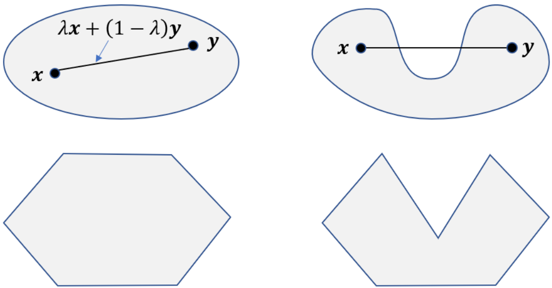
1-2　凸集和凸函数

凸集

定义1.2.1： 设集合$C \subseteq \mathbb{R}^n$，若对$\forall \boldsymbol{x}$，$\boldsymbol{y} \in C$，以及任意$\lambda \in [0,1]$，满足：$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in C$，则称$C$为**凸集**(convex set)。



从几何上看，若$\boldsymbol{x}$，$\boldsymbol{y} \in C$，则连接$\boldsymbol{x}$与$\boldsymbol{y}$的直线段包含在C内。规定空集是凸集。

常见凸集：超平面，直线，球。

超平面与闭半空间

例1.2.1：超平面

设$\boldsymbol{a} \in \mathbb{R}^n$为非零向量和$\beta$为实数，如下集合：

$$H = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{a}^T \boldsymbol{x} = \beta\}$$

称为超平面(hyperplane)。

例1.2.2：正负闭半空间

对应于超平面H的如下两个集合：

$$C_1 = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{a}^T \boldsymbol{x} \geq \beta\} \quad C_2 = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{a}^T \boldsymbol{x} \leq \beta\},$$

分别为正负闭半空间。

注意： 超平面H分离C_1与C_2。

凸集运算

定理1.2.1 凸集运算性质.

设C_1, C_2为$\mathbb{R}^n$中凸集，则

- 两个凸集的交是凸集;
$$C_1 \cap C_2 = \{\boldsymbol{x} | \boldsymbol{x} \in C_1 \text{且} \boldsymbol{x} \in C_2\}$$
- 两个凸集的和是凸集;
$$C_1 + C_2 = \{\boldsymbol{x} + \boldsymbol{y} | \boldsymbol{x} \in C_1, \boldsymbol{y} \in C_2\}$$
- 两个凸集的差是凸集;
$$C_1 - C_2 = \{\boldsymbol{x} - \boldsymbol{y} | \boldsymbol{x} \in C_1, \boldsymbol{y} \in C_2\}$$
- 凸集数乘集合为凸集.
$$\alpha C_1 = \{\alpha \boldsymbol{x} | \boldsymbol{x} \in C_1\},$$
其中α为非零实数

凸函数

定义1.2.2： 设$C \subseteq \mathbb{R}^n$为非空凸集，函数$f(\boldsymbol{x}): C \to \mathbb{R}$，对$\forall \boldsymbol{x}$，$\boldsymbol{y} \in C$，以及$\forall \lambda \in [0,1]$，有

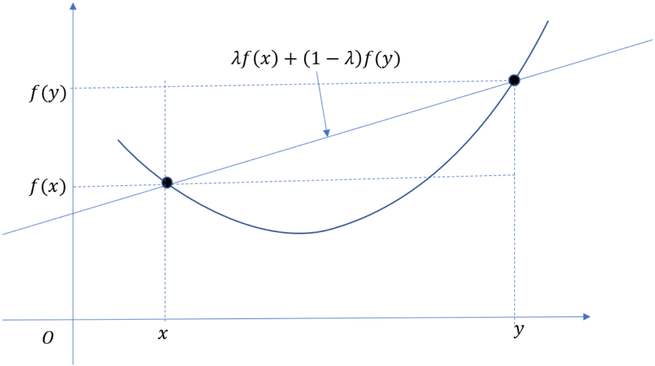
$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \le \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}),$$

成立，则称f为凸函数(convex function)。

定义1.2.3： 如果上述不等式对$\boldsymbol{x} \neq \boldsymbol{y}$与任意$\lambda \in (0,1)$严格成立，则称$f$为严格凸函数。

注意：若$-f(\boldsymbol{x})$是凸集C上的(严格)凸函数，则称$f(\boldsymbol{x})$为凸集C上的(严格)凹函数.

凸函数图示

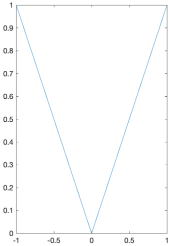


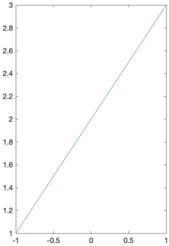
对$\forall \boldsymbol{x}, \boldsymbol{y} \in C$, 连结$(\boldsymbol{x}, f(\boldsymbol{x}))$与$(\boldsymbol{y}, f(\boldsymbol{y}))$之间的直线段位于凸函数$f(x)$图形(曲线或曲面)的上方。

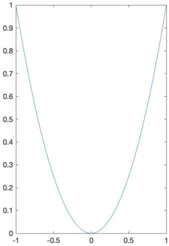
凸函数举例

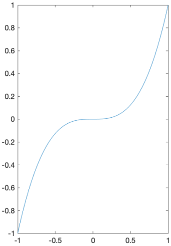
例1.2.3：一维凸函数例子

- $f_1(x) = |x|$ 为$\mathbb{R}$上凸函数；
- $f_2(x) = x + 2$ 为$\mathbb{R}$上凸函数;
- $f_3(x) = x^2$ 为$\mathbb{R}$上凸函数;
- $f_4(x) = x^3$在$\mathbb{R}$上非凸，在$\{x|x \geq 0\}$上是凸的。









凸函数运算

定理1.2.2 凸函数运算性质.

设$C \subseteq \mathbb{R}^n$为非空凸集：

- 若f是定义在C上的凸函数，α为非负实数(≥ 0)，则αf也是C上的凸函数; (非负数乘)
- 若f_1, f_2是定义在C上的凸函数，则$f_1 + f_2$也是C上的凸函数; (和函数)
- 若$f_i(\boldsymbol{x})(i = 1, 2)$是定义在$C$上的凸函数，则:
$$\max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$$
也C上的凸函数; (取极大)
- 若$f_i(x)(i = 1, 2)$是定义在C上的凸函数，则:
$$\alpha_1 f_1(\boldsymbol{x}) + \alpha_2 f_2(\boldsymbol{x})$$
也C上的凸函数，其中$\alpha_i \geq 0(i = 1, 2)$。(非负组合)

定义1.2.4 若最优化问题的可行域是凸集，且目标函数是凸函数，则称该问题为**凸规划(优化)问题**。

定理1.2.3 凸规划全局最优性.

设$\boldsymbol{x}^*$为凸规划问题的局部最优解，则$\boldsymbol{x}^*$也是全局最优解。

证明： **反证法** 设$\boldsymbol{x}^*$是局部最优解，非全局最优解，则存在$\boldsymbol{y} \neq \boldsymbol{x}^*$满足$f(\boldsymbol{y}) < f(\boldsymbol{x}^*)$. 由可行集的凸性，对$\forall \lambda \in (0,1)$，点$\lambda \boldsymbol{x}^* + (1-\lambda)\boldsymbol{y}$都是可行点。由目标函数的凸性有：

$$\begin{aligned} f(\lambda \boldsymbol{x}^* + (1-\lambda)\boldsymbol{y}) &\leq \lambda f(\boldsymbol{x}^*) + (1-\lambda) f(\boldsymbol{y}) \\ &< \lambda f(\boldsymbol{x}^*) + (1-\lambda) f(\boldsymbol{x}^*) = f(\boldsymbol{x}^*). \end{aligned}$$

与$\boldsymbol{x}^*$是局部最优矛盾。因此, $\boldsymbol{x}^*$必为全局最优解。 ■

可行域凸性

定理1.2.4 可行域凸性.

考虑非空可行域：

$$\Omega = \{\boldsymbol{x} | c_i(\boldsymbol{x}) \geq 0,\ i = 1, 2, \cdots, m\}.$$

若所有约束函数$c_i(\boldsymbol{x})$是凹函数，则可行域Ω 是凸集。

证明： 设$\boldsymbol{x}$, $\boldsymbol{y}$为两不同可行点，即有$c_i(\boldsymbol{x}) \geq 0, c_i(\boldsymbol{y}) \geq 0, i = 1, 2, \cdots, m$. 易知$-c_i(\boldsymbol{x}), i = 1, 2, \cdots, m$是凸函数。由凸函数的定义，对$\forall \lambda \in (0, 1)$，有:

$$-c_i(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq -\lambda c_i(\boldsymbol{x}) - (1-\lambda) c_i(\boldsymbol{y}) \leq 0$$

即对$\forall \lambda \in (0, 1), i = 1, 2, \cdots, m$. 有:

$$c_i(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq 0,$$

于是证明了$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \Omega$, 即$\Omega$是凸集。 ■

定理1.2.5 函数值判断.

函数$f(\boldsymbol{x})$是$\mathbb{R}^n$上凸函数的充要条件是对$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,$\psi(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{y})$是关于$\alpha$的凸函数。

证明： **必要性** 设α_1，α_2是α的任意两个值，由$\psi(\alpha)$ 的定义和$f(x)$的凸性，对任意$\lambda \in [0,1]$, 有

$$\begin{aligned}
\psi(\lambda\alpha_1 + (1-\lambda)\alpha_2) &= f(\boldsymbol{x} + (\lambda\alpha_1 + (1-\lambda)\alpha_2)\boldsymbol{y}) \\
&= f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{x} + (\lambda\alpha_1 + (1-\lambda)\alpha_2)\boldsymbol{y}) \\
&= f(\lambda(\boldsymbol{x} + \alpha_1\boldsymbol{y}) + (1-\lambda)(\boldsymbol{x} + \alpha_2\boldsymbol{y})) \\
&\leq \lambda f(\boldsymbol{x} + \alpha_1\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x} + \alpha_2\boldsymbol{y}) \\
&= \lambda\psi(\alpha_1) + (1-\lambda)\psi(\alpha_2).
\end{aligned}$$

由凸函数的定义知$\psi(\alpha)$是关于α的凸函数。

定理1.2.5证明续

证明续： **充分性** 设对$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,$\psi(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{y})$是关于$\alpha$的凸函数。任取$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,设$\bar{\boldsymbol{z}} = \boldsymbol{x} + \alpha_1 \boldsymbol{y}$, $\hat{\boldsymbol{z}} = \boldsymbol{x} + \alpha_2 \boldsymbol{y}$为方向$\boldsymbol{x} + \alpha \boldsymbol{y}$上的两个点，则对$\forall \lambda \in [0, 1]$, 有:

$$\begin{aligned}
f(\lambda \bar{\boldsymbol{z}} + (1-\lambda)\hat{\boldsymbol{z}}) &= f(\lambda(\boldsymbol{x} + \alpha_1 \boldsymbol{y}) + (1-\lambda)(\boldsymbol{x} + \alpha_2 \boldsymbol{y})) \\
&= f(\boldsymbol{x} + (\lambda \alpha_1 + (1-\lambda)\alpha_2)\boldsymbol{y}) \\
&= \psi(\lambda \alpha_1 + (1-\lambda)\alpha_2) \\
&\leq \lambda \psi(\alpha_1) + (1-\lambda)\psi(\alpha_2) \\
&= \lambda f(\boldsymbol{x} + \alpha_1 \boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x} + \alpha_2 \boldsymbol{y}) \\
&= \lambda f(\bar{\boldsymbol{z}}) + (1-\lambda) f(\hat{\boldsymbol{z}}).
\end{aligned}$$

由$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$的任意性即知$f(\boldsymbol{x})$是凸函数。 ■

定理1.2.6 一阶判断条件.

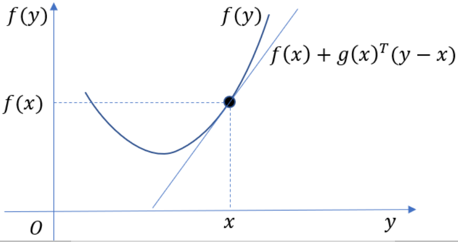
设$f(\boldsymbol{x})$是定义在凸集C上的一阶连续可微函数，则:

1. $f(\boldsymbol{x})$是凸函数的充要条件是：

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y} \in C. \tag{a}$$

2. $f(\boldsymbol{x})$是严格凸函数的充要条件是:

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) > \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y} \in C,\ \boldsymbol{x} \neq \boldsymbol{y}.$$



定理1.2.6证明

证明： ① **必要性** 设$f(\boldsymbol{x})$是凸集C上的凸函数，则对$\forall \boldsymbol{x}, \boldsymbol{y} \in C$和任意$\lambda \in (0,1)$有:$f(\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x}) \leq \lambda f(\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x})$,由此得:

$$\frac{f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda} \leq f(\boldsymbol{y}) - f(\boldsymbol{x}). \tag{b}$$

由泰勒展开式有:

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

代入(b)式得

$$\nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) + \frac{o(\lambda \|\boldsymbol{y} - \boldsymbol{x}\|)}{\lambda} \leq f(\boldsymbol{y}) - f(\boldsymbol{x}).$$

两边关于$\lambda \to 0$取极限即得: $f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x})$.

定理1.2.6证明续

证明续： **充分性** 设函数$f(\boldsymbol{x})$满足条件(a). 对$\forall \boldsymbol{x}, \boldsymbol{y} \in C$取$\bar{\boldsymbol{x}} = \lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}$，$\lambda \in (0,1)$，由$C$是凸集知$\bar{\boldsymbol{x}} \in C$.
根据条件(a) 分别有:

$$f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}) \le f(\boldsymbol{x}), \forall \boldsymbol{x} \in C,$$
$$f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{y} - \bar{\boldsymbol{x}}) \le f(\boldsymbol{y}), \forall \boldsymbol{y} \in C.$$

λ乘以第1式，$1-\lambda$乘以第2式后再相加得:

$$f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} - \bar{\boldsymbol{x}}) \le \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}),$$

由于$\bar{\boldsymbol{x}} = \lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}$，于是有：$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \le \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$.
由凸函数的定义知$f(\boldsymbol{x})$是凸集C上的凸函数。 ■

定理1.2.7 二阶判断条件.

设$f(\boldsymbol{x})$是非空开凸集$C\subseteq\mathbb{R}^n$上的二阶连续可微函数，则:

1. $f(\boldsymbol{x})$是凸函数的充分必要条件是$f(\boldsymbol{x})$的Hesse矩阵(二阶导数矩阵)$\nabla^2 f(\boldsymbol{x})$在$C$上半正定;
2. 若$f(\boldsymbol{x})$的Hesse矩阵$\nabla^2 f(\boldsymbol{x})$在$C$上正定，则$f(\boldsymbol{x})$是$C$上的严格凸函数; 反之, 如果$f(\boldsymbol{x})$是$C$上的严格凸函数，则$\nabla^2 f(\boldsymbol{x})$在$C$上半正定。

证明： ① **必要性** 任取$\bar{\boldsymbol{x}}\in C$(开凸集)，对$\forall \boldsymbol{s}\in\mathbb{R}^n$存在$\delta>0$ 使对任意$\alpha\in[0,\delta)$有$\bar{\boldsymbol{x}}+\alpha\boldsymbol{s}\in C$。$f(\boldsymbol{x})$是$C$上的凸函数，根据前面一阶凸函数判断定理有：

$$f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{s})\geq f(\bar{\boldsymbol{x}})+\alpha\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{s},\forall\boldsymbol{s}\in\mathbb{R}^n,\quad \alpha\in[0,\delta). \tag{c}$$

定理1.2.7证明续

证明续： 将$f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{s})$在$\bar{x}$处二阶泰勒展开, 得:

$$f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{s})=f(\bar{\boldsymbol{x}})+\alpha\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{s}+\frac{1}{2}\alpha^2\boldsymbol{s}^T\nabla^2 f(\bar{\boldsymbol{x}})\boldsymbol{s}+o(\|\alpha\boldsymbol{s}\|^2). \qquad \text{(d)}$$

将(d)代入(c)得:

$$\frac{1}{2}\alpha^2\boldsymbol{s}^T\nabla^2 f(\bar{\boldsymbol{x}})\boldsymbol{s}+o(\|\alpha\boldsymbol{s}\|^2)\geq 0,\forall\boldsymbol{s}\in\mathbb{R}^n.$$

上式两边除以α^2，并令$\alpha\to 0$, 取极限得

$$\boldsymbol{s}^T\nabla^2 f(\bar{\boldsymbol{x}})\boldsymbol{s}\geq 0,\forall\boldsymbol{s}\in\mathbb{R}^n,\bar{\boldsymbol{x}}\in C,$$

于是证明了$\nabla^2 f(\boldsymbol{x})$在$C$上是半正定的。

定理1.2.7证明续

证明续： **充分性** 设$\nabla^2 f$在C上半正定. 任取$\boldsymbol{x} \in C$, 将$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}} \in C$展开, 有

$$f(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \nabla^2 f(\boldsymbol{\xi})(x - \bar{\boldsymbol{x}}) \qquad \text{(e)}$$

其中$\boldsymbol{\xi} = \bar{\boldsymbol{x}} + \theta(\boldsymbol{x} - \bar{\boldsymbol{x}}), 0 < \theta < 1$。
由集合的凸性知$\boldsymbol{\xi} \in C$, 因此$\nabla^2 f$正半定, 故有:

$$(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \nabla^2 f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) \geq 0$$

代入(e)既得:

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

对$\forall \boldsymbol{x}, \bar{\boldsymbol{x}} \in C$成立。根据一阶判断定理即知$f$是$C$上的凸函数。

定理1,2.7证明续

证明续： ② 设$f(\boldsymbol{x})$的Hesse矩阵$\nabla^2 f(\boldsymbol{x})$在$C$上正定，任取两不同点$\boldsymbol{x},\boldsymbol{y}\in C$,将$f(\boldsymbol{y})$在点$\boldsymbol{x}$处展开, 有

$$f(\boldsymbol{y})=f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x})+\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^T\nabla^2 f(\boldsymbol{\xi})(\boldsymbol{y}-\boldsymbol{x}), \qquad \text{(f)}$$

其中$\boldsymbol{\xi}=\boldsymbol{x}+\theta(\boldsymbol{y}-\boldsymbol{x})\in C,\ \theta\in(0,1)$。
由$\nabla^2 f(\boldsymbol{x})$在$C$上的正定性以及$\boldsymbol{x}\neq\boldsymbol{y}$ 有$(\boldsymbol{y}-\boldsymbol{x})^T\nabla^2 f(\boldsymbol{\xi})(\boldsymbol{y}-\boldsymbol{x})>0$，代入(f)即得

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

对任意不同的$\boldsymbol{x},\boldsymbol{y}\in C$成立，根据凸函数一阶判断定理结论②知$f(\boldsymbol{x})$是$C$上的严格凸函数。

定理1.2.7证明续

证明续： 当$f(\boldsymbol{x})$是C上的严格凸函数时，$f(\boldsymbol{x})$也是C上的凸函数，由本定理的第一部分立得$\nabla^2 f(\boldsymbol{x})$在$C$上半正定的结论。 ■

注意： 严格凸函数的Hesse矩阵在C上不一定正定。考虑

$$f(x) = x^4$$

此函数显然是严格凸的，但在$x = 0$处有

$$\nabla^2 f(x) = 12x^2 = 0$$

即$\nabla^2 f(x)$在$x = 0$处不是正定的，只是正半定。

凸函数判断

例1.2.4 判断二次函数：

$$f(x_1,x_2)=x_1+3x_2+\frac{3}{2}x_1^2+\frac{3}{2}x_1x_2+2x_2^2$$

的凸性。

改写以上函数为矩阵向量形式：

$$f(x_1,x_2)=(1,3)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}+\frac{1}{2}(x_1,x_2)\begin{pmatrix} 3 & 3/2 \\ 3/2 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

二阶海森矩阵:$G(\boldsymbol{x})=\begin{pmatrix} 3 & 3/2 \\ 3/2 & 4 \end{pmatrix}$为正定矩阵(计算特征值或者各阶顺序主子式), 所以f为严格凸函数。

1-3 无约束优化最优性条件

最优性条件

最优性条件对于最优化理论的研究具有重要意义，且对最优化**算法设计**和**终止条件**的确定起重要作用。

最优性条件是指最优化问题的(局部或全局)最优解所必须满足的条件，如

- 一阶必要条件
- 二阶必要条件
- 充分必要条件(只对一些特殊的最优化问题存在)

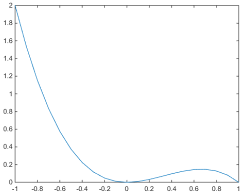
平稳点(驻点)

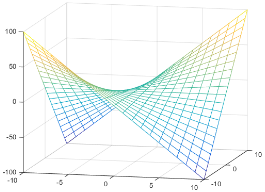
定义1.3.1 对应梯度向量为零的点称为平稳点或者驻点。

例1.3.1 分析函数的极值点与驻点：

(a) $f(x) = x^2 - x^3$极值点.

(b) $f(x, y) = xy$





(a) 根据$f'(x) = 2x - 3x^2 = 0$得到其驻点为：$x_1^* = 0, x_2^* = \frac{2}{3}$; 局部极值点一定是驻点; (b) 驻点可以不是局部极值点

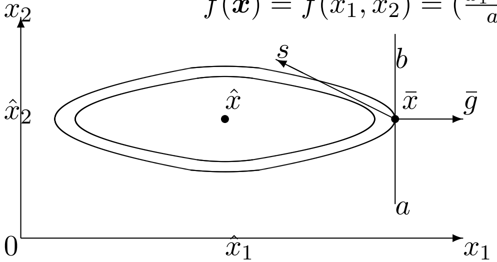
下降方向

定义1.3.2： 设$f(\boldsymbol{x})$为定义在空间$\mathbb{R}^n$上的连续函数，点$\bar{\boldsymbol{x}} \in \mathbb{R}^n$，若对于方向$\boldsymbol{s} \in \mathbb{R}^n$, 且存在数$\delta > 0$,使得

$$f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{s}) < f(\bar{\boldsymbol{x}}), \quad \forall \quad \alpha \in (0, \delta),$$

成立, 则称$\boldsymbol{s}$为$f(\boldsymbol{x})$在$\bar{\boldsymbol{x}}$处的一个下降方向。点$\bar{\boldsymbol{x}}$处的所有下降方向的全体记为$\mathcal{D}(\bar{\boldsymbol{x}})$。

$$f(\boldsymbol{x}) = f(x_1, x_2) = \left(\frac{x_1 - \hat{x}_1}{a_1}\right)^2 + \left(\frac{x_2 - \hat{x}_2}{a_2}\right)^2$$



定理1.3.1 (下降方向条件).

设函数$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}}$处连续可微，如存在非零向量$\boldsymbol{s} \in \mathbb{R}^n$使得$\nabla f(\bar{\boldsymbol{x}})^T \boldsymbol{s} < 0$成立，则$\boldsymbol{s}$是$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}}$处的下降方向。

证明: 对于充分小的$\alpha > 0$, 将$f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{s})$在点$\bar{\boldsymbol{x}}$处作Taylor展开, 有

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

由$\alpha > 0$以及$\nabla f(\bar{\boldsymbol{x}})^T \boldsymbol{s} < 0$知存在$\delta > 0$, 使对任意$\alpha \in (0, \delta)$有

$$\alpha \nabla f(\bar{\boldsymbol{x}})^T \boldsymbol{s} + o(\|\alpha \boldsymbol{s}\|) < 0$$

结合以上两式有$f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{s}) < f(\bar{\boldsymbol{x}}), \forall \alpha \in (0, \delta)$ 由此证明了$\boldsymbol{s}$是$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}}$处的下降方向. ■

一阶必要条件

定理1.3.2: (一阶必要条件).

设集合$C \subseteq \mathbb{R}^n$为开集，函数$f : C \to \mathbb{R}$连续可微，若$\boldsymbol{x}^*$为局部极小点，则$\boldsymbol{g}(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*) = 0$。

证明：设$\boldsymbol{x}^*$是局部极小点，考虑序列$\boldsymbol{x}_k = \boldsymbol{x}^* - \alpha_k \boldsymbol{g}(\boldsymbol{x}^*)$，利用Taylor展开，对于充分大的$k$，有

$$0 \leq f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) = -\alpha_k \nabla f(\boldsymbol{\eta}_k)^T \boldsymbol{g}(\boldsymbol{x}^*)$$

其中$\boldsymbol{\eta}_k$是$\boldsymbol{x}_k$与$\boldsymbol{x}^*$的凸组合。

两边同除以α_k，并取极限($\alpha_k \to 0$)。由于$f \in C^1$，故$0 \leq -\|\nabla f(\boldsymbol{x}^*)\|^2$，这意味着$\nabla f(\boldsymbol{x}^*) = 0$。 ■

定理1.3.3 (二阶必要条件).

设$f: C \subseteq \mathbb{R}^n \to \mathbb{R}$在开集$C$上二阶连续可微，若$\boldsymbol{x}^* \in C$是$f$的一个局部极小点，则:

$$\boldsymbol{g}(\boldsymbol{x}^*) = 0, \quad \boldsymbol{d}^T G(\boldsymbol{x}^*)\boldsymbol{d} \geq 0(\text{半正定})$$

证明: 定理中第1式在定理1.3.2中已证明，故只证明第2式。对任意的向量$\boldsymbol{d}$，考虑序列$\boldsymbol{x}_k = \boldsymbol{x}^* + \alpha_k \boldsymbol{d}$，由于$f \in C^2$且$\boldsymbol{g}(\boldsymbol{x}^*) = 0$，根据Taylor展开，对充分大的$k$有:

$$0 \leq f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) = \frac{1}{2}\alpha_k^2 \boldsymbol{d}^T G(\boldsymbol{\eta}_k)\boldsymbol{d}$$

其中$\boldsymbol{\eta}_k$是$\boldsymbol{x}_k$与$\boldsymbol{x}^*$的凸组合。两边同除以$\frac{1}{2}\alpha_k^2$，取极限($\alpha_k \to 0$)，得: $\boldsymbol{d}^T G(\boldsymbol{x}^*)\boldsymbol{d} \geq 0, \forall \boldsymbol{d} \in \mathbb{R}^n$，结论得证。 ■

二阶充分条件

定理1.3.3 (二阶充分条件).

设$f: C \subseteq \mathbb{R}^n \to \mathbb{R}$在开集$C$上二阶连续可微，则$\boldsymbol{x}^* \in C$是$f$的一个严格局部极小点的充分条件为：

$$\boldsymbol{g}(\boldsymbol{x}^*) = 0, \boldsymbol{d}^T G(\boldsymbol{x}^*)\boldsymbol{d} > 0(\text{正定})$$

证明: 根据条件，对任意的向量$\boldsymbol{d}$，考虑Taylor展开

$$f(\boldsymbol{x}^* + \epsilon\boldsymbol{d}) = f(\boldsymbol{x}^*) + \frac{1}{2}\epsilon^2 \boldsymbol{d}^T G(\boldsymbol{x}^* + \theta\epsilon\boldsymbol{d})\boldsymbol{d}, \epsilon > 0.$$

由于G正定，$f \in C^2$，考虑足够小的ϵ，使得$\boldsymbol{x}^* + \epsilon\boldsymbol{d} \in N_\delta(\boldsymbol{x}^*)$，有$\boldsymbol{d}^T G(\boldsymbol{x}^*)\boldsymbol{d} > 0$成立，故有：$f(\boldsymbol{x}^* + \epsilon\boldsymbol{d}) > f(\boldsymbol{x}^*)$。因此，$\boldsymbol{x}^*$为严格局部极小点. ■

定理1.3.4 凸最优性.

设$f : C \subseteq \mathbb{R}^n \to \mathbb{R}$是凸集$C$上的凸函数，且$f \in C^1$，则$\boldsymbol{x}^*$总体极小点的充要条件是$\boldsymbol{g}(\boldsymbol{x}^*) = 0$。

证明： 由凸函数判断一阶条件，f为可微凸函数,则有：

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \boldsymbol{g}(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*), \forall \boldsymbol{x} \in C$$

故$\boldsymbol{x}^*$为全局极小点. ■

注意： 一般情形下，平稳点不一定是极小点。若目标函数为凸函数时，平稳点即是极小点。

基于方向导数最优性条件

定理1.3.5 (一阶必要条件).

设集合$C \subseteq \mathbb{R}^n$为开集, 函数$f : C \to \mathbb{R}$连续可微，若$\boldsymbol{x}^*$为局部极小点，则f在$\boldsymbol{x}^*$处沿任意方向的一阶方向导数为零。

定理1.3.6 (二阶充分条件).

设集合$C \subseteq \mathbb{R}^n$为开集, 函数$f : C \to \mathbb{R}$二阶连续可微, 若f在$\boldsymbol{x}^* \in C$处沿任意方向的一阶方向导数为零，且二阶方向导数为正，则$\boldsymbol{x}^*$是f的一个严格局部极小点。