矩阵论

2024年秋学期

第七讲

2024年9月30日

第3章 矩阵微分

第4章 梯度分析与最优化

一阶实矩阵微分：Jacobian矩阵辨识

标量函数的Jacobian矩阵辨识

$$\mathrm{d}f(\boldsymbol{x})=\mathrm{tr}(\boldsymbol{A}\mathrm{d}\boldsymbol{x}) \Longleftrightarrow \mathrm{D}_{\boldsymbol{x}}f(\boldsymbol{x})=\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}=\boldsymbol{A}$$

$$\mathrm{d}f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{A}\mathrm{d}\boldsymbol{X}) \Longleftrightarrow \mathrm{D}_{\boldsymbol{X}}f(\boldsymbol{X})=\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^{\mathrm{T}}}=\boldsymbol{A}$$

$$\mathrm{D}_{\boldsymbol{X}}f(\boldsymbol{X})=\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^{\mathrm{T}}}=\boldsymbol{A} \Longleftrightarrow \nabla_{\boldsymbol{X}}f(\boldsymbol{X})=\boldsymbol{A}^{\mathrm{T}}$$

习题1：利用矩阵微分证明二次型函数$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 变元向量 $\boldsymbol{x}$ 的梯度向量

$$\nabla_{\boldsymbol{x}}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\right)=\left(\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\right)\boldsymbol{x}$$

习题2：利用Jacobian矩阵辨识，求包含逆矩阵的迹函数 $\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\right)$的梯度矩阵

一阶实矩阵微分：Jacobian矩阵辨识

示例 已知 $\frac{\partial f}{\partial \boldsymbol{Y}}$ 和 $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$，求 $\frac{\partial f}{\partial \boldsymbol{X}}$?

解 $$\mathrm{d}f = \mathbf{tr}\left[\left(\frac{\partial f}{\partial \boldsymbol{Y}}\right)^{\mathrm{T}} \mathrm{d}\boldsymbol{Y}\right] = \mathbf{tr}\left[\left(\frac{\partial f}{\partial \boldsymbol{Y}}\right)^{\mathrm{T}} \boldsymbol{A}\mathrm{d}\mathbf{X}\boldsymbol{B}\right] = \mathbf{tr}\left[\boldsymbol{B}\left(\frac{\partial f}{\partial \boldsymbol{Y}}\right)^{\mathrm{T}} \boldsymbol{A}\mathrm{d}\mathbf{X}\right]$$

$$= \mathbf{tr}\left[\left(\boldsymbol{A}^{\mathrm{T}} \frac{\partial f}{\partial \boldsymbol{Y}} \boldsymbol{B}^{\mathrm{T}}\right)^{\mathrm{T}} \mathrm{d}\mathbf{X}\right],$$

迹的性质：交换顺序

得 $\frac{\partial f}{\partial \boldsymbol{X}} = \boldsymbol{A}^{\mathrm{T}} \frac{\partial f}{\partial \boldsymbol{Y}} \boldsymbol{B}^{\mathrm{T}}$。

一阶实矩阵微分：Jacobian矩阵辨识

示例 已知 $f=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{b}$，求 $\frac{\partial f}{\partial \boldsymbol{X}}$?

解 $\mathrm{d}f=\mathrm{d}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{b}+\boldsymbol{a}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\boldsymbol{b}+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\mathrm{d}\boldsymbol{b}=\mathbf{tr}\left[\boldsymbol{a}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\boldsymbol{b}\right]$

0

$=\mathbf{tr}\left[\boldsymbol{b}\boldsymbol{a}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]=\mathbf{tr}\left[\left(\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}\right)^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]$

得 $\frac{\partial f}{\partial \boldsymbol{X}}=\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$。

笔记 上述两个例子反复应用了 $\mathbf{tr}(\boldsymbol{A}\boldsymbol{B})=\mathbf{tr}(\boldsymbol{B}\boldsymbol{A})$（$\boldsymbol{A}\boldsymbol{B}$ 和 $\boldsymbol{B}\boldsymbol{A}$ 特征值相同）。

一阶实矩阵微分：Jacobian矩阵辨识

示例 已知 $f=\mathbf{tr}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{X}\right]$，求 $\frac{\partial f}{\partial \boldsymbol{X}}$?

解 $\mathrm{d}f=\mathbf{tr}\left[\mathrm{d}\left(\boldsymbol{X}^{\mathrm{T}}\right)\boldsymbol{S}\boldsymbol{X}\right]+\mathbf{tr}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\mathrm{d}\boldsymbol{X}\right]=\mathbf{tr}\left[(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{S}\boldsymbol{X}\right]+\mathbf{tr}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\mathrm{d}\boldsymbol{X}\right]$
$=\mathbf{tr}\left[(\boldsymbol{S}\boldsymbol{X})^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]+\mathbf{tr}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\mathrm{d}\boldsymbol{X}\right]=\mathbf{tr}\left[2(\boldsymbol{S}\boldsymbol{X})^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]$，得 $\frac{\partial f}{\partial \boldsymbol{X}}=2\boldsymbol{S}\boldsymbol{X}$。

示例 已知 $f=\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2$（实数域），求 $\frac{\partial f}{\partial \boldsymbol{X}}$?

解 $\mathrm{d}f=\mathrm{d}\left[(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\right]=\mathrm{d}(\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})+(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{x}$
$=(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{x}+(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{x}=\mathbf{tr}\left[2\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}\right)^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]$，
得 $\frac{\partial f}{\partial \boldsymbol{X}}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$。

Hessian矩阵

定义 假设有一实值函数 $f(x_1,\cdots,x_n)$，如果 f 的所有二阶偏导数都存在并在定义域内连续，那么函数 f 的**海森矩阵**（Hessian matrix）为

$$\boldsymbol{H}[f(\boldsymbol{x})]=\frac{\partial^2 f(\boldsymbol{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}=\frac{\partial}{\partial\mathbf{x}}\left[\frac{\partial f(\boldsymbol{x})}{\partial\mathbf{x}^{\mathrm{T}}}\right]\in\mathbb{R}^{m\times m}$$

$$\boldsymbol{H}[f(\boldsymbol{x})]=\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})=\nabla_{\boldsymbol{x}}\left(\mathrm{D}_{\boldsymbol{x}} f(\boldsymbol{x})\right)$$

显然海森矩阵是一个对称矩阵

$$[\boldsymbol{H}f(\boldsymbol{x})]_{i,j}=\left[\frac{\partial^2 f(\boldsymbol{x})}{\partial\mathbf{x}\partial\mathbf{x}^T}\right]_{i,j}=\frac{\partial}{\partial x_i}\left[\frac{\partial f(\boldsymbol{x})}{\partial x_j}\right]$$

$$\boldsymbol{H}[f(\boldsymbol{x})]=\frac{\partial^2 f(\boldsymbol{x})}{\partial\mathbf{x}\partial\mathbf{x}^T}=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_m}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2 f}{\partial x_m\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_m\partial x_m}\end{bmatrix}\in\mathbb{R}^{m\times m}$$

示例 $f=x^2-y^2$，$\boldsymbol{H}_f=\begin{bmatrix}2 & 0\\ 0 & -2\end{bmatrix}$。

共轭梯度

实解析函数：对于实变量域内都是实解析的，但对于复变量不一定是复解析（全纯）的。

复解析在现代数学中常用“全纯”代替，复解析函数常称为全纯函数。

令复变函数 $f(z)$ 可以用实部 $u(x,y)$ 和虚部 $v(x,y)$ 写作

$$f(z)=u(x,y)+\mathrm{j}v(x,y)$$

式中 $z=x+\mathrm{j}y$ ，并且 $u(x,y)$ 和 $v(x,y)$ 分别是实值函数。

如果 $f(z)$在某区域内满足柯西-黎曼方程:

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y},\quad \frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}$$

那么 $f(z)$ 在该区域内解析 (Analytic)，即 $f(z)$ 在该区域内可微。

共轭梯度

形式偏导定义(在复变函数论中，形式偏导（Formal Partial Derivatives）的概念需要特别注意，因为复变函数的性质和实变函数的性质有显著不同。)

共轭梯度算子 $$\nabla_{z^*} = \frac{\partial}{\partial z^*} = \frac{1}{2}\left(\frac{\partial}{\partial x} + \mathrm{j}\frac{\partial}{\partial y}\right)$$

利用柯西黎曼条件

梯度算子 $$\nabla_{z} = \frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \mathrm{j}\frac{\partial}{\partial y}\right)$$

共轭梯度

实部与虚部的独立性假设

$$\frac{\partial x}{\partial y}=0 \quad 和 \quad \frac{\partial y}{\partial x}=0$$

$$\frac{\partial z}{\partial z^*}=\frac{\partial x}{\partial z^*}+\mathrm{j}\frac{\partial y}{\partial z^*}=\frac{1}{2}\left(\frac{\partial x}{\partial x}+\mathrm{j}\frac{\partial x}{\partial y}\right)+\mathrm{j}\frac{1}{2}\left(\frac{\partial y}{\partial x}+\mathrm{j}\frac{\partial y}{\partial y}\right)=\frac{1}{2}(1+0)+\mathrm{j}\frac{1}{2}(0+\mathrm{j})$$

$$\frac{\partial z^*}{\partial z}=\frac{\partial x}{\partial z}-\mathrm{j}\frac{\partial y}{\partial z}=\frac{1}{2}\left(\frac{\partial x}{\partial x}-\mathrm{j}\frac{\partial x}{\partial y}\right)-\mathrm{j}\frac{1}{2}\left(\frac{\partial y}{\partial x}-\mathrm{j}\frac{\partial y}{\partial y}\right)=\frac{1}{2}(1-0)-\mathrm{j}\frac{1}{2}(0-\mathrm{j})$$

$$\frac{\partial z}{\partial z^*}=0 \qquad \frac{\partial z^*}{\partial z}=0$$

z 和 z^* 是两个相互独立的变量

共轭梯度

任何一个非全纯的复变函数 $f(z)$ 写成 $f(z,z^*)$ 后，都变成了全纯函数

单个复变量的梯度

$$\nabla_z f(z,z^*) = \left.\frac{\partial f(z,z^*)}{\partial z}\right|_{z^*=\text{常数}}, \quad \nabla_{z^*} f(z,z^*) = \left.\frac{\partial f(z,z^*)}{\partial z^*}\right|_{z=\text{常数}}$$

单个复变量的微分 $\mathrm{d}f(z,z^*) = \frac{\partial f(z,z^*)}{\partial z}\mathrm{d}z + \frac{\partial f(z,z^*)}{\partial z^*}\mathrm{d}z^*$

共轭梯度

标量函数的梯度向量和共轭梯度向量

$$\nabla_{\mathbf{z}} f\left(\mathbf{z},\mathbf{z}^*\right)=\left.\frac{\partial f\left(\mathbf{z},\mathbf{z}^*\right)}{\partial \mathbf{z}}\right|_{\mathbf{z}^*=\text{常数向量}}=\left(\mathrm{D}_{\mathbf{z}} f\left(\mathbf{z},\mathbf{z}^*\right)\right)^{\mathrm{T}}$$

$$\nabla_{\mathbf{z}^*} f\left(\mathbf{z},\mathbf{z}^*\right)=\left.\frac{\partial f\left(\mathbf{z},\mathbf{z}^*\right)}{\partial \mathbf{z}^*}\right|_{\mathbf{z}=\text{常数向量}}=\left(\mathrm{D}_{\mathbf{z}^*} f\left(\mathbf{z},\mathbf{z}^*\right)\right)^{\mathrm{T}}$$

其中梯度算子 $$\nabla_{\mathbf{z}}=\frac{\partial}{\partial \mathbf{z}}=\frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{x}}-\mathrm{j}\frac{\partial}{\partial \boldsymbol{y}}\right)$$

共轭梯度算子 $$\nabla_{\mathbf{z}^*}=\frac{\partial}{\partial \mathbf{z}^*}=\frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{x}}+\mathrm{j}\frac{\partial}{\partial \boldsymbol{y}}\right)$$

共轭梯度

复变元向量的微分

$$\mathrm{d}f\left(\mathbf{z},\mathbf{z}^*\right)=\left[\frac{f\left(\mathbf{z},\mathbf{z}^*\right)}{\partial z_1},\cdots,\frac{f\left(\mathbf{z},\mathbf{z}^*\right)}{\partial z_m}\right]\begin{bmatrix}\mathrm{d}z_1\\ \vdots\\ \mathrm{d}z_m\end{bmatrix}+\left[\frac{f\left(\mathbf{z},\mathbf{z}^*\right)}{\partial z_1^*},\cdots,\frac{f\left(\mathbf{z},\mathbf{z}^*\right)}{\partial z_m^*}\right]\begin{bmatrix}\mathrm{d}z_1^*\\ \vdots\\ \mathrm{d}z_m^*\end{bmatrix}$$

$$=\frac{\partial f\left(\mathbf{z},\mathbf{z}^*\right)}{\partial \mathbf{z}^{\mathrm{T}}}\mathrm{d}\mathbf{z}+\frac{\partial f\left(\mathbf{z},\mathbf{z}^*\right)}{\partial \mathbf{z}^{\mathrm{H}}}\mathrm{d}\mathbf{z}^*$$

$\boldsymbol{z}$ 和 $\boldsymbol{z}^*$ 是两个相互独立的向量

$$\frac{\partial \mathbf{z}^{\mathrm{T}}}{\partial \mathbf{z}}=\frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \mathbf{z}}+\mathrm{j}\frac{\partial \mathbf{y}^{\mathrm{T}}}{\partial \mathbf{z}}=\frac{1}{2}\left(\frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \mathbf{x}}-\mathrm{j}\frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \mathbf{y}}\right)+\mathrm{j}\frac{1}{2}\left(\frac{\partial \mathbf{y}^{\mathrm{T}}}{\partial \mathbf{x}}-\mathrm{j}\frac{\partial \mathbf{y}^{\mathrm{T}}}{\partial \mathbf{y}}\right)=\boldsymbol{I}_{m\times m}$$

$$\frac{\partial \mathbf{z}^{\mathrm{T}}}{\partial \mathbf{z}^*}=\frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \mathbf{z}^*}+\mathrm{j}\frac{\partial \mathbf{y}^{\mathrm{T}}}{\partial \mathbf{z}^*}=\frac{1}{2}\left(\frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \mathbf{x}}+\mathrm{j}\frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \mathbf{y}}\right)+\mathrm{j}\frac{1}{2}\left(\frac{\partial \mathbf{y}^{\mathrm{T}}}{\partial \mathbf{x}}+\mathrm{j}\frac{\partial \mathbf{y}^{\mathrm{T}}}{\partial \mathbf{y}}\right)=\boldsymbol{O}_{m\times m}$$

第四章：梯度分析与最优化

最优化：极大值或极小值

主要讨论：

- **极值存在的条件（梯度分析）**
- **优化算法**

典型的优化问题

$$\min_{\boldsymbol{x} \in \mathcal{D}} f(\boldsymbol{x})$$ n维实数空间上的实值函数，逼近原始目标的简化目标函数

无约束优化问题的解决方法涉及计算目标函数的梯度或导数，并使用数值优化算法来找到梯度为零的点或局部最小值/最大值。常见的无约束优化算法包括梯度下降法、牛顿法、拟牛顿法等。

- 梯度下降法：一种迭代方法，每一步沿目标函数梯度的负方向更新变量，直到收敛。
- 牛顿法：利用目标函数的一阶和二阶导数信息来加速寻找最优点，特别适用于二次可微函数。
- 拟牛顿法：牛顿法的变种，不直接计算海森矩阵，而是通过近似来求解，减少了计算量

单变量函数

一个简单的例子是求解二次函数的最小值: $\min f(x) = ax^2 + bx + c$

其中 a,b,c 是常数，而 x 是变量。通过求解 $f'(x) = 0$ ，可以找到最小值点的 x 值。

函数 $f(x)$ 有最小值的条件是什么?

单变量函数

单变量函数的平稳点与极值点

- **全局极小点** $f\left(x^{\star}\right) \leqslant f(x), \quad \forall x \in \mathcal{D}, x \neq x^{\star}$

- **严格全局极小点** $f\left(x^{\star}\right) < f(x), \quad \forall\, x \in \mathcal{D}$

- **（开）邻域** $B_o(c;r) = \{x \mid x \in \mathcal{D}, |x-c| < r\}$ ⟵ **正数**

- **闭邻域** $B_c(c;r) = \{x \mid x \in \mathcal{D}, |x-c| \leqslant r\}$

单变量函数

单变量函数的平稳点与极值点

点 c 称为函数 $f(x)$ 的一个局部极小 (或极大) 点, 若

$$f(c) \leqslant f(c+\Delta x) \quad \text{或} \quad f(c) \geqslant f(c+\Delta x))$$

$f(c)$ ：**极小值（或极大值）**

对满足 $0<|\Delta x|\leqslant r$ 的所有 Δx 均成立。

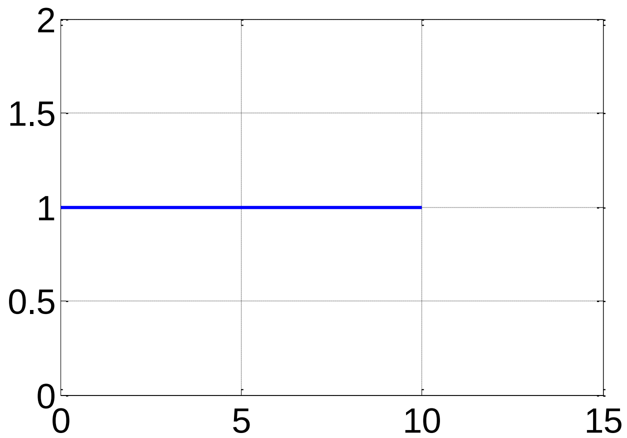
点 c 称为函数 $f(x)$ 的一个严格局部极小点, 若

$$f(c) < f(c+\Delta x)$$

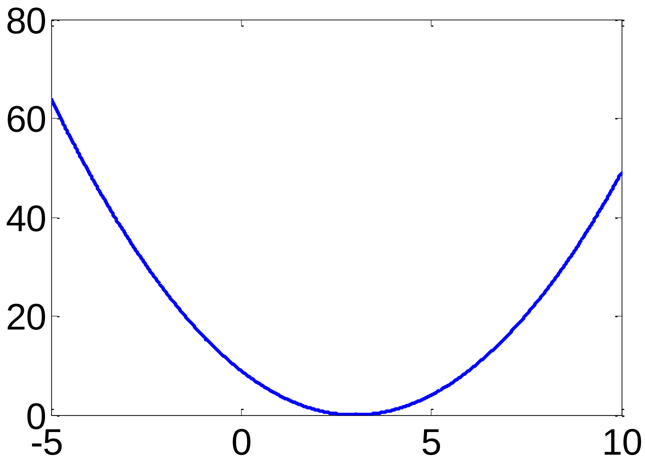
对满足 $0<|\Delta x|\leqslant r$ 的所有 Δx 均成立。

单变量函数

单变量函数的平稳点与极值点



（弱）局部极小点



严格局部极小点

单变量函数

单变量函数的平稳点与极值点

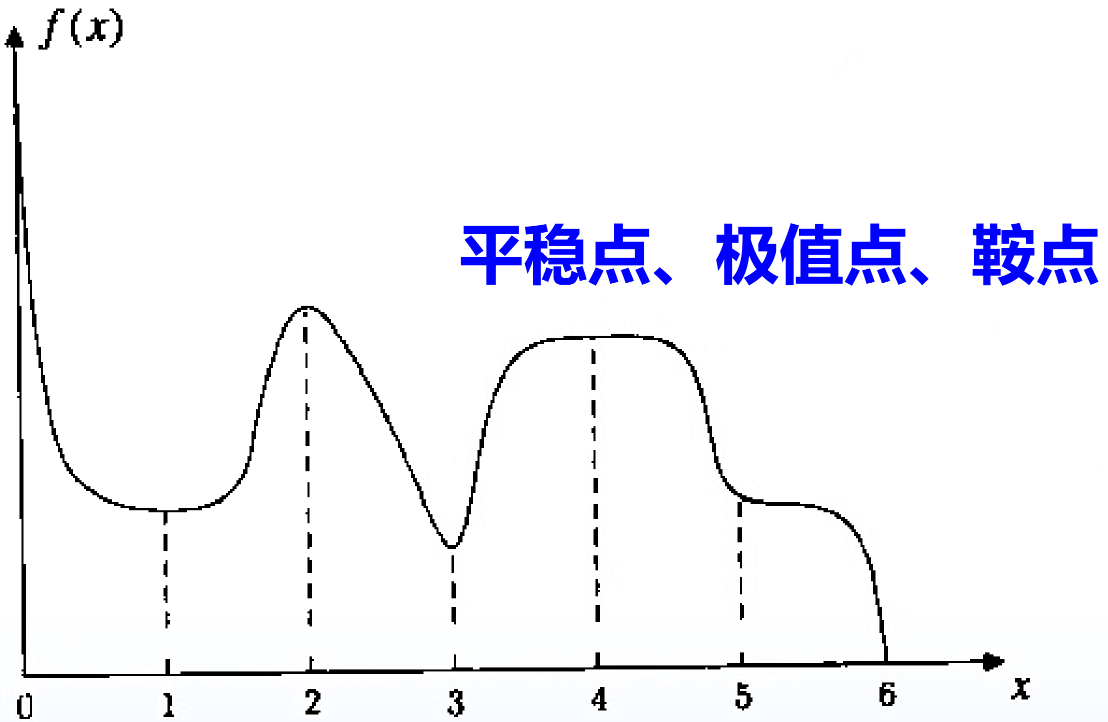


图 4.1.1 单变量函数的平稳点与极值点

如何寻找极值点?

单变量函数

单变量函数的平稳点与极值点

平稳点： $f'(c)=0$

- 平稳点是函数图像上的一个点，在该点处，函数的导数为零。直观上，这意味着函数在这一点的切线是水平的。
- 平稳点是函数局部极值可能出现的地方，但不是所有平稳点都是极值点。

局部极小点: $f'(c)=0$ $\quad f''(c)=\left.\frac{\mathrm{d}^2 f(x)}{dx^2}\right|_{x=c}\geq 0$

局部极大点: $f'(c)=0$ $\quad f''(c)=\left.\frac{\mathrm{d}^2 f(x)}{dx^2}\right|_{x=c}\leq 0$

鞍点（saddle point）： $f'(c)=0$

$$f''(c+\Delta x)\leq 0 \qquad f''(c+\Delta x)\geq 0$$

多变量函数

多变量函数的平稳点与极值点

多变量函数无约束极小化问题

$$\min_{x \in S} f(\boldsymbol{x}) \quad f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$$

邻域 $B(\boldsymbol{c};r) = \left\{ \boldsymbol{x} \middle| \ \boldsymbol{x} \in \mathbb{R}^n, \| \boldsymbol{x} - \boldsymbol{c} \|_2 < r \right\}$

闭合邻域 $B(\boldsymbol{c};r) = \left\{ \boldsymbol{x} \middle| \ \boldsymbol{x} \in \mathbb{R}^n, \| \boldsymbol{x} - \boldsymbol{c} \|_2 \le r \right\}$

多变量函数

多变量函数的平稳点与极值点

二阶泰勒级数逼近

$$
\begin{aligned}
f(\boldsymbol{c}+\Delta\boldsymbol{x}) &= f(\boldsymbol{c})+\left(\frac{\partial f(\boldsymbol{c})}{\partial \boldsymbol{c}}\right)^{\mathrm{T}}\Delta\boldsymbol{x}+\frac{1}{2}(\Delta\boldsymbol{x})^{\mathrm{T}}\cdot\frac{\partial^2 f(\boldsymbol{c})}{\partial \boldsymbol{c}\partial \boldsymbol{c}^{\mathrm{T}}}\Delta\boldsymbol{x} \\
&= f(\boldsymbol{c})+(\nabla f(\boldsymbol{c}))^{\mathrm{T}}\Delta\boldsymbol{x}+\frac{1}{2}(\Delta\boldsymbol{x})^{\mathrm{T}}\boldsymbol{H}(f(\boldsymbol{c}))\Delta\boldsymbol{x}
\end{aligned}
$$

梯度向量

$$
\nabla f(\boldsymbol{c}) = \frac{\partial f(\boldsymbol{c})}{\partial \boldsymbol{c}} = \left.\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{c}}
$$

Hessian矩阵

$$
\boldsymbol{H}(f(c)) = \frac{\partial^2 f(\boldsymbol{c})}{\partial \boldsymbol{c}\partial \boldsymbol{c}^{\mathrm{T}}} = \left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right|_{\boldsymbol{x}=\boldsymbol{c}}
$$

多变量函数

多变量函数的平稳点与极值点

$$f(c)\leqslant f(c+\Delta x)\quad \forall\, 0<\|\Delta x\|_2\leqslant r$$

局部极小

$$\boldsymbol{H}(f(\boldsymbol{c}))=\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right|_{\boldsymbol{x}=\boldsymbol{c}}\succeq 0$$

局部极小

$$f(c)< f(c+\Delta x)\quad \forall\, 0<\|\Delta x\|_2\leqslant r$$

严格局部极小

$$f(\boldsymbol{c})\leqslant f(\boldsymbol{x})\quad \forall\,\boxed{\boldsymbol{x}\in S}$$

全局极小

$$f(\boldsymbol{c})< f(\boldsymbol{x})\quad \forall \boldsymbol{x}\in S,\boldsymbol{x}\neq \boldsymbol{c}$$

严格全局极小

多变量函数

多变量函数 $f(\boldsymbol{X})$ 的平稳点与极值点

邻域

$$B(\boldsymbol{C};r)=\left\{\boldsymbol{X}\,\middle|\,\boldsymbol{X}\in\mathbb{R}^{m\times n},\|\mathrm{vec}(\boldsymbol{X})-\mathrm{vec}(\boldsymbol{C})\|_2<r\right\}$$

二阶泰勒级数逼近

$$f(\boldsymbol{C}+\Delta\boldsymbol{X})=f(\boldsymbol{C})+\left(\frac{\partial f(\boldsymbol{C})}{\partial\,\mathrm{vec}(\boldsymbol{C})}\right)^{\mathrm{T}}\mathrm{vec}(\Delta\boldsymbol{X})+\frac{1}{2}(\mathrm{vec}(\Delta\boldsymbol{X}))^{\mathrm{T}}\frac{\partial^2 f(\boldsymbol{C})}{\partial\,\mathrm{vec}(\boldsymbol{C})\partial(\mathrm{vec}\,\boldsymbol{C})^{\mathrm{T}}}\mathrm{vec}(\Delta\boldsymbol{X})$$

$$=f(\boldsymbol{C})+\left(\nabla_{\mathrm{vec}\boldsymbol{C}}\,f(\boldsymbol{C})\right)^{\mathrm{T}}\mathrm{vec}(\Delta\boldsymbol{X})+\frac{1}{2}(\mathrm{vec}(\Delta\boldsymbol{X}))^{\mathrm{T}}\boldsymbol{H}(f(\boldsymbol{C}))\,\mathrm{vec}(\Delta\boldsymbol{X})$$

$$\nabla_{\mathrm{vec}\boldsymbol{C}}\,f(\boldsymbol{C})=\left.\frac{\partial f(\boldsymbol{X})}{\partial\,\mathrm{vec}(\boldsymbol{X})}\right|_{\boldsymbol{X}=\boldsymbol{C}}\in\mathbb{R}^{mn}\qquad \boldsymbol{H}(f(\boldsymbol{C}))=\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial\,\mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right|_{\boldsymbol{X}=\boldsymbol{C}}\in\mathbb{R}^{mn\times mn}$$

总结

表 4.1.1　实变函数的平稳点和极值点的条件

实变函数	$f(x):\mathbb{R}\to\mathbb{R}$	$f(\boldsymbol{x}):\mathbb{R}^n\to\mathbb{R}$	$f(\boldsymbol{X}):\mathbb{R}^{m\times n}\to\mathbb{R}$
平稳点	$\left.\frac{\partial f(x)}{\partial x}\right\|_{x=c}=0$	$\left.\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}=\boldsymbol{0}$	$\left.\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}=\boldsymbol{O}_{m\times n}$
局部极小点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}\geqslant 0$	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}\succeq 0$	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}\succeq 0$
严格局部极小点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}> 0$	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}\succ 0$	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}\succ 0$
局部极大点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}\leqslant 0$	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}\preceq 0$	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}\preceq 0$
严格局部极大点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}< 0$	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}\prec 0$	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}\prec 0$
鞍　点	$\left.\frac{\partial^2 f(x)}{\partial x\partial x}\right\|_{x=c}$ 不定	$\left.\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}}\right\|_{\boldsymbol{x}=\boldsymbol{c}}$ 不定	$\left.\frac{\partial^2 f(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}\right\|_{\boldsymbol{X}=\boldsymbol{C}}$ 不定

4.1.4 一阶必要条件，二阶必要条件，二阶充分条件

第三章习题

见学在浙大
作业版块
Homework2

10.13 交