第一章：介绍

(3)：最优性条件

授课教师：曹语

课程主页：`https://yucaoyc.github.io/math3806`

优化问题

$$\min_{\boldsymbol{x}\in\Omega} \; f_0(\boldsymbol{x})$$

我们希望理解如下的问题:

Q1 **存在性:** 是否具有极小值?(能不能找到最低点?)

Q2 **最优性条件:** 若解有极小值，极小值点满足什么性质?(如果在谷底，周围的地势是怎么样的?)

想象我们在大雾的山中寻找最低点，我们怎么知道脚下的点是不是就是最低点呢?

回顾练习

练习： 分别描述 $\mathbb{R}^2$ 中 (画图) $S_p = \{\boldsymbol{x} \mid \|\boldsymbol{x}\|_p = 1, \boldsymbol{x} \in \mathbb{R}^2\} \quad (p = 1, 2, \infty)$

目录

基本概念

极值定理

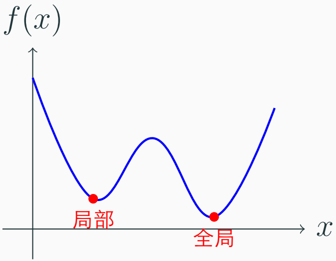
最优性条件

总结

全局最优解 vs 局部最优解

- **全局最小 (Global Minimum)**:
  在整个定义域 Ω 内最小。

- **局部最小 (Local Minimum)**:
  在附近邻域内最小。



全局最优解

定义 (最优点和最优集)

- 假设 $\boldsymbol{x}^* \in \Omega$ 满足 $f_0(\boldsymbol{x}^*) \leq f_0(\boldsymbol{z}), \forall \boldsymbol{z} \in \Omega$，则我们称该 $\boldsymbol{x}^*$ 为**最优点/全局最优解**
- 所有最优点的集合为**最优集**

练习： 请写出对如下例子的最优集：

- $f_0(x) = \log(x),\ \Omega = \mathbb{R}_{++} = (0, \infty)$
- $f_0(x) = \log(x),\ \Omega = (1, \infty)$
- $f_0(x) = \log(x),\ \Omega = [1, \infty)$

局部最优解

定义 (局部最优解)

若存在 $R > 0$ 使得对于任意的 $\boldsymbol{z} \in \Omega \cap B(\boldsymbol{x}, R)$, 皆有 $f_0(\boldsymbol{x}) \leq f_0(\boldsymbol{z})$，则我们称 $\boldsymbol{x}$ 为**局部最优解**。其中 $B(\boldsymbol{x}, R) = \{\boldsymbol{z} : \|\boldsymbol{z} - \boldsymbol{x}\|_2 \leq R\}$ 是以 $\boldsymbol{x}$ 为中心，半径为 R 的球体。

这里的 R 可以非常非常小。

该 $\boldsymbol{x}$ 是如下优化问题的最优解:

$$
\begin{aligned}
\text{minimize} \quad & f_0(\boldsymbol{z}) \\
\text{subject to} \quad & \boldsymbol{z} \in \Omega \\
& \|\boldsymbol{z} - \boldsymbol{x}\|_2 \leq R
\end{aligned}
$$

如果 $\boldsymbol{x}$ 是如上意义下的局部最优解，则存在 $R_p > 0$ 使得它同时是

$$\begin{array}{rl} \text{minimize} & f_0(\boldsymbol{z}) \\ \text{subject to} & \boldsymbol{z} \in \Omega \\ & \|\boldsymbol{z} - \boldsymbol{x}\|_p \leq R_p \end{array}$$

的最优解（即局部最优解的定义中，选择不同范数并不影响）

提示： $\|\boldsymbol{x}\|_\infty \leqslant \|\boldsymbol{x}\|_p \leqslant n^{1/p}\|\boldsymbol{x}\|_\infty$，其中 n 是向量的维数。

目录

基本概念

极值定理

最优性条件

总结

开集和闭集：直观理解

定义

- 对于集合 C，如果对于任意的 $\boldsymbol{x} \in C$，都存在 $\epsilon > 0$ 使得 $B(x, \epsilon) \subset C$，则称 $\boldsymbol{x}$ 为 C 的**内点**；
- C 的所有内点组成的集合被称为 C 的**内部**，符号为 $\mathbf{int}C$；
- 若 $C = \mathbf{int}C$，则称 C 为**开集**；(直观：不包含边界)
- 若其补集 $\mathbb{R}^n \backslash C$ 为开集，则称 C 为**闭集**。(直观：包含所有边界)

定理 (闭集的另一个定义)

集合 C 为闭集的一个充要条件是，对于任意一个序列 $\{\boldsymbol{x}_i\}_{i=1}^{\infty} \subset C$，若 $\lim_{i \to \infty} \|\boldsymbol{x} - \boldsymbol{x}_i\| = 0$，则 $\boldsymbol{x} \in C$。

直观："你怎么跑也跑不出去"——极限点也必须在集合内。

开集和闭集

例子： 请判断如下的集合是否为开集，是否为闭集?

(1) $\emptyset$

(2) $\{\boldsymbol{x}\}$

(3) $(1,2)$

(4) $(1,2]$

(5) $[1,2]\times[1,2]$

极值定理

定义 (紧集 Compact set)

若 $C \subset \mathbb{R}^n$，当 C 是有界的闭集，它又被称为紧集。

定理 (极值定理)

若 C 为紧集，函数 f 在 C 上为连续函数，则 f 的值是有界的，且在 C 中某点能取到极大值/极小值。

这就好比在一个有限的封闭山谷里，最低点一定存在。极值定理保障了对一类问题的最优解的存在性。
回应了前面提出的 Q1。

目录

基本概念

极值定理

最优性条件

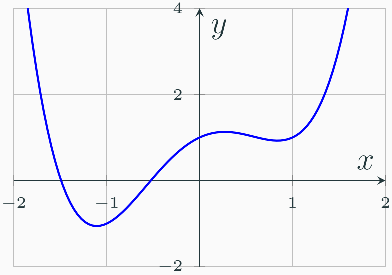
总结

一维函数的回忆

微积分里学过：考虑函数 $f:\mathbb{R}\to\mathbb{R}$，$x^*\in\mathbb{R}$,

- **必要条件**:
  - 一阶：若 x^* 是局部极小，则 $f'(x^*)=0$ (驻点)。
  - 二阶：若 x^* 是局部极小，则 $f''(x^*)\geq 0$ (开口向上)。
- **充分条件**:
  - 若 $f'(x^*)=0$ 且 $f''(x^*)>0$，则 x^* 是局部极小点。

$y(x)=(x^2-1)^2+x$



目标：将这些" 斜率为 0"、" 开口向上" 的概念推广到 n 维空间，即解决 Q2。

必要条件的反命题不一定对

我们假设函数 f 是光滑的，即任意阶导数存在且连续：

问题：$f'(x^*) = 0$ 能否推出 x^* 是局部极小点?

问题：$f'(x^*) = 0, f''(x^*) = 0$，能否推出 x^* 是局部极小点?

一阶条件：梯度

导数的 n 维推广是**梯度**，记为 ∇f

定义 (梯度)

若 $f \equiv f(x_1, \cdots, x_n)$ 偏导存在，则

$$\nabla f := \begin{bmatrix} \partial_{x_1} f \\ \vdots \\ \partial_{x_n} f \end{bmatrix}$$

例子：

- 若 $f(x) = \boldsymbol{v}^\top \boldsymbol{x} + b$，则 $\nabla f(x) = \boldsymbol{v}$；
- 若 $f(x) = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}$，则 $\nabla f(x) = (\boldsymbol{A} + \boldsymbol{A}^\top)\boldsymbol{x}$。

一阶必要条件

定理 (一阶必要条件)

- 假设目标函数 f_0 可微，$\boldsymbol{x}^*$ 是局部极小点，则对于任何合适的单位向量 $\boldsymbol{v}$，$\boldsymbol{v}^\top \nabla f_0(\boldsymbol{x}^*) \geq 0$（其中合适的 $\boldsymbol{v}$ 的含义是满足只要 ϵ 足够小，$\boldsymbol{x}^* + \epsilon \boldsymbol{v} \in \Omega$）。
- 若任意的单位向量 $\boldsymbol{v}$ 都是可行的，则 $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$；这样的点我们称为**驻点 (Stationary Point)**。

直观解释:

- 如果在某个方向 v 上，方向导数 $\nabla f^T v < 0$，说明顺着 v 走函数值会下降，那这就不是最低点。
- 如果是内部点，所有方向都能走，所以必须各个方向看来都是平的。

证明.

设 $\boldsymbol{x}^*$ 是局部极小点，则存在 $R > 0$ 使得对于任意的 $\boldsymbol{z} \in \Omega \cap B(\boldsymbol{x}^*, R)$, 皆有 $f_0(\boldsymbol{x}^*) \leq f_0(\boldsymbol{z})$。对于任意合适的单位向量 $\boldsymbol{v}$，当 ϵ 足够小时，$\boldsymbol{x}^* + \epsilon \boldsymbol{v}$ 在局部区域内，因此函数

$$g(\epsilon) := f_0(\boldsymbol{x}^* + \epsilon \boldsymbol{v})$$

应该在 $\epsilon \geq 0$ 的局部区域内，在 $\epsilon = 0$ 处取得最小值。我们有 $g'(0) \geq 0$，即

$$g'(0) = \nabla f_0(\boldsymbol{x}^*)^\top \boldsymbol{v} \geq 0$$

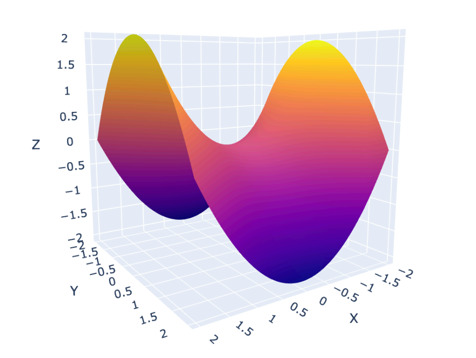
$\square$

问题： 那么反过来结论是否对呢？

反例：鞍点 (Saddle Point)

对于 $f(x,y) = (x^2 - y^2)/2$，$(0,0)$ 满足一阶条件 ($\nabla f = 0$)，但不是最优点。

- 沿 x 轴看是极小 (x^2)
- 沿 y 轴看是极大 ($-y^2$)



Hessian 矩阵

定义 (Hessian 矩阵)

二阶导数的 n 维拓展是 Hessian 矩阵，定义为

$$\nabla^2 f := \begin{bmatrix} \partial_{x_1,x_1} f & \cdots & \partial_{x_1,x_n} f \\ \vdots & \ddots & \vdots \\ \partial_{x_n,x_1} f & \cdots & \partial_{x_n,x_n} f \end{bmatrix}$$

若 f 具有连续二阶导，则 $\nabla^2 f$ 为对称矩阵。

例子：

- 若 $f(x) = \boldsymbol{v}^\top \boldsymbol{x} + b$，则 $\nabla^2 f(x) = \mathbf{0}_{n\times n}$
- 若 $f(x) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$，则 $\nabla^2 f(x) = \boldsymbol{A} + \boldsymbol{A}^\top$

类似的，若 x^* 是局部最优解，则对于函数

$$g(\epsilon) := f_0(\boldsymbol{x}^* + \epsilon \boldsymbol{v}) \tag{1}$$

其开口向上，即 $g''(0) \geq 0$，因此

$$g''(0) = \boldsymbol{v}^\top \nabla^2 f_0(\boldsymbol{x}^*) \boldsymbol{v} \geq 0$$

因此，我们得到如下结论：

定理 (二阶必要条件)

假设目标函数 f_0 具有连续的二阶导，$\boldsymbol{x}^*$ 是局部极小点，则对于任何合适的单位向量 $\boldsymbol{v}$，$\boldsymbol{v}^\top \nabla^2 f_0(x^*) \boldsymbol{v} \geq 0$（其中合适的 $\boldsymbol{v}$ 的含义是满足只要 ϵ 足够小，$\boldsymbol{x}^* + \epsilon \boldsymbol{v} \in \Omega$）。

正定与半正定：矩阵的“符号”

定义

假设矩阵 $\boldsymbol{A}$ 是对称矩阵：

- 若对于任意向量 $\boldsymbol{v}$，$\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} \geq 0$，则称 $\boldsymbol{A}$ 为**半正定**(PSD)，$\boldsymbol{A} \succeq 0$。
- 若对于任意非零 $\boldsymbol{v}$，$\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} > 0$，则称为**正定**(PD)，$\boldsymbol{A} \succ 0$。

因此，上述定理中的二阶必要条件可以表述为：$\nabla^2 f_0(\boldsymbol{x}^*) \succeq 0$，即 Hessian 矩阵是半正定的。

特征值与正定性

定义 (特征值与特征向量)

- 对于矩阵 $\boldsymbol{A}$，如果存在标量 λ 和非零向量 $\boldsymbol{v}$ 使得 $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$，则称 λ 是 $\boldsymbol{A}$ 的**特征值**，$\boldsymbol{v}$ 是对应的**特征向量**。
- 特征值满足如下方程：

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$$

- 特征向量是 $\boldsymbol{A} - \lambda\boldsymbol{I}$ 的零空间中的向量。

定理 (特征值与正定性的关系)

$\boldsymbol{A}$ 为半正定 *(正定)* 等价于其特征值非负 *(全为正)*。

证明.

设对称 $\boldsymbol{A}$ 的特征值为 $\lambda_1, \cdots, \lambda_n$，对应的单位特征向量为 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$，则 $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ 能够构成 $\mathbb{R}^n$ 的一组基底（该证明将作为作业）。

- （非负特征值 $\to$ 半正定）因此对于任意向量 $\boldsymbol{v}$，存在 $\alpha_1, \cdots, \alpha_n$ 使得 $\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$，因此

$$\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} = \sum_{i=1}^{n} \lambda_i \alpha_i^2 \geq 0$$

- （半正定 $\to$ 非负特征值）反过来，设 $\boldsymbol{v}$ 为 $\boldsymbol{A}$ 的特征向量，λ 为对应的特征值，则 $\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} = \lambda \|\boldsymbol{v}\|^2 \geq 0$，因此 $\lambda \geq 0$。

$\square$

练习： 判断 $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ 是否为半正定矩阵。

二阶充分条件：如何确认是极小值

定理 (二阶充分条件)

假设目标函数 f_0 具有连续的二阶导数，若 $\boldsymbol{x}^*$ 满足：

1. **一阶条件**： $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ （驻点）
2. **二阶条件**： $\nabla^2 f_0(\boldsymbol{x}^*) \succ \mathbf{0}$ （*Hessian* 矩阵正定）

则 $\boldsymbol{x}^*$ 一定是局部极小点。

这相当于一维中的：$f'(x) = 0$ 且 $f''(x) > 0$。
正定保证了在所有方向上，函数值都在“向上弯曲”。

练习

练习： 请使用前面的定理来说明为什么 $(0,0)$ 不可能是

$$f(x,y)=(x^2-y^2)/2$$

的局部极小点。

在上述计算中，我们已经隐式地使用了多元函数的泰勒展开，总结如下：

定理 (多元函数的泰勒展开)

设 f 在 $\boldsymbol{x}$ 的某个邻域内具有连续的二阶导数，则对于任意 $\boldsymbol{v}$ 使得当 $\|\boldsymbol{v}\|$ 足够小时，$\boldsymbol{x}+\boldsymbol{v}$ 在该邻域内，且

$$f(\boldsymbol{x}+\boldsymbol{v}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top \boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x})\boldsymbol{v} + o(\|\boldsymbol{v}\|^2).$$

目录

基本概念

极值定理

最优性条件

总结

总结

- 掌握最优点（全局最优解），局部最优解，开/闭集的定义和含义，极值定理，以及 ∇f，Hessian $\nabla^2 f$ 等概念。
- 掌握最优性条件的 3 个结论，并且能知道基本的证明思路。

阅读作业 & 参考资料：

- 课本 4.1.1 章，以及附录 A.2.1，附录 A.4

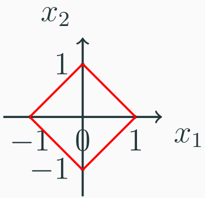
参考答案

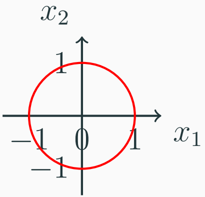
- 第3页：

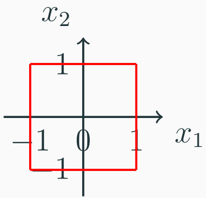
$$S_1 = \left\{ x \in \mathbb{R}^2 \;\middle|\; \|x\|_1 = 1 \right\} = \left\{ x \in \mathbb{R}^2 \;\middle|\; |x_1| + |x_2| = 1 \right\}$$

$$S_2 = \left\{ x \in \mathbb{R}^2 \;\middle|\; \|x\|_2 = 1 \right\} = \left\{ x \in \mathbb{R}^2 \;\middle|\; \sqrt{x_1^2 + x_2^2} = 1 \right\}$$

$$S_\infty = \left\{ x \in \mathbb{R}^2 \;\middle|\; \|x\|_\infty = 1 \right\} = \left\{ x \in \mathbb{R}^2 \;\middle|\; \max\{|x_1|, |x_2|\} = 1 \right\}$$







参考答案

- 第6页：$\emptyset$, $\emptyset$, $\{1\}$
- 第11页：(1) 是开集，是闭集；(2) 不是开集，是闭集；(3) 是开集，不是闭集；(4) 不是开集，不是闭集；(5) 不是开集，是闭集
- 第15页：对 $f(x) = x^3$，显然 $x = 0$ 满足一阶条件，但该点不是极小点。
- 第25页：由于 $\boldsymbol{A}$ 的特征值为 $3, -1$，因此该矩阵不是半正定。
- 第27页：假设 $(0, 0)$ 是局部极小点，则 $\nabla^2 f(0, 0) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ 需要是半正定的，但显然 $\nabla^2 f(0, 0)$ 不是，因此得到矛盾。