M05M11084 最优化理论、算法与应用

2 理论基础

M05M11084

2 理论基础

参考：

1. 矩阵分析与应用，张贤达，第3章
2. 最优化导论，Edwin K.P., Chong著，孙志强等译，第4~5章
3. Convex Optimization, Stephen Boyd, Chapter 2&3

1.导数定义及运算法则

2. 梯度信息

3. 泰勒展开与函数逼近

4. 极值点与平稳点

5. 凸集、凸函数与凸优化

6.优化问题的最优性条件

7. 下降方向的判定

1.导数定义及运算法则

- 1.1 函数与变量定义的多种形式
- 1.2 实函数（列向量为变量）
- 1.3 向量函数（列向量为变量）
- 1.4 导数运算法则
  常用函数导数公式
- 1.5 二阶导数矩阵
- 1.6 微分与导数的关系

2. 梯度信息

3. 泰勒展开与函数逼近

4. 极值点与平稳点

5. 凸集、凸函数与凸优化

6. 无约束问题的最优性条件

7. 下降方向

参考:

2. 矩阵分析与应用，张贤达，第3章

1.1 函数与变量定义的多种形式

	实函数 $\to \mathcal{R}$	向量函数 $\to \mathcal{R}^p$	矩阵函数 $\to \mathcal{R}^{p\times q}$
实数变量 $\mathcal{R}$	$f(x)$	$\boldsymbol{f}(x)$	$F(x)$
列向量变量 $\mathcal{R}^n$	$f(\boldsymbol{x})$	$\boldsymbol{f}(\boldsymbol{x})$	$F(\boldsymbol{x})$
矩阵变量 $\mathcal{R}^{m\times n}$	$f(X)$	$\boldsymbol{f}(X)$	$F(X)$

主要涉及

少量涉及

$f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x}$

$\boldsymbol{f}(\boldsymbol{x}) = A\boldsymbol{x}$

当上下文无歧义时
也简单写成

$f(x) = a^T x$

$f(x) = Ax$

注：无论是函数还是变量，对于向量和矩阵 “转置” 还有其它形式，本课程不讨论

1.2 实函数（列向量为变量）

设可微函数$f:\mathcal{R}^n \to \mathcal{R}, \boldsymbol{x} \in \mathcal{R}^n$，即，$n$元实函数$f(\boldsymbol{x})$, 自变量$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$

一阶导数向量 $Df(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$ 行向量

梯度向量 $\nabla f(\boldsymbol{x}) = \left(Df(\boldsymbol{x})\right)^T = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$ 列向量

例 $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x}$，$\boldsymbol{x} \in \mathcal{R}^n$，给定$\boldsymbol{a} \in \mathcal{R}^n$ $\quad \boldsymbol{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

$$Df(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\sum_{j=1}^{n} a_j x_j\right) = a_i$$

$$= [a_1 \ \ a_2 \ \ \cdots \ \ a_n]$$

$$= \boldsymbol{a}^T$$

1.3 向量函数（列向量为变量）

$\boldsymbol{f}$: $\mathcal{R}^n \to \mathcal{R}^m$，$\boldsymbol{x} \in \mathcal{R}^n$ $\qquad \boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$

一阶导数矩阵 　　雅克比矩阵
Jacobi

$$D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \boxed{Df_1(\boldsymbol{x})} \\ Df_2(\boldsymbol{x}) \\ \vdots \\ Df_m(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}_{m \times n}$$

$$Df_i(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f_i(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_i(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_i(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

例 $\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{x}$, $\boldsymbol{x}\in\mathcal{R}^n$

$$\frac{\partial x_i}{\partial x_j}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$$

$$Dx_i=[\underbrace{0\ \ \cdots\ \ 0}_{i-1}\ \ \underbrace{1}_{\text{第}i\text{个}}\ \ \underbrace{0\ \ \cdots\ \ 0}_{n-i}]$$

$$D=\begin{bmatrix}\frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \cdots & \frac{\partial}{\partial x_n}\end{bmatrix}$$

$$D\boldsymbol{f}(\boldsymbol{x})=D\boldsymbol{x}=D\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\begin{bmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&1\end{bmatrix}=I_n$$

例 $\boldsymbol{f}:\mathcal{R}^n\to\mathcal{R}^m$, 函数 $\boldsymbol{f}(\boldsymbol{x})=A\boldsymbol{x}$, 变量$\boldsymbol{x}\in\mathcal{R}^n$, 给定$A\in\mathcal{R}^{m\times n}$

$$\boldsymbol{f}(\boldsymbol{x})=A\boldsymbol{x}=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\begin{bmatrix}\boldsymbol{a}_1^T\boldsymbol{x}\\\boldsymbol{a}_2^T\boldsymbol{x}\\\vdots\\\boldsymbol{a}_m^T\boldsymbol{x}\end{bmatrix}$$

$$\boldsymbol{a}_i^T=[a_{i1}\ \ a_{i2}\ \ \cdots\ \ a_{in}],\ \ i=1,2,\cdots,m$$

$$D\left(\boldsymbol{a}_i^T\boldsymbol{x}\right)=\boldsymbol{a}_i^T$$

$$D\boldsymbol{f}(\boldsymbol{x})=\begin{bmatrix}D(\boldsymbol{a}_1^T\boldsymbol{x})\\D(\boldsymbol{a}_2^T\boldsymbol{x})\\\vdots\\D(\boldsymbol{a}_m^T\boldsymbol{x})\end{bmatrix}=\begin{bmatrix}\boldsymbol{a}_1^T\\\boldsymbol{a}_2^T\\\vdots\\\boldsymbol{a}_m^T\end{bmatrix}=A$$

一阶导数矩阵的几何解释

$\boldsymbol{f}: \mathcal{R}^n \to \mathcal{R}^m$，点 $\boldsymbol{x}_0 = \left[x_1^0, \dots, x_{j-1}^0, x_j^0, x_{j+1}^0, \dots, x_n^0\right]^T$ 的导数矩阵为 $\mathrm{D}\,\boldsymbol{f}(\boldsymbol{x}_0)$，$J = \{1, \dots, n\}$

$\mathrm{D}\,\boldsymbol{f}(\boldsymbol{x}_0)$的列为向量偏导数 $\dfrac{\partial \boldsymbol{f}}{\partial x_j}(\boldsymbol{x}_0)$

设仅第j个元素为变量的向量为

$\boldsymbol{x}_j^0 = \left[x_1^0, \dots, x_{j-1}^0, x_j, x_{j+1}^0, \dots, x_n^0\right]^T, j \in J$

第j个元素x_j是变量

其余元素x_i^0为常数，$\forall i \neq j, i \in J$

$\boldsymbol{x}_1^0 = [x, b]^T$，或，$\boldsymbol{x}_2^0 = [a, y]^T$

$\boldsymbol{f}\left(\boldsymbol{x}_j^0\right)$ 只有第j个元素x_j为变量的单变量函数，为曲线

$f(x, b)$，或，$f(a, y)$

曲线$\boldsymbol{f}\left(\boldsymbol{x}_j^0\right)$在点$\boldsymbol{x}_0$的切向量就是$\dfrac{\partial \boldsymbol{f}}{\partial x_j}(\boldsymbol{x}_0)$

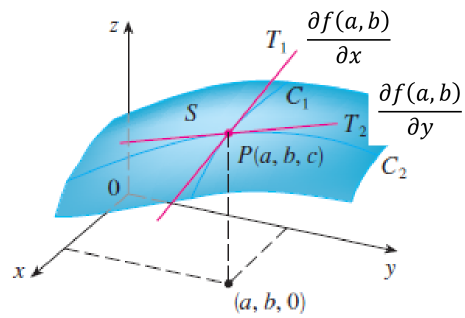


FIGURE 1
The partial derivatives of f at (a, b) are the slopes of the tangents to C_1 and C_2.

1.4 导数运算法则

设 函数$\boldsymbol{f}, \boldsymbol{g}$可微，$\alpha, \beta$为常数

① $D\boldsymbol{c} = \boldsymbol{0}$

② $D(\alpha \boldsymbol{f} + \beta \boldsymbol{g}) = \alpha D\boldsymbol{f} + \beta D\boldsymbol{g}$

③ $D(\boldsymbol{f}^T) = (D\boldsymbol{f})^T$

④ 复合函数的链式法则

⑤ 乘积法则

详见后

④链式法则

链式法则1	复合函数 $h:\mathcal{R}\to\mathcal{R}$，$h=f\circ g$，$h(t)=f\big(\boldsymbol{g}(t)\big)$ $h'(t)=Df(\boldsymbol{x})D\boldsymbol{g}(t)=\nabla f(\boldsymbol{x})^T\begin{bmatrix} g_1'(t) \\ \vdots \\ g_n'(t)\end{bmatrix}$	其中，$\boldsymbol{g}:\mathcal{R}\to\mathcal{R}^n$，$f:\mathcal{R}^n\to\mathcal{R}$ 记 $\boldsymbol{x}=\boldsymbol{g}(t)$
链式法则2	复合函数 $h:\mathcal{R}^n\to\mathcal{R}$，$h=f\circ\boldsymbol{g}$，$h(\boldsymbol{u})=f\big(\boldsymbol{g}(\boldsymbol{u})\big)$ $Dh(\boldsymbol{u})=Df(\boldsymbol{x})D\boldsymbol{g}(\boldsymbol{u})$ $D_{\boldsymbol{u}}h(\boldsymbol{u})=D_{\boldsymbol{x}}f(\boldsymbol{x})D_{\boldsymbol{u}}\boldsymbol{g}(\boldsymbol{u})$	其中，$\boldsymbol{g}:\mathcal{R}^n\to\mathcal{R}^m$，$f:\mathcal{R}^m\to\mathcal{R}$ 记 $\boldsymbol{x}=\boldsymbol{g}(\boldsymbol{u})$
链式法则3	复合函数 $\boldsymbol{h}:\mathcal{R}^n\to\mathcal{R}^m$，$\boldsymbol{h}=\boldsymbol{f}\circ\boldsymbol{g}$，$\boldsymbol{h}(\boldsymbol{u})=\boldsymbol{f}\big(\boldsymbol{g}(\boldsymbol{u})\big)$ $D\boldsymbol{h}(\boldsymbol{u})=D\boldsymbol{f}(\boldsymbol{x})D\boldsymbol{g}(\boldsymbol{u})$ $D_{\boldsymbol{u}}\boldsymbol{h}(\boldsymbol{u})=D_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x})D_{\boldsymbol{u}}\boldsymbol{g}(\boldsymbol{u})$	其中$\boldsymbol{g}:\mathcal{R}^n\to\mathcal{R}^p$，$\boldsymbol{f}:\mathcal{R}^p\to\mathcal{R}^m$ 记 $\boldsymbol{x}=\boldsymbol{g}(\boldsymbol{u})$

例 考虑函数 $f:\mathcal{R}^n\to\mathcal{R}$，$f\in C^2$　　$\boldsymbol{x}^*,\boldsymbol{d}\in\mathcal{R}^n$,

$$\boldsymbol{x}(\alpha)=\boldsymbol{x}^*+\alpha\boldsymbol{d},\quad \alpha\in\mathcal{R}$$

$$\phi(\alpha)\triangleq f\big(\boldsymbol{x}(\alpha)\big),\quad \phi:\mathcal{R}\to\mathcal{R}$$

求$\phi(0)$，$\phi'(0)$

$\boldsymbol{x}(0)=\boldsymbol{x}^*$

$\phi(0)=f(\boldsymbol{x}^*)$

$D\boldsymbol{x}(\alpha)=\boldsymbol{d}$

$\phi'(\alpha)=Df\big(\boldsymbol{x}(\alpha)\big)D\boldsymbol{x}(\alpha)=\nabla f\big(\boldsymbol{x}(\alpha)\big)^T\boldsymbol{d}=\boldsymbol{d}^T\nabla f\big(\boldsymbol{x}(\alpha)\big)$

$\phi'(0)=\boldsymbol{d}^T\nabla f(\boldsymbol{x}^*)$

④链式法则

Exercise 5.5

Consider $f(\boldsymbol{x}) = x_1x_2/2$, $\boldsymbol{g}(s,t) = [4s+3t \quad 2s+t]^T$.

Evaluate $\frac{\partial}{\partial s} f\big(\boldsymbol{g}(s,t)\big)$ and $\frac{\partial}{\partial t} f\big(\boldsymbol{g}(s,t)\big)$ using the chain rule.

$$f(\boldsymbol{x}) = x_1x_2/2$$

$$\boldsymbol{g}(s,t) \triangleq \begin{bmatrix} x_1(s,t) \\ x_2(s,t) \end{bmatrix} = \begin{bmatrix} 4s+3t \\ 2s+t \end{bmatrix}$$

$$h(s,t) \triangleq f\big(\boldsymbol{g}(s,t)\big)$$

$$D_{s,t}h(s,t) = D_{\boldsymbol{x}} f(\boldsymbol{x}) D_{s,t}\boldsymbol{g}(s,t)$$

$$D_{s,t}h(s,t) = D_{\boldsymbol{x}} f(\boldsymbol{x}) D_{s,t}\boldsymbol{g}(s,t)$$

$$= \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial s} & \frac{\partial x_2}{\partial t} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{x_2}{2} & \frac{x_1}{2} \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} x_1 + 2x_2 & \frac{x_1 + 3x_2}{2} \end{bmatrix}$$

⑤乘积法则

$$\boldsymbol{f}: \mathcal{R}^n \to \mathcal{R}^m, \quad \boldsymbol{g}: \mathcal{R}^n \to \mathcal{R}^m$$

$$h: \mathcal{R}^n \to \mathcal{R} \qquad h(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{g}(\boldsymbol{x})$$

$$Dh(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T D\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})^T D\boldsymbol{f}(\boldsymbol{x})$$

⑤乘积法则

$$h(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{g}(\boldsymbol{x})$$
$$Dh(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T D\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})^T D\boldsymbol{f}(\boldsymbol{x})$$

例 1. $h(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$, $A \in \mathcal{R}^{n\times n}, \boldsymbol{x} \in \mathcal{R}^n$

$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$, $\boldsymbol{g}(\boldsymbol{x}) = A\boldsymbol{x}$

$$D(\boldsymbol{x}^T A\boldsymbol{x}) = D\big((\boldsymbol{x})^T (A\boldsymbol{x})\big) = \boldsymbol{x}^T D(A\boldsymbol{x}) + (A\boldsymbol{x})^T D(\boldsymbol{x}) = \boldsymbol{x}^T (A + A^T)$$

若$A^T = A$，则$D(\boldsymbol{x}^T A\boldsymbol{x}) = 2\boldsymbol{x}^T A$

例 2. $h(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x}$, $\boldsymbol{x} \in \mathcal{R}^n$

$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$, $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{x}$

$$D(\boldsymbol{x}^T \boldsymbol{x}) = \boldsymbol{x}^T D(\boldsymbol{x}) + (\boldsymbol{x})^T D(\boldsymbol{x}) = 2\boldsymbol{x}^T$$

常用函数导数公式

1. 若给定 $A \in \mathcal{R}^{m\times n}, \boldsymbol{b} \in \mathcal{R}^m$ ；变量 $\boldsymbol{x} \in \mathcal{R}^n$,

$$D(\boldsymbol{b}^T A\boldsymbol{x}) = \boldsymbol{b}^T A,$$
$$D(\boldsymbol{x}^T A\boldsymbol{x}) = \boldsymbol{x}^T (A + A^T) \text{ if } m = n$$
$$D(A\boldsymbol{x}) = A$$
$$D(\boldsymbol{x}) = I$$

2. 若给定 $\boldsymbol{b} \in \mathcal{R}^n$, $D\big(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}\big) = D\big(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}\big) = \boldsymbol{b}^{\mathrm{T}}$

3. 若给定 $Q \in \mathcal{R}^{n\times n}$, $Q = Q^T$ ，那么 $D(\boldsymbol{x}^T Q\boldsymbol{x}) = 2\boldsymbol{x}^T Q$

$$D(\boldsymbol{x}^T \boldsymbol{x}) = 2\boldsymbol{x}^T$$

$$\boxed{\|\boldsymbol{x}\|^2 = \boldsymbol{x}^T \boldsymbol{x}}$$

1.5 二阶导数矩阵

$f:\mathcal{R}^n \to \mathcal{R}, \boldsymbol{x} \in \mathcal{R}^n$，二阶可微

二阶导数矩阵 $D^2 f(\boldsymbol{x}) \triangleq D(Df(\boldsymbol{x})^T)$

$\triangleq D\big(\nabla f(\boldsymbol{x})\big)$

$$Df(\boldsymbol{x})^T = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

$$D^2 f(\boldsymbol{x}) = D\left(\begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}\right) = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}\right) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}\right) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_n}\right) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_n}\right) \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}$$

向量函数的二阶导数涉及矩阵函数求导，暂时不讲；
若后面需要，再具体讲

例 考虑函数 $f:\mathcal{R}^n \to \mathcal{R}, f \in C^2 \quad \boldsymbol{x}^*, \boldsymbol{d} \in \mathcal{R}^n,$

$$\boldsymbol{x}(\alpha) = \boldsymbol{x}^* + \alpha \boldsymbol{d}, \quad \alpha \in \mathcal{R}$$

$$\phi(\alpha) \triangleq f\big(\boldsymbol{x}(\alpha)\big), \quad \phi:\mathcal{R} \to \mathcal{R}$$

求$\phi(0)$, $\phi'(0)$和$\phi''(0)$

$$f \in C^2 \Longrightarrow \Big(D^2 f\big(\boldsymbol{x}(\alpha)\big)\Big)^T = D^2 f\big(\boldsymbol{x}(\alpha)\big)$$

$$\nabla^2 f\big(\boldsymbol{x}(\alpha)\big) = D^2 f\big(\boldsymbol{x}(\alpha)\big)$$

$\boldsymbol{x}(0) = \boldsymbol{x}^*$

$\phi(0) = f(\boldsymbol{x}^*)$

$D\boldsymbol{x}(\alpha) = \boldsymbol{d}$

$\phi'(\alpha) = \boldsymbol{d}^T \nabla f\big(\boldsymbol{x}(\alpha)\big)$

$$\phi''(\alpha) = D\big[\boldsymbol{d}^T \nabla f\big(\boldsymbol{x}(\alpha)\big)\big] = \boldsymbol{d}^T D\big[\nabla f\big(\boldsymbol{x}(\alpha)\big)\big]$$

$$= \boldsymbol{d}^T D_{\boldsymbol{x}}\big[\nabla f\big(\boldsymbol{x}(\alpha)\big)\big][D_\alpha \boldsymbol{x}(\alpha)]$$

$$= \boldsymbol{d}^T \big[D^2 f\big(\boldsymbol{x}(\alpha)\big)\big]\boldsymbol{d}$$

$\phi'(0) = \boldsymbol{d}^T \nabla f(\boldsymbol{x}^*)$

$\phi''(0) = \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d}$

例 $h(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$, $A \in \mathcal{R}^{n\times n}$, $\boldsymbol{x} \in \mathcal{R}^n$

$$D(\boldsymbol{x}^T A\boldsymbol{x}) = D((\boldsymbol{x}^T)(A\boldsymbol{x}))$$
$$= \boldsymbol{x}^T(A + A^T)$$

$$D^2 h(\boldsymbol{x}) = D(Dh(\boldsymbol{x})^T)$$
$$= D\left(\left(\boldsymbol{x}^T(A + A^T)\right)^T\right)$$
$$= D\left((A + A^T)^T \boldsymbol{x}\right)$$
$$= (A + A^T)^T$$
$$= A + A^T$$
$$= 2A$$

$A^T = A$

1.6 微分与导数的关系

可微的实值函数 $f: \mathcal{R}^n \to \mathcal{R}$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \rightarrow d\boldsymbol{x} = \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}$$

$$df(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})^T}{\partial \boldsymbol{x}} d\boldsymbol{x}$$
$$= \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x})}{\partial x_i} dx_i$$
$$= Df(\boldsymbol{x})d\boldsymbol{x}$$
$$= \left(\nabla f(\boldsymbol{x})\right)^T d\boldsymbol{x}$$

如果表达式 $df(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x})d\boldsymbol{x}$ 成立，那么，导数 $Df(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x})$

常用于求导

1.6 微分与导数的关系

可微向量函数 $\boldsymbol{f}:\mathcal{R}^n \to \mathcal{R}^m$

$$df_j(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{\partial f_j(\boldsymbol{x})}{\partial x_i} dx_i$$

$$\begin{aligned} d\boldsymbol{f}(\boldsymbol{x}) &= \frac{\partial \boldsymbol{f}(\boldsymbol{x})^T}{\partial \boldsymbol{x}} d\boldsymbol{x} \\ &= D\boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} \\ &= \left(\nabla \boldsymbol{f}(\boldsymbol{x})\right)^T d\boldsymbol{x} \end{aligned}$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \to d\boldsymbol{x} = \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}$$

$$\boldsymbol{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} \to d\boldsymbol{f} = \begin{bmatrix} df_1 \\ df_2 \\ \vdots \\ df_m \end{bmatrix}$$

1.导数定义及运算法则

2. 梯度信息
   - 2.1 梯度向量
     常用函数梯度公式
   - 2.2 Hesse 矩阵
   - 2.3 矩阵的正定/半正定性质与判定方法

3. 泰勒展开与函数逼近

4. 极值点与平稳点

5. 凸集、凸函数与凸优化

6. 无约束问题的最优性条件

7. 下降方向的判定

2.1 梯度向量

用于构造搜索方向

可微函数 $f: \mathcal{R}^n \to \mathcal{R} \quad \boldsymbol{x} \in \mathcal{R}^n$

$f(\boldsymbol{x})$在$\boldsymbol{x}$处的梯度向量

$$\nabla f(\boldsymbol{x}) \triangleq \left(Df(\boldsymbol{x})\right)^T = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

$n = 1$时， $\nabla f(x) = f'(x)$

常用函数梯度公式

1. 若给定 $A \in \mathcal{R}^{m \times n}, \boldsymbol{b} \in \mathcal{R}^m$ ； 变量 $\boldsymbol{x} \in \mathcal{R}^n$

   $\nabla(\boldsymbol{b}^T A\boldsymbol{x}) = A^T\boldsymbol{b},$

   $\nabla(A\boldsymbol{x}) = A^T$

   $\nabla(\boldsymbol{x}^T A\boldsymbol{x}) = (A + A^T)\boldsymbol{x} \ \text{ if } m = n$

2. 若给定 $\boldsymbol{b} \in \mathcal{R}^n$， $\nabla\left(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}\right) = \nabla\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}\right) = \boldsymbol{b}$

3. 若给定 $Q \in \mathcal{R}^{n \times n}$， $Q = Q^T$ ，那么$\nabla(\boldsymbol{x}^T Q\boldsymbol{x}) = 2Q\boldsymbol{x}$

   $\nabla(\boldsymbol{x}^T\boldsymbol{x}) = 2\boldsymbol{x}$

4. $\nabla(\|\boldsymbol{x}\|^2) = \nabla(\boldsymbol{x}^T\boldsymbol{x}) = 2\boldsymbol{x}$

2.2 Hesse 矩阵

二阶可微函数 $f: \mathcal{R}^n \to \mathcal{R}$ $\boldsymbol{x} \in \mathcal{R}^n$

$$\nabla^2 f(\boldsymbol{x}) \triangleq \left(D^2 f(\boldsymbol{x})\right)^T = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}$$

$f(\boldsymbol{x})$在$\boldsymbol{x}$处的Hesse 矩阵

记为H，或F

$n = 1$时，$\nabla^2 f(x) = f''(x)$

$\boxed{D^2 f(\boldsymbol{x}) \triangleq D\left(\nabla f(\boldsymbol{x})\right)}$

克莱罗定理或施瓦茨定理 *Clairaut's theorem* or *Schwarz's theorem*

如果$f: \mathcal{R}^n \to \mathcal{R}$在点$\boldsymbol{x}$处是二次连续可微的，那么，点$\boldsymbol{x}$处的Hesse矩阵是对称的

如果f的二次偏导数不是连续的，那么，Hesse矩阵不一定对称

Example Consider the function $f(\boldsymbol{x}) = \begin{cases} \frac{x_1 x_2 (x_1^2 - x_2^2)}{x_1^2 + x_2^2} & \text{if } \boldsymbol{x} \neq \boldsymbol{0}, \\ 0 & \text{if } \boldsymbol{x} = \boldsymbol{0} \end{cases}$

compute its Hessian at the point $\boldsymbol{0} = [0,0]^T$.

2.2 Hesse 矩阵

$$f(\boldsymbol{x})=\begin{cases}\frac{x_1x_2(x_1^2-x_2^2)}{x_1^2+x_2^2} & \text{if } \boldsymbol{x}\neq\mathbf{0},\\ 0 & \text{if } \boldsymbol{x}=\mathbf{0}\end{cases}$$

$$\frac{\partial f}{\partial x_1}(\boldsymbol{x})=\begin{cases}\dfrac{x_2(x_1^4-x_2^4+4x_1^2x_2^2)}{(x_1^2+x_2^2)^2} & \text{if } \boldsymbol{x}\neq\mathbf{0}\\ 0 & \text{if } \boldsymbol{x}=\mathbf{0}\end{cases}$$

$$\frac{\partial f}{\partial x_2}(\boldsymbol{x})=\begin{cases}\dfrac{x_1(x_1^4-x_2^4+4x_1^2x_2^2)}{(x_1^2+x_2^2)^2} & \text{if } \boldsymbol{x}\neq\mathbf{0}\\ 0 & \text{if } \boldsymbol{x}=\mathbf{0}\end{cases}$$

$$\boxed{\frac{\partial^2 f}{\partial x_1^2}=\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right)} \qquad \boxed{\frac{\partial^2 f}{\partial x_2(\partial x_1)}=\frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)}$$

$$\frac{\partial f}{\partial x_1}([x_1,0]^{\mathrm{T}})=0 \qquad \frac{\partial f}{\partial x_1}([0,x_2]^{\mathrm{T}})=-x_2$$

$$\frac{\partial^2 f}{\partial x_1^2}(\mathbf{0})=0 \qquad \frac{\partial^2 f}{\partial x_2\partial x_1}(\mathbf{0})=-1$$

$$\boxed{\frac{\partial^2 f}{\partial x_2^2}=\frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right)} \qquad \boxed{\frac{\partial^2 f}{\partial x_1(\partial x_2)}=\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)}$$

$$\frac{\partial f}{\partial x_2}([0,x_2]^{\mathrm{T}})=0 \qquad \frac{\partial f}{\partial x_2}([x_1,0\,]^{\mathrm{T}})=x_1$$

$$\frac{\partial^2 f}{\partial x_2^2}(\mathbf{0})=0 \qquad \frac{\partial^2 f}{\partial x_1\partial x_2}(\mathbf{0})=1$$

Therefore, the Hessian evaluated at the point **0** is

$$F(0)=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$$

which is *not symmetric*

2.2 Hesse 矩阵

例 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TQ\boldsymbol{x}-\boldsymbol{b}^T\boldsymbol{x}$ $\qquad Q^T=Q$

求：1. $\nabla f(\boldsymbol{x})$

2. $\nabla^2 f(\boldsymbol{x})$

$$\begin{aligned}Df(\boldsymbol{x})&=D\left(\frac{1}{2}\boldsymbol{x}^TQ\boldsymbol{x}-\boldsymbol{b}^T\boldsymbol{x}\right)\\&=D\left(\frac{1}{2}\boldsymbol{x}^TQ\boldsymbol{x}\right)-D(\boldsymbol{b}^T\boldsymbol{x})\\&=\boldsymbol{x}^TQ-\boldsymbol{b}^T\end{aligned}$$

$$\begin{aligned}\nabla f(\boldsymbol{x})&=\big(Df(\boldsymbol{x})\big)^T\\&=(\boldsymbol{x}^TQ-\boldsymbol{b}^T)^T\\&=Q\boldsymbol{x}-\boldsymbol{b}\end{aligned}$$

$$\begin{aligned}D^2f(\boldsymbol{x})&\triangleq D\big(\nabla f(\boldsymbol{x})\big)\\&=D(Q\boldsymbol{x}-\boldsymbol{b}\,)\\&=Q\end{aligned}$$

$$\begin{aligned}\nabla^2 f(\boldsymbol{x})&\triangleq\big(D^2f(\boldsymbol{x})\big)^T\\&=(Q)^T\\&=Q\end{aligned}$$

常用记号

$$\boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})$$

$$\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k) \longleftrightarrow \boldsymbol{g}^{(k)} = \nabla f\left(\boldsymbol{x}^{(k)}\right) \qquad g^k = \nabla f\left(x^k\right)$$

$$H(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x})$$

$$H_k = \nabla^2 f(\boldsymbol{x}_k) \longleftrightarrow H^{(k)} = \nabla^2 f\left(\boldsymbol{x}^{(k)}\right) \qquad H^k = \nabla^2 f\left(x^k\right)$$

练习

Consider the function $f(\boldsymbol{x}) = x_1^3 + x_2^2 + x_3^4 + 2x_1x_2$

a. Find $\nabla f(\boldsymbol{x})$.

b. Find the Hessian $F(\boldsymbol{x})$.

Solution

a. $\nabla f(\boldsymbol{x}) = \begin{bmatrix} 3x_1^2 + 2x_2 \\ 2x_1 + 2x_2 \\ 4x_3^3 \end{bmatrix}$

b. $F(\boldsymbol{x}) = \begin{bmatrix} 6x_1 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 12x_3^2 \end{bmatrix}$

练习

Consider the function $f(x) = ax^3 + b\sin x$

a. Find $\nabla f(\boldsymbol{x})$.

b. Find the Hessian $F(\boldsymbol{x})$.

Solution

$$f'(x) = 3ax^2 + b\cos x,$$

$$f''(x) = 6ax - b\sin x$$

练习

EXERCISES 5.3 Consider the function $f(\boldsymbol{x}) = (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x})(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x})$,

where ***a***, ***b***, and ***x*** are n-dimensional vectors.

a. Find $\nabla f(\boldsymbol{x})$.

b. Find the Hessian $F(\boldsymbol{x})$.

$$\begin{aligned} f(\boldsymbol{x}) &= (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x})(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}) \\ &= (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{a})(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}) \\ &= \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}})\boldsymbol{x} \end{aligned}$$

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= \left(\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}} + (\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}})^{T}\right)\boldsymbol{x} \\ &= (\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}} + \boldsymbol{b}\boldsymbol{a}^{\mathrm{T}})\boldsymbol{x} \end{aligned}$$

$$F(\boldsymbol{x}) = \boldsymbol{a}\boldsymbol{b}^{\mathrm{T}} + \boldsymbol{b}\boldsymbol{a}^{\mathrm{T}}$$

练习

Consider

$$f\left(x_1,x_2,x_3,x_4\right)=x_1^2+2x_2^2+3x_3^2+4x_4^2+5x_1x_2+6x_1x_3+7x_1x_4+8x_2x_3+9x_2x_4+10x_3x_4$$
$$+11x_1+12x_2+13x_3+14x_4+15$$

✓ How to change the General Form of a Quadratic Function into its Standard Form

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TQ\boldsymbol{x}-\boldsymbol{b}^T\boldsymbol{x}+c \qquad Q^T=Q$$

✓ The Derivative of a Quadratic Function

$$\nabla f(\boldsymbol{x})=Q\boldsymbol{x}-\boldsymbol{b}$$

$$\nabla^2 f(\boldsymbol{x})=Q$$

$$f\left(x_1,x_2,x_3,x_4\right)=x_1^2+2x_2^2+3x_3^2+4x_4^2+5x_1x_2+6x_1x_3+7x_1x_4+8x_2x_3+9x_2x_4+10x_3x_4+11x_1+12x_2+13x_3+14x_4+15$$

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} \qquad \boldsymbol{Q}=\begin{bmatrix}2&5&6&7\\5&4&8&9\\6&8&6&10\\7&9&10&8\end{bmatrix} \qquad \boldsymbol{b}=\begin{bmatrix}-11\\-12\\-13\\-14\end{bmatrix}$$

symmetric

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TQ\boldsymbol{x}-\boldsymbol{b}^T\boldsymbol{x}+15$$

2.3 矩阵的正定、半正定、负定、半负定、不定性质与判定

设矩阵$A \in \mathcal{R}^{n\times n}, A = A^T$，则

	$\forall\, \boldsymbol{d} \in \mathcal{R}^n$	总有	则称A为	记
①	$\boldsymbol{d} \neq \boldsymbol{0}$	$\boldsymbol{d}^T A\boldsymbol{d} > 0$	正定矩阵	$A \succ 0$
②		$\boldsymbol{d}^T A\boldsymbol{d} \geq 0$	半正定矩阵	$A \succcurlyeq 0$
③	$\boldsymbol{d} \neq \boldsymbol{0}$	$\boldsymbol{d}^T A\boldsymbol{d} < 0$	负定矩阵	$A \prec 0$
④		$\boldsymbol{d}^T A\boldsymbol{d} \leq 0$	半负定矩阵	$A \preccurlyeq 0$
⑤	其它		不定矩阵	

Hesse矩阵的正定、半正定、负定、半负定、不定性质与判定

$f: \mathcal{R}^n \to \mathcal{R}$，$\boldsymbol{x} \in \mathcal{R}^n$，$f \in C^2$，$H(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x})$

	$\forall\, \boldsymbol{d} \in \mathcal{R}^n$	总有	则称$H(\boldsymbol{x})$为	记
①	$\boldsymbol{d} \neq \boldsymbol{0}$	$\boldsymbol{d}^T H(\boldsymbol{x})\boldsymbol{d} > 0$	正定矩阵	$H(\boldsymbol{x}) \succ 0$
②		$\boldsymbol{d}^T H(\boldsymbol{x})\boldsymbol{d} \geq 0$	半正定矩阵	$H(\boldsymbol{x}) \succcurlyeq 0$
③	$\boldsymbol{d} \neq \boldsymbol{0}$	$\boldsymbol{d}^T H(\boldsymbol{x})\boldsymbol{d} < 0$	负定矩阵	$H(\boldsymbol{x}) \prec 0$
④		$\boldsymbol{d}^T H(\boldsymbol{x})\boldsymbol{d} \leq 0$	半负定矩阵	$H(\boldsymbol{x}) \preccurlyeq 0$
⑤	其它		不定矩阵	

$f: \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}$，若存在常数$c > 0, \forall\, \boldsymbol{d} \in \mathcal{R}^n, \boldsymbol{x} \in \mathcal{D}$, 有$\boldsymbol{d}^T H(\boldsymbol{x})\boldsymbol{d} \geq c\|\boldsymbol{d}\|^2$，称$H(\boldsymbol{x})$在$\mathcal{D}$上是一致正定的

如何判定n阶对称矩阵A为正（负）定矩阵?

方法一： 特征值

设λ_i，$i = 1,2,\dots,n$是矩阵A的特征值

① $\lambda_i > 0$，$i = 1,2,\dots,n$，则矩阵A是正定的，　$A \succ 0$

② $\lambda_i \geq 0$，$i = 1,2,\dots,n$，则矩阵A是半正定的，$A \succcurlyeq 0$

③ $\lambda_i < 0$，$i = 1,2,\dots,n$，则矩阵A是负定的，　$A \prec 0$

④ $\lambda_i \leq 0$，$i = 1,2,\dots,n$，则矩阵A是半负定的，$A \preccurlyeq 0$

⑤ 否则，矩阵A是不定的

如何判定n阶对称矩阵A为正（负）定矩阵?

方法二： 主子式　　$A^T = A$

① A各阶**顺序主子式**都大于0，则 $A \succ 0$

② A所有的**主子式**都是非负的，则 $A \succcurlyeq 0$

③ A各阶**顺序主子式**负正相间，则 $A \prec 0$

④ A的所有**奇数阶主子式**都非正、**偶数阶主子式**都非负，则 $A \preccurlyeq 0$

⑤ 否则，矩阵A是不定的

例 判定$f(x) = -x_1^2 + x_2^2 + x_3^2 + x_1x_2 + 2x_1 - 2x_2 + 1$的Hesse矩阵是否为正定矩阵?

$$H = \begin{bmatrix} -2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

顺序主子式: $\Delta_1 = -2 < 0$

$$\Delta_2 = \begin{vmatrix} -2 & 1 \\ 1 & 2 \end{vmatrix} = -5 < 0$$

$$\Delta_3 = \begin{vmatrix} -2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{vmatrix} = -10 < 0$$

H不定矩阵

例 判定$f(x) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + 2x_1x_3 + 2x_1 - 2x_2 + 1$的Hesse矩阵是否为正定矩阵?

$$H = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

一阶主子式: $\Delta_{11} = 2 > 0, \Delta_{22} = 2 > 0,\ \Delta_{33} = 2 > 0$

二阶主子式: $\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0,\ \begin{vmatrix} 2 & 2 \\ 2 & 2 \end{vmatrix} = 0,\ \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3 > 0$

三阶主子式: $\Delta_3 = \begin{vmatrix} 2 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 2 \end{vmatrix} = -2 < 0$

H不定矩阵

M05M11084 最优化理论、算法与应用

2 理论基础

一元实函数的Taylor展开式

设 $f: \mathcal{R} \to \mathcal{R}, x \in \mathcal{R}, f$在点$x$的某邻域内具有$K$阶连续导数（整数$K \geq 1$），则$f(x+\delta)$在点$x$处的$K$阶Taylor展开式

$$f(x+\delta) = f(x) + f'(x)\delta + \frac{f''(x)}{2!}\delta^2 + \cdots + \frac{f^{(K)}(x)}{K!}\delta^K + o(\delta^K)$$

式中，$o(\delta^K)$是关于δ^K的高阶无穷小

多元实函数的一阶与二阶Taylor展开式

设 $f: \mathcal{R}^n \to \mathcal{R}, \boldsymbol{x} \in \mathcal{R}^n, f$在点$\boldsymbol{x}$的某邻域内一阶连续可导，则$f(\boldsymbol{x}+\boldsymbol{\delta})$在点$\boldsymbol{x}$处的一阶Taylor展开式

$$f(\boldsymbol{x}+\boldsymbol{\delta}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|)$$

式中，$o(\|\boldsymbol{\delta}\|)$是关于$\|\boldsymbol{\delta}\|$的高阶无穷小

f在点$\boldsymbol{x}$的某邻域内二阶连续可导，则$f(\boldsymbol{x}+\boldsymbol{\delta})$在点$\boldsymbol{x}$处的二阶Taylor展开式

$$f(\boldsymbol{x}+\boldsymbol{\delta}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^T H(\boldsymbol{x})\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|^2)$$

式中，$o(\|\boldsymbol{\delta}\|^2)$是关于$\|\boldsymbol{\delta}\|^2$的高阶无穷小

2-范数 $\|\boldsymbol{x}\| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$

1.导数定义及运算法则

2. 梯度信息

3. 泰勒展开与函数逼近

4. 极值点与平稳点

5. 凸集、凸函数与凸优化

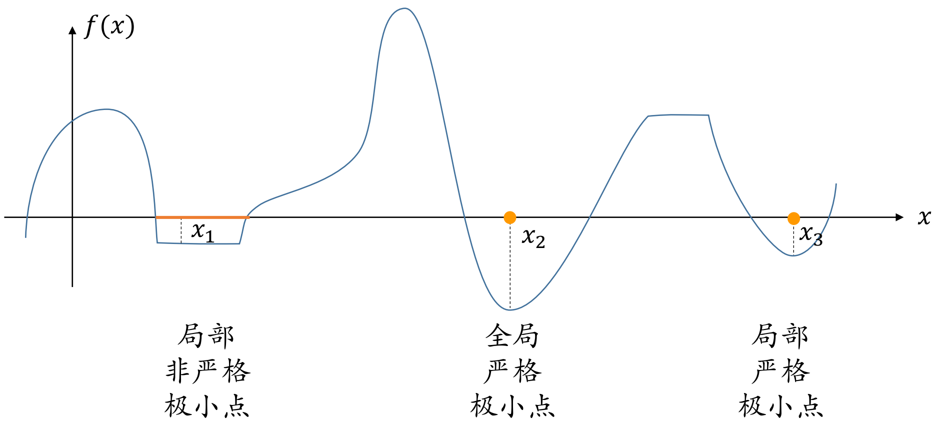
6.优化问题的最优性条件

7. 下降方向的判定

极值点

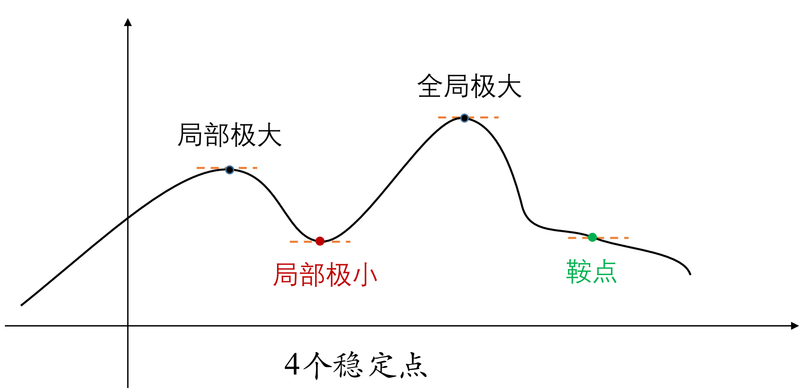
设$f: \mathcal{R}^n \to \mathcal{R},\ X \subseteq \mathcal{R}^n$有

<table>
<tr><th rowspan="2">情况</th><th colspan="3">条件</th><th colspan="2">称</th></tr>
<tr><th colspan="2">$\forall\ \boldsymbol{x} \in X$</th><th>$\exists \boldsymbol{x}^* \in X$</th><th>$\boldsymbol{x}^*$极小点</th><th>$f(\boldsymbol{x}^*)$极小值</th></tr>
<tr><td>①</td><td colspan="2"></td><td>$f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$</td><td>全局</td><td>全局</td></tr>
<tr><td>②</td><td colspan="2">$\boldsymbol{x} \ne \boldsymbol{x}^*$</td><td>$f(\boldsymbol{x}^*) < f(\boldsymbol{x})$</td><td>严格全局</td><td>严格全局</td></tr>
<tr><td>③</td><td rowspan="2">$\|\boldsymbol{x} - \boldsymbol{x}^*\| < \varepsilon$
$\varepsilon > 0$</td><td></td><td>$f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$</td><td>局部</td><td>局部</td></tr>
<tr><td>④</td><td>$\boldsymbol{x} \ne \boldsymbol{x}^*$</td><td>$f(\boldsymbol{x}^*) < f(\boldsymbol{x})$</td><td>严格局部</td><td>严格局部</td></tr>
</table>

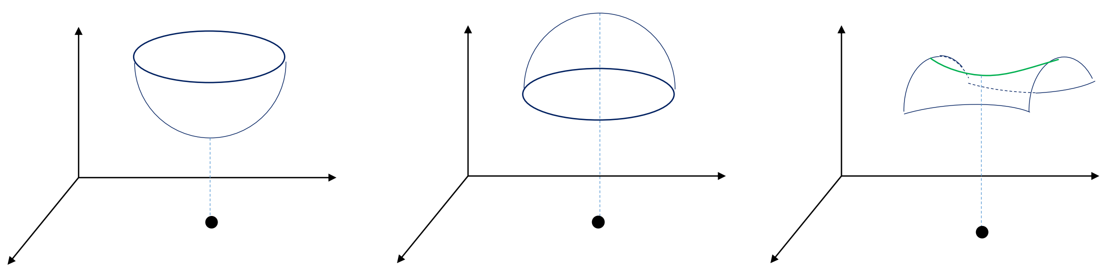


稳定点

设$f: \mathcal{R}^n \to \mathcal{R}$, $X \subseteq \mathcal{R}^n$，$\tilde{\boldsymbol{x}} \in X$，在点$\tilde{\boldsymbol{x}}$处梯度$\boldsymbol{g}(\tilde{\boldsymbol{x}})$存在，若$\boldsymbol{g}(\tilde{\boldsymbol{x}}) = \mathbf{0}$，则称$\tilde{\boldsymbol{x}}$是函数$f$的稳定点或驻点



4个稳定点



1.导数定义及运算法则

2. 梯度信息

3. 泰勒展开与函数逼近

4. 极值点与平稳点

5. 凸集、凸函数与凸优化

- 5.1 凸集
- 5.2 凸函数
- 5.3凸函数的判别定理
- 5.4 凸优化问题

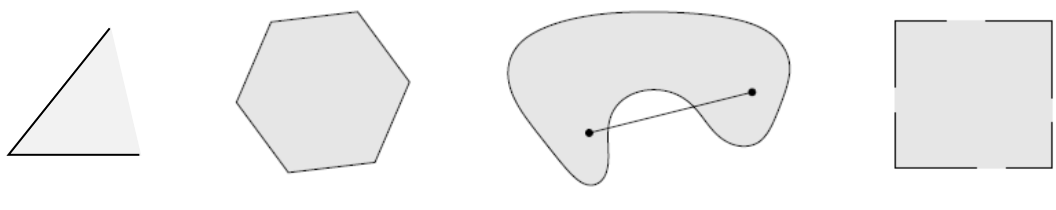
6.优化问题的最优性条件

7. 下降方向的判定

5.1 凸集

设集合$X \subset \mathcal{R}^n$. 若对任意的$\boldsymbol{x}, \boldsymbol{y} \in X$及任意的实数$\lambda \in [0,1]$, 都有$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in X$, 则集合$X$ 为凸集.

凸集的几何意义: 对非空集合 $X \subset \mathcal{R}^n$, 若连接其中任意两点的线段仍属于该集合, 则称该集合X为凸集.



Convex Sets

Nonconvex Sets

凸集的基本性质

设D, D_1, D_2是凸集, α 是一实数, 那么

(1) $\alpha D = \{\boldsymbol{y} | \boldsymbol{y} = \alpha \boldsymbol{x}, \boldsymbol{x} \in D\}$ 是凸集.

(2) 交集$D_1 \cap D_2$是凸集.

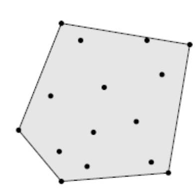
(3) 和集$D_1 + D_2 = \{\boldsymbol{z} | \boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} \in D_1, \boldsymbol{y} \in D_2\}$ 也是凸集.

(4) 差集$D_1 - D_2 = \{\boldsymbol{z} | \boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} \in D_1, \boldsymbol{y} \in D_2\}$ 也是凸集.

例 n 维欧氏空间中的m个点的凸组合是一个凸集. 即集合

$$\left\{\boldsymbol{x} = \sum_{i=1}^{m} \alpha_i \boldsymbol{x}_i \,\middle|\, \boldsymbol{x}_i \in \mathcal{R}^n, \alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i = 1\right\}$$

是凸集.

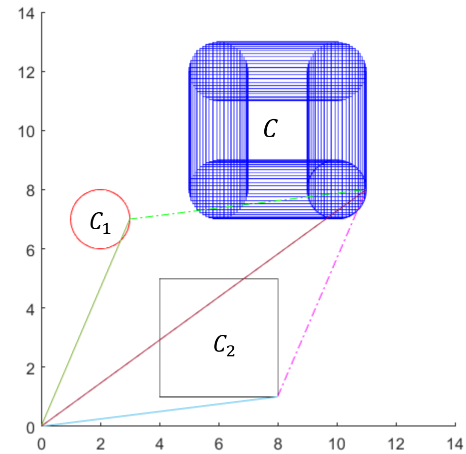


例：集合的和与差

$C=C_1+C_2=\{z\in\mathcal{R}^2|z=x+y,x\in C_1,y\in C_2\}$

$C_1=\{x\in\mathcal{R}^2|(x_1-1)^2+(x_2-7)^2=1\}$

$C_2=\{x\in\mathcal{R}^2|4\le x_1\le 8,1\le x_2\le 5\}$



$C_2=C-C_1$

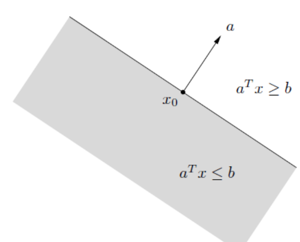
例 n 维欧氏空间中的超平面$H\triangleq\{x|a^Tx=b\}$是一个凸集，其中$b\in\mathcal{R},a\in\mathcal{R}^n\backslash\{0\}$是超平面的法向量. 此外，下面的四个半空间都是凸集.

(1)正的闭半空间 $H^+\triangleq\{x|a^Tx\ge b\}$

(2)负的闭半空间$H^-\triangleq\{x|a^Tx\le b\}$

(3)正的开半空间$\dot{H}^+\triangleq\{x|a^Tx>b\}$

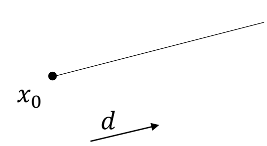
(4)负的开半空间$\dot{H}^-\triangleq\{x|a^Tx<b\}$



例 以$\boldsymbol{x}_0\in\mathcal{R}^n$为起点，$\boldsymbol{d}\in\mathcal{R}^n\backslash\{\boldsymbol{0}\}$为方向的射线

$\boldsymbol{r}(\boldsymbol{x}_0;\boldsymbol{d})\triangleq\{\boldsymbol{x}\in\mathcal{R}^n|\boldsymbol{x}=\boldsymbol{x}_0+\alpha\boldsymbol{d},\alpha\ge 0\}$

是凸集

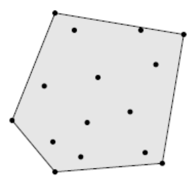


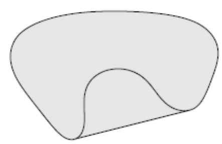
凸包

集合$\mathcal{D} \subset \mathcal{R}^n$的凸包(convex hull) 是指所有包含$\mathcal{D}$的凸集的交集，记为

$$\mathrm{conv}(\mathcal{D}) \triangleq \bigcap_{\mathcal{G} \supseteq \mathcal{D}} \mathcal{G}$$

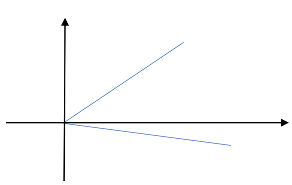
其中$\mathcal{G}$为凸集

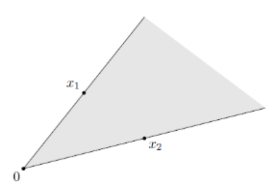




锥与凸锥

设非空集合$\mathcal{G} \subset \mathcal{R}^n$. 若对任意的$\boldsymbol{x} \in \mathcal{G}$和任意的实数$\lambda \geq 0$，有$\lambda\boldsymbol{x} \in \mathcal{G}$,
则称$\mathcal{G}$为一个锥(cone, nonnegative homogeneous). $\mathbf{0} \in \mathcal{G}$





若锥$\mathcal{G}$是凸集, 称$\mathcal{G}$为凸锥(convex cone)

例 多面体$\{\boldsymbol{x} \in \mathcal{R}^n | \boldsymbol{A}\boldsymbol{x} \geq 0\}$是凸锥, 通常称之为多面锥(polyhedral cone)

例 集合

$$\mathcal{R}^n_+ \triangleq \{\boldsymbol{x} \in \mathcal{R}^n | x_i \geq 0, i = 1, \dots, n\}$$

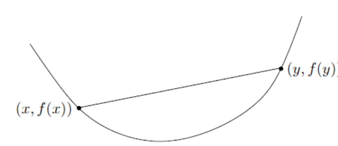
是凸锥, 通常称之为非负锥(nonnegative cone).

n维正实数集是凸集

$$\mathcal{R}^n_{++} \triangleq \{\boldsymbol{x} \in \mathcal{R}^n | x_i > 0, i = 1, \dots, n\}$$

5.2 凸函数

设函数$f: \mathcal{D} \subset \mathcal{R}^n \to \mathcal{R}$, $\mathcal{D}$为凸集



(1) f是$\mathcal{D}$上的凸函数, $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$及$\forall\, \lambda \in [0,1]$，有

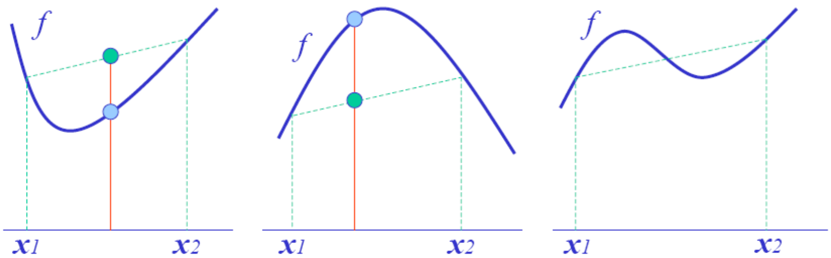
$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

(2) f是$\mathcal{D}$上的严格凸函数, $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$, $\boldsymbol{x} \neq \boldsymbol{y}$及$\forall\, \lambda \in [0,1]$, 有

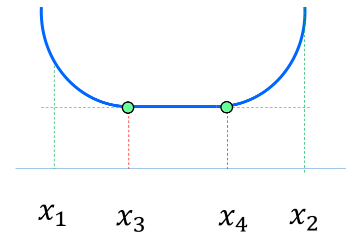
$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) < \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

(3) f是$\mathcal{D}$上的一致凸函数, $\exists$常数$\gamma > 0$, 使$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$及$\forall \lambda \in [0,1]$, 有

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) + \frac{1}{2}\lambda(1-\lambda)\gamma\|\boldsymbol{x} - \boldsymbol{y}\|^2 \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$



convex function　concave function　neither convex nor concave



非严格凸函数

严格凸函数

凸函数的基本性质

设$f_i, i=1,2,\dots,n$都是凸集$\mathcal{D}$上的凸函数，$c_i\in\mathcal{R}^+, i=1,2,\dots,n$, $\alpha\in\mathcal{R}$, 则

有 $\sum_{i=1}^{n} c_i f_i(\boldsymbol{x})$ 是$\mathcal{D}$上的凸函数

设f都是凸集$\mathcal{D}$上的凸函数, $\alpha\in\mathcal{R}$

（下）水平集

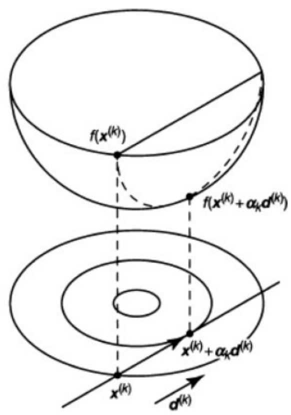
$$\mathcal{L}(f,\alpha)\triangleq\{\boldsymbol{x}|\boldsymbol{x}\in\mathcal{D}, f(\boldsymbol{x})\le\alpha\}$$

是凸集

函数是凸的 ⟺ 函数在与其定义域相交的任何直线上都是凸的

- ✓ 考虑函数$f(\boldsymbol{x})$，$g(t)=f(\boldsymbol{x}+t\boldsymbol{d})$，$g(t)$是凸函数 ⟺ $f(\boldsymbol{x})$是凸函数

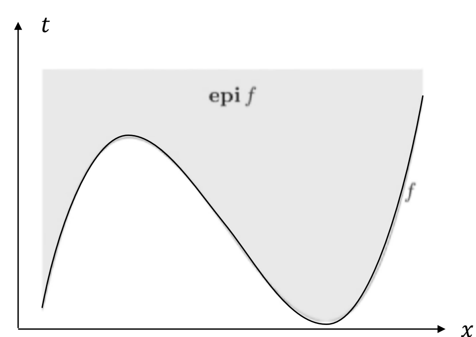
例： $\min_{x\in\mathcal{R}^n} f(x) \xrightarrow{\text{给定点}x^k\text{和方向}d^k} \min_{\alpha>0}\phi(\alpha)=f\left(x^k+\alpha d^k\right)$



函数的上境图 epigraph

$\operatorname{epi} f=\{(x,t)|x\in\operatorname{dom} f, f(x)\le t\}\subset\mathcal{R}^{n+1}$

f是凸函数 ⟺ $\operatorname{epi} f$ 是凸集



5.3凸函数的判别定理

一阶判定条件

设f 在凸集$\mathcal{D} \subset \mathcal{R}^n$上一阶连续可微，则

(1) f 在$\mathcal{D}$上为凸函数的充要条件是

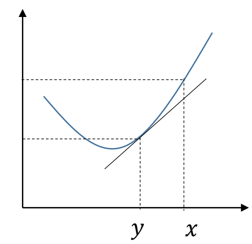
$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}), \qquad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$$

(2) f 在$\mathcal{D}$上严格凸的充要条件是

$$f(\boldsymbol{x}) > f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}), \qquad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}, \boldsymbol{x} \neq \boldsymbol{y}$$

(3) f 在$\mathcal{D}$上一致凸的充要条件是，存在常数$m > 0$

$$f(\boldsymbol{x}) > f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) + m\|\boldsymbol{x}-\boldsymbol{y}\|^2, \qquad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$$



证明： （1） 必要性

f 在$\mathcal{D}$上为凸函数，从定义出发，$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$，$\alpha \in (0,1)$

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) = f\big(\boldsymbol{y} + \alpha(\boldsymbol{x}-\boldsymbol{y})\big) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})$$

在$\boldsymbol{y}$处作一阶Taylor展开

$$f\big(\boldsymbol{y} + \alpha(\boldsymbol{x}-\boldsymbol{y})\big) = f(\boldsymbol{y}) + \alpha\nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) + o(\alpha\|\boldsymbol{x}-\boldsymbol{y}\|)$$

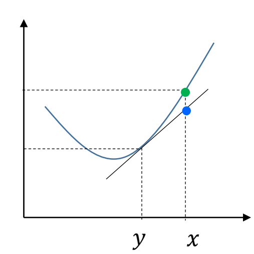
当α充分小时，

$$f\big(\boldsymbol{y} + \alpha(\boldsymbol{x}-\boldsymbol{y})\big) \approx f(\boldsymbol{y}) + \alpha\nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})$$

$$\alpha\nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) - \alpha f(\boldsymbol{y})$$

$$\nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) \leq f(\boldsymbol{x}) - f(\boldsymbol{y})$$

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y})$$



点$(y, f(y))$处的切线在x处的值在$(x, f(x))$的下方

充分性（略）

一致凸的解释

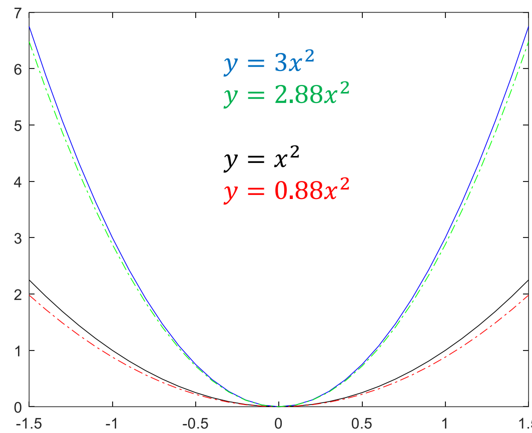
f 在$\mathcal{D}$上一致凸的充要条件是，存在常数$m>0$，使$\forall\, \boldsymbol{x},\boldsymbol{y}\in\mathcal{D}$，成立

$$f(\boldsymbol{x}) > f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) + m\|\boldsymbol{x}-\boldsymbol{y}\|^2$$

$$\nabla^2 f(\boldsymbol{x}) - \frac{m}{2}I \succ 0$$

m越大，曲线在极小点附近越陡、下降的越快

且在红色线之上



一致凸的解释

f 在$\mathcal{D}$上一致凹的充要条件是，存在常数$M>0$，使$\forall\, \boldsymbol{x},\boldsymbol{y}\in\mathcal{D}$，成立

$$f(\boldsymbol{x}) < f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) + M\|\boldsymbol{x}-\boldsymbol{y}\|^2$$

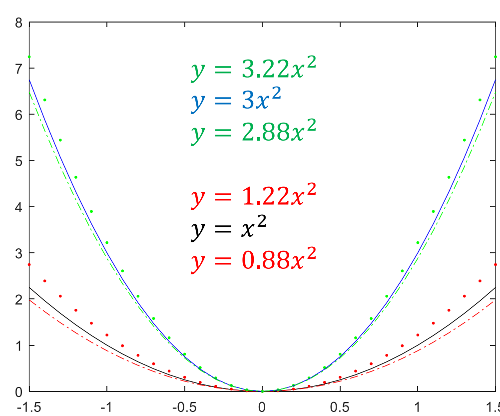
$$\nabla^2 f(\boldsymbol{x}) - \frac{M}{2}I \prec 0$$

M越大，曲线在极小点附近越陡、下降的越快

且在红色线之下

一般地，称为大M、小m条件

对函数分别提供了一个上界和一个下界



二阶判定条件

设函数f 在凸集$\mathcal{D} \subset \mathcal{R}^n$上二阶连续可微，则

(1) f 在$\mathcal{D}$上是凸的　　充要条件　是$\nabla^2 f(\boldsymbol{x}) \succcurlyeq 0, \forall\, \boldsymbol{x} \in \mathcal{D}$

(2) f 在$\mathcal{D}$上是严格凸的　充分条件　是$\nabla^2 f(\boldsymbol{x}) \succ 0, \forall\, \boldsymbol{x} \in \mathcal{D}$

(3) f 在$\mathcal{D}$上是一致凸的　充要条件　是$\nabla^2 f(\boldsymbol{x})$ 一致正定，$\forall\, \boldsymbol{x} \in \mathcal{D}$

注意，$\nabla^2 f(\boldsymbol{x}) \succ 0$是$f$ 严格凸的充分条件而非必要条件

例　$f(x) = \frac{1}{2}x^T Q x$是凸函数　$\Leftrightarrow$　$\nabla^2 f(\boldsymbol{x}) = Q \succcurlyeq 0$

例 $f(X) = \ln\det X$，$X \succ 0$ 是凹函数$\Leftrightarrow$　$\nabla^2 f(X) = -X^{-2} \prec 0$

说明　(1)　f 在$\mathcal{D}$上是凸的 $\Leftrightarrow$　$\nabla^2 f(\boldsymbol{x}) \succcurlyeq 0$

$\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$，$\boldsymbol{x} = \boldsymbol{y} + \alpha \boldsymbol{d}$，$\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{d} \neq \boldsymbol{0}$，$\alpha \in (0,1)$

充分性　$\Leftarrow$ $\nabla^2 f(\boldsymbol{x}) \succcurlyeq 0$

$f(\boldsymbol{x})$在$\boldsymbol{y}$处的二阶Taylor展开

$$f(\boldsymbol{x}) = f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) + \frac{1}{2}\alpha^2 \boldsymbol{d}^T \nabla^2 f(\boldsymbol{y}) \boldsymbol{d} + o(\|\alpha \boldsymbol{d}\|^2)$$

当 α 充分小时，$f(\boldsymbol{x}) \approx f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y}) + \frac{1}{2}\boxed{\alpha^2 \boldsymbol{d}^T \nabla^2 f(\boldsymbol{y}) \boldsymbol{d}}$

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y})$$

必要性　f 是凸的　$\Longrightarrow$　$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y})$

$$\boldsymbol{d}^T \nabla^2 f(\boldsymbol{y}) \boldsymbol{d} \geq 0$$

$$\nabla^2 f(\boldsymbol{x}) \succcurlyeq 0$$

例 判定$f(x) = 2x_1^2 + x_2^2 + x_3^2 - x_1x_2 - 4x_1 + 1$在$\mathcal{R}^3$是否为严格凸函数?

$$H = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

特征值：$\lambda_1 = 2 > 0$，$\lambda_2 = 3 - \sqrt{2} > 0$，$\lambda_3 = 3 + \sqrt{2} > 0$

H是正定矩阵，$f(x)$是严格凸函数

例 判定$f(x) = x_1^2 + x_2^2 + x_3^2 + 4x_1x_2 + 4x_1x_3 + 1$在$\mathcal{R}^3$上的凹凸性。

$$H = \begin{bmatrix} 2 & 4 & 4 \\ 4 & 2 & 0 \\ 4 & 0 & 2 \end{bmatrix}$$

特征值：$\lambda_1 = 2 > 0$，$\lambda_2 = 2 - 4\sqrt{2} < 0$，$\lambda_3 = 2 + 4\sqrt{2} > 0$

H是不定矩阵，$f(x)$既不是凸函数、也不是凹函数

例 判定$f(x) = a_1x_1 + a_2x_2 + a_3x_3 + b$在$\mathcal{R}^3$上的凹凸性。

其中，非零实数$a_i \in \mathcal{R}, i = 1,2,3$，非零实数$b \in \mathcal{R}$

$$H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

特征值：$\lambda_1 = \lambda_2 = \lambda_3 = 0$

H既是半正定的也是半负定矩阵，$f(x)$既是凸函数、也是凹函数

直线、平面既是凹集也是凸集

强凸与强光滑性

Strong Convexity

若函数$f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|^2$是一个凸函数，那么$f(\boldsymbol{x})$就是一个凸性量度为$m$的强凸函数

$$\nabla^2 f(\boldsymbol{x}) - mI \succcurlyeq 0$$

$$MI - \nabla^2 f(\boldsymbol{x}) \succcurlyeq 0$$

$$M \geq \lambda\big(\nabla^2 f(\boldsymbol{x})\big) \geq m$$

m小于Hesse阵的最小特征值

M大于Hesse阵的最大特征值

Smoothness

如果$\nabla f(\boldsymbol{x})$满足$\forall \boldsymbol{x}, \boldsymbol{y}$，$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|, L > 0$，那么称它具有$L$的光滑性量度

其实质就是**Lipschitz连续**

强凸性的以下性质等价

1. f强凸，且凸性量度为m
2. $\left(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\right)^T(\boldsymbol{x}-\boldsymbol{y}) \geq m\|\boldsymbol{x}-\boldsymbol{y}\|^2, \forall \boldsymbol{x}, \boldsymbol{y}$
3. $\nabla^2 f(\boldsymbol{x}) \succcurlyeq mI$，$\forall \boldsymbol{x}$
4. $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2$

强光滑性的以下性质等价

1. $\nabla f(\boldsymbol{x})$是Lipschitz连续的，且常数为L 。
2. $\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq L\|\boldsymbol{x}-\boldsymbol{y}\|, L > 0, \forall \boldsymbol{x}, \boldsymbol{y}$
3. $\nabla^2 f(\boldsymbol{x}) \succcurlyeq LI, \forall \boldsymbol{x}$
4. $f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) + \frac{L}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2$

参考 Amir Beck (2017), *First-order Methods in Optimization,* Chapter 5

5.4 凸优化问题

$$\begin{aligned} &\min f(\boldsymbol{x}) \\ &\text{s.t.}\, c_i(\boldsymbol{x}) = 0, i \in \mathcal{E} \\ &\quad\ \ c_i(\boldsymbol{x}) \geq 0, i \in \mathcal{I} \end{aligned}$$

$$\mathcal{I} = \{1,2, \dots, m_1\}$$
$$\mathcal{E} = \{m_1 + 1, m_1 + 2, \dots, m_1 + m_2\}$$

可行域 $X = \left\{ \boldsymbol{x} \in \mathcal{R}^n \middle| \begin{matrix} c_i(\boldsymbol{x}) = 0,\ i \in \mathcal{E} \\ c_i(\boldsymbol{x}) \geq 0,\ i \in \mathcal{I} \end{matrix} \right\}$

凸优化问题的定义

- 优化问题的可行域X是凸集
- 目标函数$f(\boldsymbol{x})$是X上的凸函数

凸优化问题的判定准则

优化问题

$$\begin{aligned} &\min f(\boldsymbol{x}) \\ &\text{s.t.}\, c_i(\boldsymbol{x}) = 0, i \in \mathcal{E} \\ &\quad\ \ c_i(\boldsymbol{x}) \geq 0, i \in \mathcal{I} \end{aligned}$$

同时满足:

①目标函数$f(\boldsymbol{x})$为凸函数

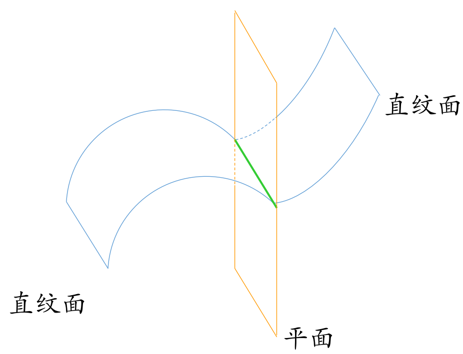
②等式约束函数$c_i(\boldsymbol{x})$，$i \in \mathcal{E}$都是线性函数

③不等式约束函数 $-c_i(\boldsymbol{x}), i \in \mathcal{I}$都是凸函数

那么，此优化问题为凸优化问题

充分条件，不是必要条件 下例说明

例 由线性等式约束和非线性等式约束的交集构成的可行域是凸集



例 判断下面的优化问题是否为凸优化问题

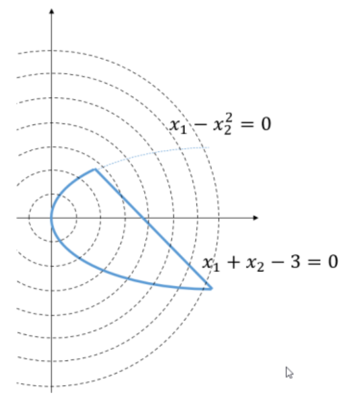
$$\min f(x) = x_1^2 + x_2^2 + 8$$

解：$H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$为正定矩阵，$f(x)$为凸函数，可行域$X = \mathcal{R}^2$为凸集

所以，该问题为凸优化问题

例 判断下面的优化问题是否为凸优化问题，并画出可行域

$$\min f(x) = x_1^2 + x_2^2 + 8$$
$$\text{s.t.}\, x_1 - x_2^2 \geq 0$$
$$x_1 + x_2 - 3 \leq 0$$



解：$f(x)$为严格凸函数

约束函数

$c_1(\boldsymbol{x}) = x_1 - x_2^2 \ \rightarrow\ -c_1(\boldsymbol{x}) = x_2^2 - x_1$ 　凸函数 $\leftarrow$ $H_{-c_1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 半正定

$c_2(\boldsymbol{x}) = -x_1 - x_2 + 3 \ \rightarrow\ -c_2(\boldsymbol{x}) = x_1 + x_2 - 3$ 　凸函数 $\leftarrow$ $H_{-c_2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ 半正定

所以，该问题为凸优化问题

说明

关于凸集、保凸运算、广义不等式、集合分离定理，以及凸函数、保凸变换等

将按照Convex Optimization的第2章和第3章的内容，专门做视频详细讲解

我会发链接到群里，供大家参考

6.1 一阶条件

$$\min f(\boldsymbol{x})$$
$$\text{s.t. } \boldsymbol{x} \in X$$

设$f:\mathcal{R}^n \to \mathcal{R}, \boldsymbol{x}^* \in X$，$f(\boldsymbol{x})$在$\boldsymbol{x}^*$处一阶可导. 若$\boldsymbol{x}^*$为$f$ 的局部极小点，则对$\boldsymbol{x}^*$处任意可行方向$\boldsymbol{d} \in \mathcal{R}^n$，总有 $\nabla f(\boldsymbol{x}^*)^T\boldsymbol{d} \geq \mathbf{0}$

证明：对$\boldsymbol{x}^*$处任意可行方向$\boldsymbol{d}$，$\exists\gamma > 0$，使$\alpha \in (0,\gamma)$满足$\boldsymbol{x}^* + \alpha\boldsymbol{d} \in X$

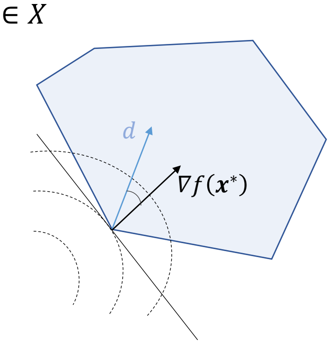
$f(\boldsymbol{x}^* + \alpha\boldsymbol{d})$在$\boldsymbol{x}^*$处Taylor一阶展开

$$f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha\nabla f(\boldsymbol{x}^*)^T\boldsymbol{d} + o(\alpha\|\boldsymbol{d}\|)$$

当α充分小时，有

$$f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) - f(\boldsymbol{x}^*) \approx \alpha\nabla f(\boldsymbol{x}^*)^T\boldsymbol{d} \geq 0$$

$\boldsymbol{x}^*$为f 的局部极小点



设$f:\mathcal{R}^n \to \mathcal{R}, \boldsymbol{x}^* \in \operatorname{int} X$, $f(\boldsymbol{x})$在$\boldsymbol{x}^*$处一阶可导. 若$\boldsymbol{x}^*$为f 的局部极小点，则$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

内点

证明： 反证法

设$\nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}$，取$\boldsymbol{d} = -\nabla f(\boldsymbol{x}^*)$，则

$$\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} = -\nabla f(\boldsymbol{x}^*)^T \nabla f(\boldsymbol{x}^*) = -\|\nabla f(\boldsymbol{x}^*)\|^2 < 0$$

所以，$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$

对无约束优化问题，$f(\boldsymbol{x})$在$\boldsymbol{x}^*$处一阶可导. 若$\boldsymbol{x}^*$为f 的局部极小点，则$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$

$X = \mathcal{R}^n$

设$f(\boldsymbol{x})$在开集$X \subset \mathcal{R}^n$上一阶连续可微. 若$\boldsymbol{x}^* \in X$是一个局部极小点, 则$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$

对开集X，$\operatorname{int} X = X$

全局极小点的充要条件

设$f(\boldsymbol{x})$ 在$\mathcal{R}^n$上是凸函数并且是一阶连续可微的. $\boldsymbol{x}^* \in \mathcal{R}^n$是$\min f(\boldsymbol{x})$的全局极小点的充要条件是$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$

证 只需证明充分性，必要性是显然的.

设$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$. 由凸函数的判别定理，可得

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*),$$

$$\forall \boldsymbol{x}^* \in \mathcal{R}^n$$

这表明$\boldsymbol{x}^*$是全局极小点.

6.2 二阶条件

设$f(\boldsymbol{x})$在开集$X \subset \mathcal{R}^n$上二阶连续可微. 若$\boldsymbol{x}^* \in X$是优化问题的一个局部极小点,
则必有$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ 且 $H(\boldsymbol{x}^*) \succcurlyeq 0$

证 设$\boldsymbol{x}^*$是一局部极小点, 那么，$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$

下面证明 $H(\boldsymbol{x}^*) \succcurlyeq 0$

任取$\boldsymbol{x} = \boldsymbol{x}^* + \alpha \boldsymbol{d} \in \mathcal{D}$, $\alpha > 0$，$\boldsymbol{d} \in \mathcal{R}^n$

由泰勒展开式，得

$$0 \le f(\boldsymbol{x}) - f(\boldsymbol{x}^*) = \frac{1}{2}\alpha^2 \boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} + o(\alpha^2)$$

$$0 \le \boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} + \frac{o(2\alpha^2)}{\alpha^2}$$

对上式令$\alpha \to 0$, 得$\boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} \ge 0$, $H(\boldsymbol{x}^*) \succcurlyeq 0$

二阶充分条件

设$f(\boldsymbol{x})$在开集$X \subset \mathcal{R}^n$上二阶连续可微. 若$\boldsymbol{x}^* \in X$满足$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$及$H(\boldsymbol{x}^*) \succ 0$,
则$\boldsymbol{x}^*$是优化问题的一个局部极小点

证 任取$\boldsymbol{x} = \boldsymbol{x}^* + \alpha \boldsymbol{d} \in \mathrm{X}$, $\alpha > 0$，$\boldsymbol{d} \in \mathcal{R}^n$

由泰勒公式得

$$f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} + \frac{1}{2}\alpha^2 \boldsymbol{d}^T H(\boldsymbol{x}^* + \theta\alpha \boldsymbol{d})\boldsymbol{d} \qquad \theta \in (0,1)$$

注意，$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$

$H(\boldsymbol{x}^*) \succ 0$和$f$ 二阶连续可微，故存在$\delta > 0$

使 $H(\boldsymbol{x}^* + \theta\alpha \boldsymbol{d}) \succ 0$ 在$\|\theta\alpha \boldsymbol{d}\| < \delta$范围内

因此

$$f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) > f(\boldsymbol{x}^*)$$

设$f:\mathcal{R}^n \to \mathcal{R}, \boldsymbol{x}^{(k)} \in X$， 开集$X \subseteq \mathcal{R}^n, f(\boldsymbol{x})$在$\boldsymbol{x}^{(k)}$处一阶可导，则$f(\boldsymbol{x})$在点$\boldsymbol{x}^{(k)}$的搜索方向$\boldsymbol{d}^{(k)} \in \mathcal{R}^n$是下降方向的充分条件是$\nabla f\left(\boldsymbol{x}^{(k)}\right)^T \boldsymbol{d}^{(k)} < 0$

证明： $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)} \in X, \alpha > 0$

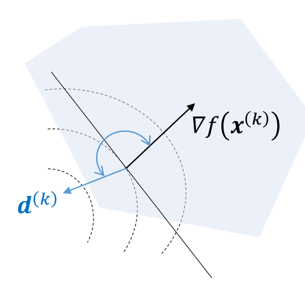
$f\left(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}\right)$在$\boldsymbol{x}^{(k)}$处Taylor一阶展开

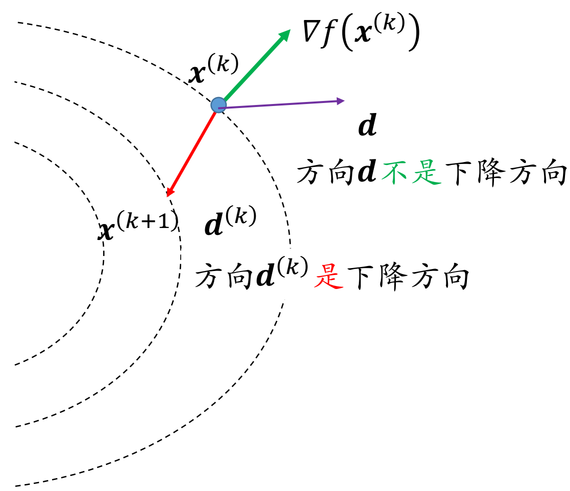
$$f\left(\boldsymbol{x}^{(k+1)}\right) = f\left(\boldsymbol{x}^{(k)}\right) + \alpha \nabla f\left(\boldsymbol{x}^{(k)}\right)^T \boldsymbol{d}^{(k)} + o\left(\alpha \left\|\boldsymbol{d}^{(k)}\right\|\right)$$

当α充分小时，$\alpha > 0$，有

$f\left(\boldsymbol{x}^{(k+1)}\right) - f\left(\boldsymbol{x}^{(k)}\right) \approx \alpha \nabla f\left(\boldsymbol{x}^{(k)}\right)^T \boldsymbol{d}^{(k)} < 0 \;\; \to \;\; \nabla f\left(\boldsymbol{x}^{(k)}\right)^T \boldsymbol{d}^{(k)} < 0$

所以，$\boldsymbol{d}^{(k)}$是下降方向





习题1

2. 判断下列函数为凸(凹) 函数或严格凸(凹) 函数:

(4) $f(\boldsymbol{x}) = 2x_1^2 + x_2^2 + 2x_3^2 + x_1x_2 - 3x_1x_3 + x_1 - x_3$， $\boldsymbol{x} \in \mathcal{R}^3$

3. 证明: $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^TQ\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$为严格凸函数当且仅当Hessian 矩阵$Q$ 正定.

11.设$A \in \mathcal{R}^{m\times n}$, $\boldsymbol{b} \in \mathcal{R}^m$, $f(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$. 试给出无约束优化问题

$$\min_{x\in\mathcal{R}^n} f(\boldsymbol{x})$$

的一阶最优性条件，并验证该条件是否是充分的，它的最优解是否唯一.

12.设$A \in \mathcal{R}^{n\times n}$, $\boldsymbol{b} \in \mathcal{R}^n$, $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^TA\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$. 试给出无约束优化问题

$$\min_{x\in\mathcal{R}^n} f(\boldsymbol{x})$$

的最优性条件.

下载安装matlab
CVX

习题课

3. $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x}-\boldsymbol{b}^T\boldsymbol{x}$

$\nabla f(\boldsymbol{x})=Q\boldsymbol{x}-\boldsymbol{b}$

$\nabla^2 f(\boldsymbol{x})=Q$

定理9 设$f(x)$ 在$\mathcal{R}^n$上是凸函数并且是一阶连续可微的. 则$\boldsymbol{x}^* \in \mathcal{R}^n$是(1.14)的全局极小点的充要条件是$\boldsymbol{g}(\boldsymbol{x}^*)=\boldsymbol{0}$.

$h(\boldsymbol{x})=\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{g}(\boldsymbol{x})$

$Dh(\boldsymbol{x})=\boldsymbol{f}(\boldsymbol{x})^T D\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{g}(\boldsymbol{x})^T D\boldsymbol{f}(\boldsymbol{x})$

11. $f(\boldsymbol{x})=\|A\boldsymbol{x}-\boldsymbol{b}\|^2=(A\boldsymbol{x}-\boldsymbol{b})^{\boldsymbol{T}}(A\boldsymbol{x}-\boldsymbol{b})$

设 $\boldsymbol{h}(\boldsymbol{x})=A\boldsymbol{x}-\boldsymbol{b}$, $D\boldsymbol{h}(\boldsymbol{x})=A$

$f(\boldsymbol{x})=\boldsymbol{h}(\boldsymbol{x})^T\boldsymbol{h}(\boldsymbol{x})$

$Df(\boldsymbol{x})=\boldsymbol{h}(\boldsymbol{x})^T D\boldsymbol{h}(\boldsymbol{x})+\boldsymbol{h}(\boldsymbol{x})^T D\boldsymbol{h}(\boldsymbol{x})$

$=2(A\boldsymbol{x}-\boldsymbol{b})^T A$

$\nabla f(\boldsymbol{x})=Df(\boldsymbol{x})^T$

$=2A^T(A\boldsymbol{x}-\boldsymbol{b})$

$=2(A^T A\boldsymbol{x}-A^T\boldsymbol{b})$

令 $\nabla f(\boldsymbol{x})=2(A^T A\boldsymbol{x}-A^T\boldsymbol{b})=\boldsymbol{0}$

$A^T A\boldsymbol{x}-A^T\boldsymbol{b}=\boldsymbol{0}$

$A^T A\boldsymbol{x}=A^T\boldsymbol{b}$

$\boldsymbol{x}^*=(A^T A)^{-1}A^T\boldsymbol{b}$

An alternative method of arriving at the least-squares solution is to proceed as follows.

$$f(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = (A\boldsymbol{x} - \boldsymbol{b})^T(A\boldsymbol{x} - \boldsymbol{b}) = \frac{1}{2}\boldsymbol{x}^T(2A^TA)\boldsymbol{x} - \boldsymbol{x}^T(2A^T\boldsymbol{b}) + \boldsymbol{b}^T\boldsymbol{b}$$

Therefore, f is a quadratic function. The quadratic term is positive definite because rank $\boldsymbol{A} = n$.

Thus, the unique minimizer of f is obtained by solving the FONC

$$\nabla f(\boldsymbol{x}) = (2A^TA)\boldsymbol{x} - (2A^T\boldsymbol{b}) = 0$$

$$\boldsymbol{x}^* = (A^TA)^{-1}A^T\boldsymbol{b}$$

11.

Theorem 12.1 The unique vector $\boldsymbol{x}^*$ that minimizes $\|\boldsymbol{Ax} - \boldsymbol{b}\|^2$ is given by the solution to the equation $\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{A}^T\boldsymbol{b}$; that is, $\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$.

Proof. Let $\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$.

$$\begin{aligned}\|\boldsymbol{Ax} - \boldsymbol{b}\|^2 &= \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*) + (\boldsymbol{Ax}^* - \boldsymbol{b})\|^2 \\ &= [\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*) + (\boldsymbol{Ax}^* - \boldsymbol{b})]^T[\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*) + (\boldsymbol{Ax}^* - \boldsymbol{b})] \\ &= \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*)\|^2 + \|\boldsymbol{Ax}^* - \boldsymbol{b}\|^2 + 2[\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*)]^T(\boldsymbol{Ax}^* - \boldsymbol{b})\end{aligned}$$

$$\begin{aligned}&[\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*)]^T(\boldsymbol{Ax}^* - \boldsymbol{b}) \\ &= (\boldsymbol{x} - \boldsymbol{x}^*)^T\boldsymbol{A}^T[\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T - \boldsymbol{I}_n]\boldsymbol{b} \\ &= (\boldsymbol{x} - \boldsymbol{x}^*)^T[(\boldsymbol{A}^T\boldsymbol{A})(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T - \boldsymbol{A}^T]\boldsymbol{b} \\ &= (\boldsymbol{x} - \boldsymbol{x}^*)^T[\boldsymbol{A}^T - \boldsymbol{A}^T]\boldsymbol{b} \\ &= 0\end{aligned}$$

$$\|\boldsymbol{Ax} - \boldsymbol{b}\|^2 = \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*)\|^2 + \|\boldsymbol{Ax}^* - \boldsymbol{b}\|^2$$

If $\boldsymbol{x} \neq \boldsymbol{x}^*$, then $\|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*)\|^2 > 0$, because rank $\boldsymbol{A} = n$.

Thus, if $\boldsymbol{x} \neq \boldsymbol{x}^*$, we have

$$\|\boldsymbol{Ax} - \boldsymbol{b}\|^2 > \|\boldsymbol{Ax}^* - \boldsymbol{b}\|^2$$

Thus, $\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$ is the unique minimizer of $\|\boldsymbol{Ax} - \boldsymbol{b}\|^2$.

12. $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$

$$\nabla f(\boldsymbol{x}) = \boldsymbol{Ax} - \boldsymbol{b} = \boldsymbol{0}$$

$$\boldsymbol{Ax} = \boldsymbol{b}$$

$$\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$$

- 第一定义（基本定义）
- 第二定义（函数值法）
- 一阶条件（梯度法）
- 二阶条件（Hesse矩阵法）

epi f

第一定义

凸函数：

设函数$f(x)$在凸集D上有定义，如果对$\forall x, y \in D, \forall \lambda \in [0,1]$,

有$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$，则称$f($

严格凸函数：

设函数$f(x)$在凸集D上有定义，如果对$\forall x, y \in D$

有$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$，则称$f($

二阶条件

设$f(x)$为定义在非空开凸集D上二阶可微函数，$D \subset R^n$，则

(1) $f(x)$为D上凸函数 $\Leftrightarrow$ $Hesse$矩阵$\nabla^2 f(x)$为半正定阵

(1) $f(x)$为D上严格凸函数 $\Rightarrow$ $Hesse$矩阵$\nabla^2 f(x)$为半正定阵

$f(x)$为D上严格凸函数 $\Leftarrow$ $Hesse$矩阵$\nabla^2 f(x)$为正定阵

第二定义

函数$f(x)$是R^n上凸函数的充要条件为对$\forall x, y \in R^n$，单变量函数

$\varphi(x) = f(x + \alpha y)$是关于$\alpha$的凸函数

一阶条件

设$f(x)$为定义在非空开凸集D上可微函数，$D \subset R^n$，则

(1) $f(x)$为D上凸函数 $\Leftrightarrow$ $f(y) \ge f(x) + \nabla f^T(x)(y-x), \forall x, y \in D$

(1) $f(x)$为D上严格凸函数 $\Leftrightarrow$ $f(y) > f(x) + \nabla f^T(x)(y-x), \forall x, y \in D, x \ne y$

$f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathrm{dom}\, f$ is a convex set and

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

for all $x, y \in \mathrm{dom}\, f$, $0 \leq \theta \leq 1$

$f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function $g : \mathbb{R} \to \mathbb{R}$,

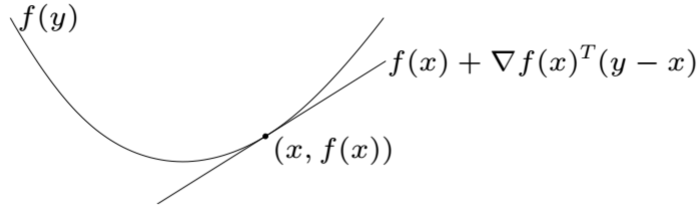
$$g(t) = f(x + tv), \quad \mathrm{dom}\ g = \{t | x + tv \in \mathrm{dom}\, f\}$$

is convex (in t) for any $x \in \mathrm{dom}\, f$, $v \in \mathbb{R}^n$

can check convexity of f by checking convexity of functions of one variable

1st-order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \operatorname{dom} f$$



first-order approximation of f is global underestimator

2nd-order conditions: for twice differentiable f with convex domain

- f is convex if and only if

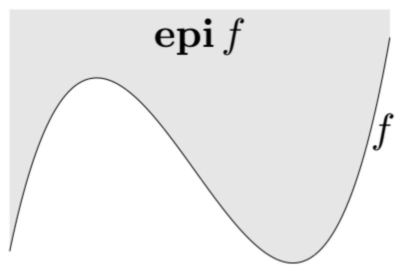
$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \operatorname{dom} f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \operatorname{dom} f$, then f is strictly convex

epigraph of $f : \mathbb{R}^n \to \mathbb{R}$:

$$\mathrm{epi}\, f = \{(x,t) \in \mathbb{R}^{n+1} | x \in \mathrm{dom}\, f, f(x) \leq t\}$$

f is convex if and only if $\mathrm{epi}\, f$ is a convex set



梯度是单调的

- A mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is monotone if

$$\langle F(x) - F(y), x - y \rangle \geq 0, \qquad x, y \in \mathbb{R}^n.$$

- A mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is uniformly monotone if there exists a constant $c > 0$ such that

$$\langle F(x) - F(y), x - y \rangle \geq c\|x - y\|^2, \qquad x, y \in \mathbb{R}^n.$$

- Suppose that $f(x) : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then $f(x)$ is convex if and only if $\nabla f(x)$ is monotone.

保凸运算

show that f is obtained from simple convex functions by operations that preserve convexity

- nonnegative weighted sum
- composition with affine function
- pointwise maximum and supremum
- composition
- minimization
- perspective

复合仿射映射

nonnegative multiple: αf is convex if f is convex, $\alpha \geq 0$

sum: $f_1 + f_2$ convex if f_1, f_2 convex (extends to infinite sums, integrals)

composition with affine function: $f(Ax + b)$ is convex if f is convex

examples

- log barrier for linear inequalities

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x), \quad \mathbf{dom}\, f = \{x | a_i^T x < b_i, i = 1, \ldots, m\}$$

- (any) norm of affine function: $f(x) = \|Ax + b\|$

逐点最大

if $f_1, \dots, f_m$ are convex, then $f(x) = \max\{f_1(x), \dots, f_m(x)\}$ is convex

examples

- piecewise-linear function: $f(x) = \max_{i=1,\dots,m}(a_i^T x + b_i)$ is convex
- sum of r largest components of $x \in \mathbb{R}^n$:

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

is convex ($x_{[i]}$ is ith largest component of x)

proof:

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} | 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$$

逐点上确界

if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, then

$$g(x) = \sup_{y \in \mathcal{A}} f(x, y)$$

is convex

examples

- support function of a set C : $S_C(x) = \sup_{y \in C} y^T x$ is convex
- distance to farthest point in a set C :

$$f(x) = \sup_{y \in C} \|x - y\|$$

复合函数

composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$:

$$f(x) = h(g(x))$$

f is convex if $\begin{array}{l} g \text{ convex}, h \text{ convex}, \tilde{h} \text{ nondecreasing} \\ g \text{ concave}, h \text{ convex}, \tilde{h} \text{ nonincreasing} \end{array}$

- proof (for $n = 1$, differentiable g, h)

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

- note: monotonicity must hold for extended-value extension $\tilde{h}$

composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), ..., g_k(x))$$

f is convex if $\begin{array}{l} g_i \text{ convex}, h \text{ convex}, \tilde{h} \text{ nondecreasing in each argument} \\ g_i \text{ concave}, h \text{ convex}, \tilde{h} \text{ nnonincreasing in each argument} \end{array}$

proof (for $n = 1$, differentiable g, h)

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

if $f(x, y)$ is convex in (x, y) and C is a convex set, then

$$g(x) = \inf_{y \in C} f(x, y)$$

is convex

the **perspective** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the function $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$,

$$g(x, t) = tf(x/t), \quad \mathrm{dom}\, g = \{(x, t) | x/t \in \mathrm{dom}\, f, t > 0\}$$

g is convex if f is convex

保凸运算: 透视函数

给定函数$f: \mathcal{R}^n \rightarrow \mathcal{R}$，则它的**透视函数** $g: \mathcal{R}^{n+1} \rightarrow \mathcal{R}$定义为

$$g(x,t) = tf\left(\frac{x}{t}\right)$$

透视函数是保凸(凹)运算
这里的t代表透视面在y上的坐标