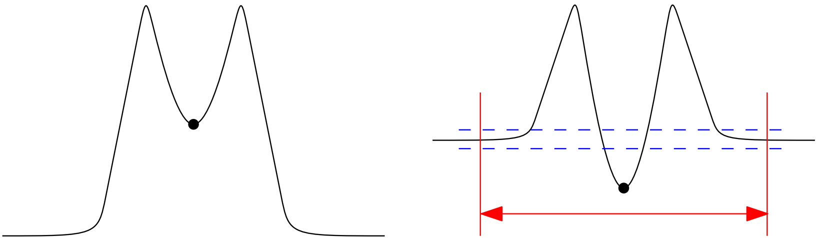
1. KKT 条件：KKT 点．
2. 传统方法：
   - 区域内部：f 的驻点．
   - 区域边界：各条边界上的可能的条件极值点 (Lagrange 函数的驻点)，边界的交点．
   - 性质不好的点：不可微点．

最值问题总结

未知最值存在性

若未知存在性 (例如: 可行域无界), 则需进一步分析, 需要额外考虑无穷远.

- 可尝试用某处函数值 (作为过渡) 控制足够远处, 讨论可行域的有界闭子集, 其上的最值为全局最值, 得到最值存在性.



总结

关系

最小值点 (全局极小值点) $\Longrightarrow$ 局部极小值点 $\xrightarrow{\text{正则点}}$ KKT 点.

极值问题核心思想

用充分条件证明是局部极小值点.
用必要条件证明不是局部极小值点.

最值问题核心思想

证明最值存在性, 找全最值点候选, 无需使用二阶条件.
不要墨守成规, 也可尝试直接放缩找最值 (需说明可以取等):

- 二元化一元、极坐标换元、均值不等式···