第一章：介绍

(2)：最优性条件

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 是否具有极小值?

Q2 若解有极小值，极小值点满足什么性质?

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

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**，所有最优点的集合为**最优集**。最优点也被称为**全局最优解**。

练习：请写出对如下例子的最优集：

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

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 $B(\boldsymbol{x}, R) = \left\{\boldsymbol{z} : \|\boldsymbol{z} - \boldsymbol{x}\|_2 \leq R\right\}$ 是以 $\boldsymbol{x}$ 为中心，半径为 R 的球体。

即该 $\boldsymbol{x}$ 是如下优化问题的最优解:

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

探究：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

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

的最优解（即局部最优解的定义中，选择不同范数并不影响结论）。

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

目录

开集和闭集

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 $\textbf{int}C$；
- 若 $C = \textbf{int}C$，则称 C 为开集；
- 若其补集 $\mathbb{R}^n \backslash C$ 为开集，则称 C 为闭集。

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 $\boldsymbol{x} \in C$。

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

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。

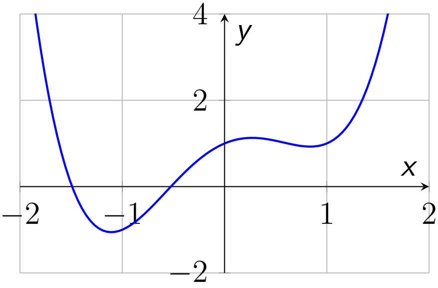
目录

一维函数的最优性条件

在微积分里我们学过：考虑函数 $f:\mathbb{R}\to\mathbb{R}$，$x^\star\in\mathbb{R}$，

- 若 f' 存在并连续，且 $x^\star$ 是局部极小点，则 $f'(x^\star)=0$；
- 若 f'' 存在并连续，且 $x^\star$ 是局部极小点，则 $f''(x^\star)\geq 0$；
- 若 $f'(x^\star)=0$，且 $f''(x^\star)>0$，则 $x^\star$ 一定是局部极小点。

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



目标：现实中我们往往需要处理多个变量，所以需要将这些概念拓展到 n 维

测试

我们假设函数 f 是光滑的，即任意阶导数存在且连续：

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

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

一阶条件

定义

若 $f \equiv f(x_1, x_2, \cdots, x_n)$ 是 n 元函数，且其偏导存在，则我们定义

$$\nabla f := \begin{bmatrix} \partial_{x_1} f \\ \partial_{x_2} f \\ \vdots \\ \partial_{x_n} f \end{bmatrix}$$

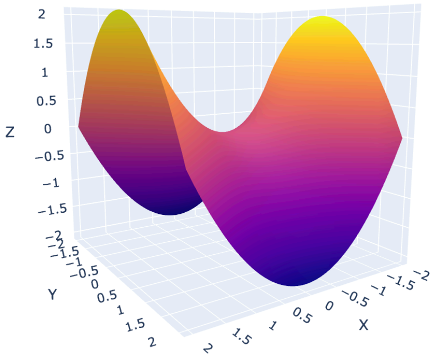
- 若 $f(x) = \boldsymbol{v}^\top \boldsymbol{x} + b$，则 $\nabla f(x) = \boldsymbol{v}$；
- 若 $f(x) = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}$，则 $\nabla f(x) = (\boldsymbol{A} + \boldsymbol{A}^\top)\boldsymbol{x}$。

定理 (一阶条件)

- 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 $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$。

证明见课内

例子：对于 $f(x,y) = (x^2 - y^2)/2$，$(0,0)$ 满足一阶条件，但该点不是最优点（它被称为鞍点）



二阶条件

定义 (Hessian 矩阵)

若 $f = f(x_1, x_2, \cdots, x_n)$ 是 n 元函数，且其二阶偏导存在，则我们定义 Hessian 矩阵 $\nabla^2 f$ 为

$$\nabla^2 f := \begin{bmatrix} \partial_{x_1,x_1} f & \partial_{x_1,x_2} f & \cdots & \partial_{x_1,x_n} f \\ \partial_{x_2,x_1} f & \partial_{x_2,x_2} f & \cdots & \partial_{x_2,x_n} f \\ \vdots & \vdots & \cdots & \vdots \\ \partial_{x_n,x_1} f & \partial_{x_n,x_2} f & \cdots & \partial_{x_n,x_n} f \end{bmatrix}$$

若 f 具有连续的二阶导，则 $\partial_{x_i,x_j} f = \partial_{x_j,x_i} f$，因此 $\nabla^2 f$ 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 $\nabla^2 f(x) = \boldsymbol{A} + \boldsymbol{A}^\top$

定义

假设矩阵 $\boldsymbol{A}$ 是对称矩阵：

- 若对于任意的向量 $\boldsymbol{v}$，$\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} \geq 0$，则我们称该矩阵 $\boldsymbol{A}$ 为半正定矩阵(positive semi-definite)，符号为 $\boldsymbol{A} \succeq \mathbf{0}_{n\times n}$（或者 $\mathbf{0}_{n\times n} \preceq \boldsymbol{A}$）。
- 若对于任意非零向量 $\boldsymbol{v}$，$\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} > 0$，则我们称该矩阵为（严格）正定矩阵（positive definite)，符号为 $\boldsymbol{A} \succ \mathbf{0}_{n\times n}$（或者 $\mathbf{0}_{n\times n} \prec \boldsymbol{A}$）。

定理

假设矩阵 $\boldsymbol{A}$ 是对称矩阵，

- $\boldsymbol{A}$ 为半正定矩阵 $\Longleftrightarrow$ 其特征值非负；
- $\boldsymbol{A}$ 为正定矩阵 $\Longleftrightarrow$ 其特征值为（严格）正数。

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

定理

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，$\boldsymbol{x}^* + \epsilon\boldsymbol{v} \in \Omega$）。

证明见课内。

练习： 请使用面前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。

定理

假设目标函数 f_0 具有连续的二阶导数，$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$，且 $\nabla^2 f_0(\boldsymbol{x}^*) \succ \mathbf{0}_{n\times n}$，则 $\boldsymbol{x}^*$ 为局部极小点。

证明见课内

目录

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
- https://jhc.sjtu.edu.cn/public/home/kuanyang/teaching/MATH3806/notes/lec02.pdf

参考答案

- 第5页：$\emptyset$, $\emptyset$, $\{1\}$, $\{e^{-1}\}$
- 第10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14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20页：由于 $\boldsymbol{A}$ 的特征值为 $3,-1$，因此该矩阵不是半正定。
- 第22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。