最优化理论与方法

第四讲 凸函数与

最优性条件

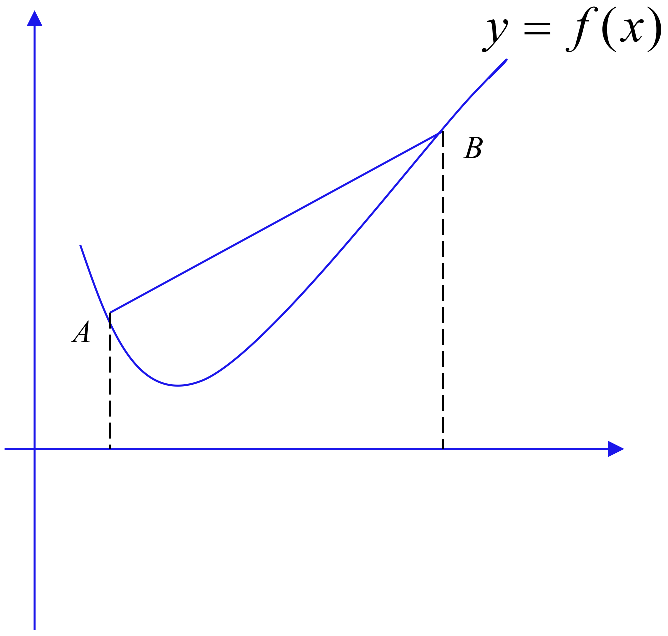
上机实验：

周二下午2:30-4：00，数学院一楼机房（3，4，5，6，7，8，12，13周）
上机作业提交截止时间：周日21：00

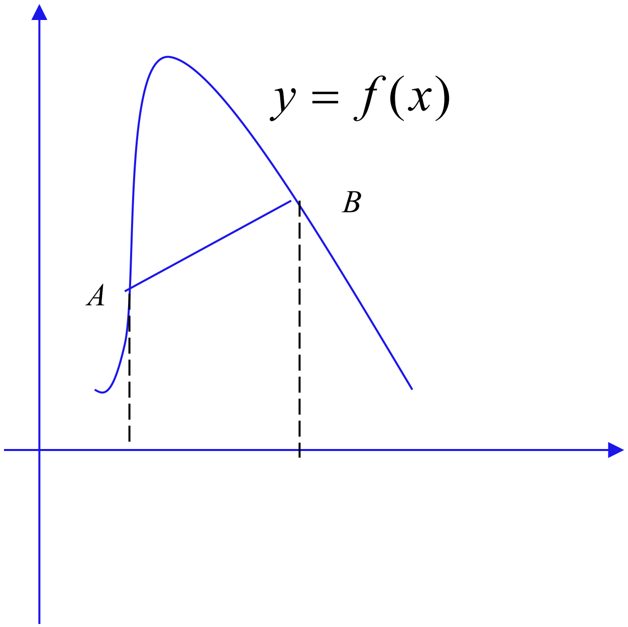
请用Python,或Matlab, 或C, C++编程

二、 凸函数

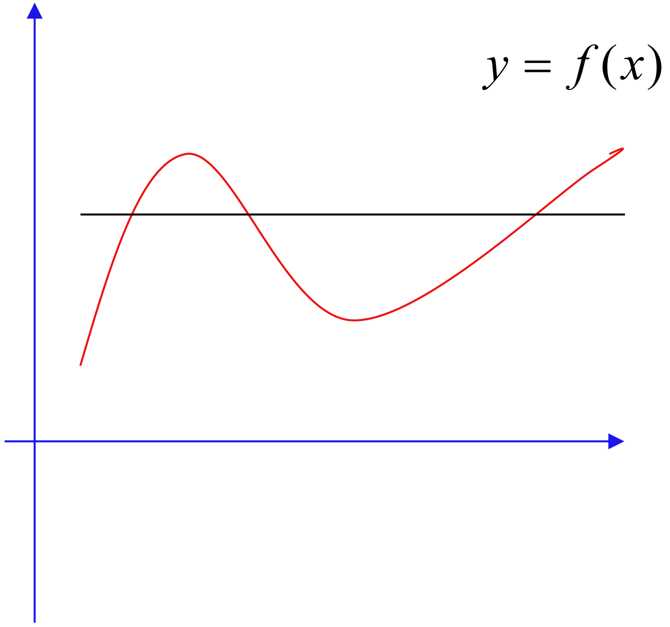
先观察三个图形：



凸函数的图形



凹函数的图形



非凸非凹函数的图形

凸函数的等价定义

凸函数 $f(\lambda \boldsymbol{x}+(1-\lambda)\boldsymbol{y}) \leqslant \lambda f(\boldsymbol{x})+(1-\lambda) f(\boldsymbol{y}), \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n},\ \lambda \in[0,1]$

$\Downarrow$ 连续可微

理论分析

$$f(\boldsymbol{y})-f(\boldsymbol{x}) \geqslant \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}), \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n}$$

或

$$(\nabla f(\boldsymbol{y})-\nabla f(\boldsymbol{x}))^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) \geqslant 0, \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n}$$

$\Downarrow$ 二阶连续可微

$$\boldsymbol{h}^{\mathrm{T}} \nabla^{2} f(\boldsymbol{x}) \boldsymbol{h} \geqslant 0, \quad \forall\ \boldsymbol{x}, \boldsymbol{h} \in \mathbb{R}^{n}$$

Hesse阵半正定

最常用的判定准则

严格凸函数：取严格不等号

一致凸函数（凸函数的加强版）

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leqslant \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) - \lambda(1-\lambda)\eta\|\boldsymbol{x}-\boldsymbol{y}\|^2,$$

$$\Big\Downarrow \text{ 连续可微}$$

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geqslant \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) + \frac{1}{2}\eta\|\boldsymbol{y}-\boldsymbol{x}\|^2, \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

或 $$(\boldsymbol{y}-\boldsymbol{x})^{\mathrm{T}}(\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x})) \geqslant \eta\|\boldsymbol{y}-\boldsymbol{x}\|^2, \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

$$\Big\Downarrow \text{ 二阶连续可微}$$

$$\boldsymbol{h}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x})\boldsymbol{h} \geqslant \eta\|\boldsymbol{h}\|^2, \quad \forall\ \boldsymbol{x},\ \boldsymbol{h} \in \mathbb{R}^n$$

- 线性函数是凸函数但不一致凸。对二次函数，一致凸与严格凸等价。
- $f(\boldsymbol{x})$ 一致凸 $\Longleftrightarrow$ $f(\boldsymbol{x}) - \eta\|\boldsymbol{x}\|^2$ 为凸函数，其中 $\eta > 0$

凸函数的简单性质

$f_1(x), f_2(x)$是凸函数 $\Rightarrow \alpha_1 f_1(x)+\alpha_2 f_2(x)$是凸函数，$\forall \alpha_1, \alpha_2 \geq 0$

推广：

$f_i(x), \mathrm{i}=1,2,\cdots,m,$ 是凸函数 $\Rightarrow$

$\alpha_1 f_1(x)+\cdots+\alpha_{\mathrm{i}} f_{\mathrm{i}}(x)$是凸函数，$\forall \alpha_{\mathrm{i}} \geq 0, i=1,2,\cdots,m$

思考：

α_i中有负数会怎么样？

凸函数的判别：

定理1.2.1 设函数$f: R^n \to R$二次连续可微. 则下列命题等价：

(1) 函数f是凸函数

(2) 对任意的$x, y \in R^n$,一元函数$\phi(t) = f(tx + (1-t)y)$是$[0,1]$上的凸函数

(3) 对任意的$x, y \in R^n$,下列的不等式成立

$$f(x) - f(y) \geq \nabla f(y)^T (x - y)$$

(4)梯度函数f单调,即

$$[\nabla f(x) - \nabla f(y)]^T (x - y) \geq 0, \forall x, y \in R^n$$

(5)对所有$x \in R^n$,$\nabla^2 f(x)$半正定.

证明：首先证命题 (1) 与命题(2)等价：

若(1)成立，对给定的$x, y \in R^n, \forall t_1, t_2 \in (0,1), \forall \alpha \in (0,1)$, 则有

$$\begin{aligned}\phi(\alpha t_1 + (1-\alpha)t_2) &= f[(\alpha t_1 + (1-\alpha)t_2)x + (1-\alpha)(\alpha t_1 + (1-\alpha)t_2)y]\\ &= f[\alpha(t_1 x + (1-t_1)y) + (1-\alpha)(t_2 x + (1-t_2)y)]\\ &\le \alpha f(t_1 x + (1-t_1)y) + (1-\alpha) f(t_2 x + (1-t_2)y)\\ &= \alpha\phi(t_1) + (1-\alpha)\phi(t_2)\end{aligned}$$

即命题(2)成立。

反之，若命题(2)成立，$\forall x, y \in R^n, \alpha \in (0,1)$, 有

$$\begin{aligned}f(\alpha x + (1-\alpha)y) &= \phi(\alpha) = \phi(\alpha \cdot 1 + (1-\alpha)\cdot 0) \le \alpha\phi(1) + (1-\alpha)\phi(0)\\ &= \alpha f(x) + (1-\alpha) f(y).\end{aligned}$$

即命题(1)成立。

其它命题证明： 按顺序：命题(1) $\Rightarrow$ 命题 (3) $\Rightarrow$ 命题(4) $\Rightarrow$ 命题(5) $\Rightarrow$ 命题(1)

例如命题(1) $\Rightarrow$ 命题(3)的证明：

若命题(1)成立，则$\forall x, y \in R^n, \alpha \in (0,1)$, 有

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y) \qquad (1)$$

又由泰勒展开式

$$f(\alpha x + (1-\alpha)y) = f(y + \alpha(x-y))$$

$$= f(y) + \alpha \nabla f(y)^T (x-y) + o(\alpha) \qquad (2)$$

由(1)和（2）可得

$$\alpha f(x) + (1-\alpha)f(y) \ge f(y) + \alpha \nabla f(y)^T (x-y) + o(\alpha)$$

即 $$\alpha[f(x) - f(y)] \ge \alpha \nabla f(y)^T (x-y) + o(\alpha)$$

上式两边除以α并令$\alpha \to 0^+$，则有

$$f(x) - f(y) \ge \nabla f(y)^T (x-y)$$

即命题(3)成立。

由凸集的定义，我们可以得到下面一个重要的凸集：

定理1.2.2 设$f(x)$是定义在凸集$S \subseteq R^n$上的凸函数，则对任意的$\alpha \in R$，水平集

$$S_\alpha = \{x \in R^n \mid f(x) \le \alpha\}$$

是R^n中的凸集

证明：$\forall x_1, x_2 \in S_\alpha$，及$\lambda \in [0,1]$，由凸函数定义，可得

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &\le \lambda f(x_1) + (1-\lambda) f(x_2) \\ &\le \lambda\alpha + (1-\lambda)\alpha = \alpha \end{aligned}$$

即$\lambda x_1 + (1-\lambda)x_2 \in \alpha$，故$S_\alpha$是凸集

进一步，

如果f(x) 是连续的凸函数，则水平集 S_α是闭凸集。

如果f(x) 是连续可微的一致凸函数，则水平集 S_α是有界闭凸集

定义1.2.6 若 $-f$是凸函数，则称函数f为凹函数，对于凹函数有类似于凸函数的判别定理

定理1.2.3 设函数$f: R^n \to R$二次连续可微. 则下列命题等价：

(1) 函数f是凹函数

(2) 对任意的$x, y \in R^n$,一元函数$\phi(t) = f(tx + (1-t)y)$是$[0,1]$上的凹函数

(3) 对任意的$x, y \in R^n$,下列的不等式成立

$$f(x) - f(y) \le \nabla f(y)^T (x - y)$$

(4)梯度函数 f 单调,即

$$[\nabla f(y) - \nabla f(x)]^T (x - y) \le 0, \forall x, y \in R^n$$

(5)对所有$x \in R^n$, $\nabla^2 f(x)$半负定.

定理1.2.5 设函数$f: R^n \to R$二次连续可微.若下面的条件之一成立，则f是严格凸函数.

(1) $f(x) - f(y) > \nabla f(y)^T (x - y), \qquad \forall x \neq y \in R^n$;

(2) 对任何$x \in R^n$,矩阵$\nabla^2 f(x)$正定.

证明: $(1) \Rightarrow f$ 严格凸

$\forall x \neq y \in R^n, \alpha \in (0,1), z = \alpha x + (1-\alpha) y$,则由(1)有

$$f(x) > f(z) + \nabla f(z)^T (x - z) \qquad f(y) > f(z) + \nabla f(z)^T (y - z)$$

从而

$$\begin{aligned} \alpha f(x) + (1-\alpha) f(y) &> f(z) + \alpha \nabla f(z)^T (x - z) + (1-\alpha) \nabla f(z)^T (y - z) \\ &= f(z) + \nabla f(z)^T [\alpha (x - z) + (1-\alpha)(y - z)] = f(z) \\ &= f\left(\alpha x + (1-\alpha) y\right) \end{aligned}$$

所以f 是严格凸函数

$(2) \Rightarrow (1) \Rightarrow f$ 严格凸

例题　已知n阶矩阵Q，n维向量q及实数c，则二次函数

$$f(x)=\frac{1}{2}x^TQx+q^Tx+c$$

是严格凸函数的充要条件是矩阵Q正定

证明：充分性：由于$\nabla f^2(x)=\frac{1}{2}(Q+Q^T)$,

若Q正定，则$\nabla f^2(x)$正定，由定理1.2.5，显然f是严格凸.

必要性：已知f严格凸，对$\forall x\neq y\in R^n$,先证

即 $$f(y)-f(x)>\nabla f(x)^{\mathrm{T}}(y-x) \qquad (1)$$

$\forall\alpha\in(0,1)$,由严格凸函数的定义，有

$$f(x+\alpha(y-x))=f(\alpha y+(1-\alpha)x)<\alpha f(y)+(1-\alpha)f(x)$$

变形可得 $$\frac{f(x+\alpha(y-x))-f(x)}{\alpha}<f(y)-f(x) \qquad (2)$$

或 $$f(y)-f(x) > \frac{f(x+\alpha(y-x))-f(x)}{\alpha} \qquad (2)$$

又由泰勒展开，得

$$f(x+\alpha(y-x))-f(x)=\alpha\nabla f(x)^{\mathrm{T}}(y-x) + \frac{\alpha^2}{2}(y-x)^{\mathrm{T}}\nabla^2 f(x)(y-x)$$

考虑到$\nabla^2 f(x)$至少半正定,将上式代入(2),得

$$f(y)-f(x) > \nabla f(x)^{\mathrm{T}}(y-x)+\frac{\alpha}{2}(y-x)^{\mathrm{T}}\nabla^2 f(x)(y-x) \geq \nabla f(x)^{\mathrm{T}}(y-x)$$

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

$\forall 0 \neq p \in R^n$,并令$y=p+x$，则有

$$f(x+p)-f(x)-\nabla f(x)^{\mathrm{T}}p=\frac{1}{2}p^{\mathrm{T}}\nabla^2 f(x)p > 0$$

这表明$\nabla^2 f(x)$正定,从而矩阵Q正定

凸规划—求凸函数在凸集上的最小值问题

一个凸规划的例子：

设$f: R^n \to R$是凸函数，$g_i: R^n \to R, i=1,2,\cdots,m_1$都是凹函数，
$h_j: R^n \to R, j=1,2,\cdots,m$都是线性函数，则下面的下面的最优化问题

$$
\begin{aligned}
\min \quad & f(x) \\
s.t. \quad & g_i(x) \ge 0, \quad i=1,2,\cdots,m_1 \\
& h_j(x) = 0, \quad j=1,2,\cdots,m
\end{aligned}
$$

是凸规划问题

只需验证可行域为凸集

例如　所有线性规划属于凸规划

非线性规划的例子：

$$
\begin{aligned}
\min \quad & 2x_1^2 + 4x_1x_2 + 16x_2^2 + -3x_1 + x_2 - 4 \\
s.t. \quad & -x_1^2 - 4x_2^2 + 10 \ge 0 \\
& x_1 - 2x_2 = 1
\end{aligned}
$$

例题1

试证明凸规划问题的局部最优解必定是其整体最优解

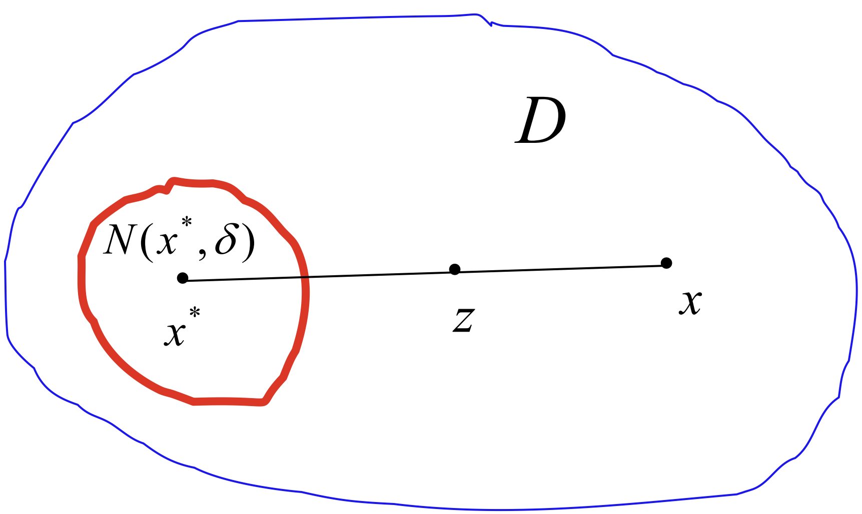
证明：设凸规划为 $\min\ f(x), x \in D \subseteq R^n$，其中 f, D 是凸的。

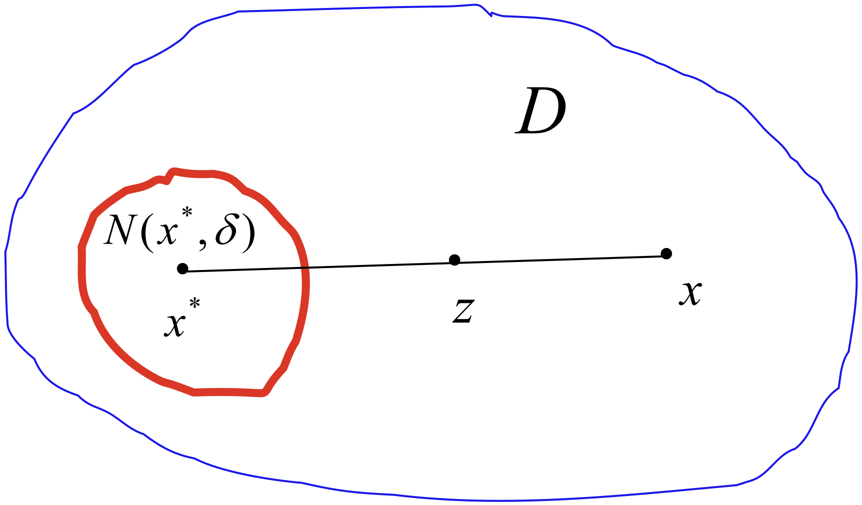
设 $x^* \in D$ 是其一个局部最优解，$\forall x \in D$，令

$$z = x^* + \alpha(x - x^*), \forall \alpha \in (0,1)$$

由于 f 是凸函数，故

$$f(z) \le \alpha f(x) + (1-\alpha) f(x^*). \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$$





另一方面由局部最优解的定义，则存在$\delta > 0$,

使得$f(z) \ge f(x^*), \forall z \in N(x^*,\delta) \bigcap D$,

由于当α充分小时，$z = x^* + \alpha(x - x^*) \in N(x^*,\delta) \bigcap D$, 故有

$$f(z) = f(x^* + \alpha(x - x^*)) \ge f(x^*) \qquad \text{..................(2)}$$

由(1), (2)得：$\alpha f(x) + (1-\alpha) f(x^*) \ge f(x^*)$，即可得

$$f(x) \ge f(x^*)$$

所以，x^*是整体最优解.

进一步，关于凸规划，我们有

例题 2

证明：已知凸规划 $\min\ f(x), x \in D \subseteq R^n$，其中$f, D$是凸的且 f 连续可微. 则$x^* \in D$是凸规划问题的最优解的充分必要条件是

$$\nabla f(x^*)^T (x - x^*) \geq 0, \forall\, x \in D$$

证明 :(1)必要性：设$x^* \in D$是凸规划问题的最优解,

若 $\exists x \in D$, 使得 $\nabla f(x^*)^T (x - x^*) < 0$, 令$p = x - x^*$,

则$\nabla f(x^*)^T p < 0$. 因此当$t > 0$充分小时,

$$f(x^* + tp) = f(x^*) + t\nabla f(x^*)^T p + o(t) < f(x^*)$$

这与$x^* \in D$是凸规划问题的最优解矛盾,故结论成立.

(2)充分性：若$\nabla f(x^*)^{\mathrm{T}} (x - x^*) \geq 0, \forall x \in D$. 由于 f 是凸函数，有

$$f(x) - f(x^*) \geq \nabla f(x^*)^{\mathrm{T}} (x - x^*) \geq 0.$$

这表明$x^* \in D$是凸规划问题的最优解

特别的，对于无约束优化问题

$$\min f(x),\quad x \in R^n \qquad (2.1)$$

定理2.1.4 若函数f是连续可微的凸函数,则x^*是问题(2.1)的最优解的充要条件是x^*满足$\nabla f(x^*)=0.$

定理2.1.4分解形式

(1）若函数f是连续可微的, x^*是问题(2.1)的局部最优解，则 x^* 满足$\nabla f(x^*)=0.$ （一阶必要性条件）

(2）若函数f是连续可微的凸函数, x^* 满足$\nabla f(x^*)=0$，则 x^* 是问题(2.1)的最优解 . （一阶充分性条件）

定理 若函数 f 是凸函数,则 f 的局部极小值点也是全局最小值点。

注意一阶必要性条件不是充分的。

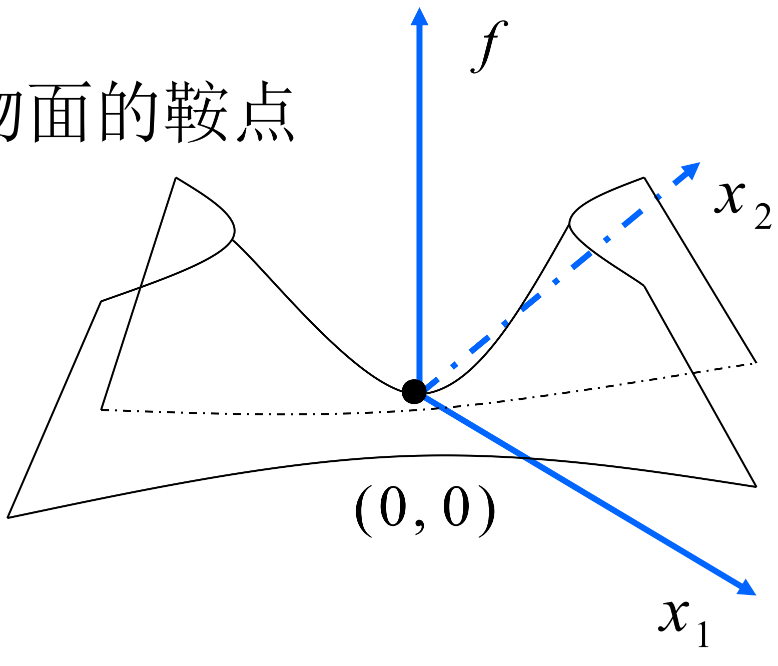
$$\nabla f(x^*) = 0.$$

例如：函数$f = x_1x_2$的图形是一双曲抛物面，

在$x^* = (0,\ 0)^T$处，显然

$$\nabla f(0,\ 0) = (0,0)^T$$

但x^*不是极小点，而是双曲抛物面的鞍点



定理：若函数f是连续可微的, x^*是问题(2.1)的局部最优解，则 x^* 满足

$$\nabla f(x^*)=0,\quad \nabla^2 f(x^*)\text{半正定}$$ （二阶必要性条件）

- 证明

利用二阶Taylor展式 $\boxed{f(x^*+tp)=f(x^*)+\frac{1}{2}t^2p^T\nabla^2 f(x^*)p+o(t^2)}$

$$f(x^*+\alpha d)-f(x^*)=\alpha\nabla f(x^*)'d+\frac{\alpha^2}{2}d'\nabla^2 f(x^*)d+o(\alpha^2)$$

由于x*是局部最小解，故 $\nabla f(x^*)=0$ 且存在充分小的正数 $\epsilon>0$ ，对于所有的 $\alpha\in(0,\epsilon)$

$$0\le\frac{f(x^*+\alpha d)-f(x^*)}{\alpha^2}=\tfrac{1}{2}d'\nabla^2 f(x^*)d+\frac{o(\alpha^2)}{\alpha^2}$$

令 $\alpha\to 0$ ，得到 $\nabla^2 f(x^*)$ 半正定

局部最优解的充分性条件

定理2.1.4 (无约束问题的二阶充分条件)

设函数$f:R^n \to R$二次连续可微, 若

$$\nabla f(x^*) = 0 \,, \quad 且 \quad \nabla^2 f(x^*) 正定,$$

则 x^*是无约束问题(2.1)的一个严格局部最优解.

$\boldsymbol{x} = \boldsymbol{x}^* + \alpha \boldsymbol{d}$ $\qquad \alpha > 0$ 充分小

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \frac{1}{2}\alpha^2 \boldsymbol{d}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} + o(\alpha^2) > f(\boldsymbol{x}^*).$$

注意： 定理2.1.4的条件不是必要的.如函数

$$f(x) = x_1^4 + x_2^4$$

显然$x^* = (0,0)^T$是严格局部极小点(最小点), 但$\nabla^2 f(x^*)$不正定

例2.1.1 利用极值条件求解下面的问题

$$\min\ f(x)=\frac{1}{3}x_1^3+\frac{1}{3}x_2^3-x_1^2-x_2$$

解：$\nabla f(x)=\begin{pmatrix} x_1^2-2x_1 \\ x_2^2-1 \end{pmatrix},\ \nabla^2 f(x)=\begin{pmatrix} 2x_1-2 & 0 \\ 0 & 2x_2 \end{pmatrix}$

由一阶必要条件$\nabla f(x)=0$，得稳定点：

$$x^{(1)}=\begin{pmatrix} 0 \\ 1 \end{pmatrix}, x^{(2)}=\begin{pmatrix} 0 \\ -1 \end{pmatrix}, x^{(3)}=\begin{pmatrix} 2 \\ 1 \end{pmatrix}, x^{(4)}=\begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

相应的Hessian矩阵为：

$\nabla^2 f(x^{(1)})=\begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$，不定　　$\nabla^2 f(x^{(2)})=\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$，负定

$\nabla^2 f(x^{(3)})=\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$，正定　　$\nabla^2 f(x^{(4)})=\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$，不定

由二阶条件知，$x^{(3)}$严格局部最优，其它三点不是极值点.

最优性条件的应用 优化问题的解析解

例

$$\min_{x>0,y\geqslant 0} f(x,y)=\frac{10}{x}+\frac{(x-y)^2}{2x}+\frac{3y^2}{2x}$$

分析与求解： 先忽略约束。

利用 $$\min_{x,y} f(x,y)=\min_{x}\min_{y} f(x,y)$$

先对固定的 x，求解关于 y 的内层优化，再求解关于 x 的外层优化 $$\min_{y} f(x,y)=\frac{10}{x}+\frac{(x-y)^2}{2x}+\frac{3y^2}{2x}$$

目标函数关于 y 为凸函数，利用最优性条件得最优解

$$y=\frac{1}{4}x$$

将上述最优解代入目标函数得外层优化问题

$$\min \ f(x) = \frac{10}{x} + \frac{3}{8}x$$

目标函数关于x为凸函数。再利用最优性条件得

$$x = \frac{4}{3}\sqrt{15}$$

这样 $y = \frac{1}{4}x = \frac{1}{3}\sqrt{15}$

它们满足约束条件，自然为原问题的最优解。

总结

- 凸函数及其性质
- 凸规划问题的最优性条件
- 无约束优化问题的:

  必要性条件(一阶,二阶)

  充分条件

  充分必要条件

作 业

- 课堂讲过的定理和性质及证明,请自己独立证明.

- 习题1:

  9，14，16