西安电子科技大学
XIDIAN UNIVERSITY

智能感知与图像理解

最优化理论

第二章 最优化问题的数学基础 II

人工智能学院
智能感知与图像理解实验室

第二章 最优化问题的数学基础 II

1 极小点的判定条件

2 锥，凸集及其性质

3 凸函数及其性质

4 函数的正定性判别和举例

极小点的判定条件

邻域的定义

定义 2.7 对于任意给定的实数$\delta > 0$,满足不等式$\|X - X_0\| < \delta$的X的集合称为点X_0的邻域，记为

$$N(X_0, \delta) = \left\{ X \mid \|X - X_0\| < \delta, \delta > 0 \right\}$$

极小点的定义

定义 2.8　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,若存在点$X^* \in D$和数$\delta > 0, \forall X \in N(X^*, \delta) \cap D$都有$f(X^*) \le f(X)$,则称$X^*$为$f(X)$的局部极小点(非严格).

定义 2.9　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,若存在点$X^* \in D$和数$\delta > 0, \forall X \in N(X^*, \delta) \cap D$但$X \ne X^*$, 都有$f(X^*) < f(X)$,则称$X^*$为$f(X)$的严格局部极小点.

定义 2.10　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,若存在点$X^* \in D, \forall X \in D$都有$f(X^*) \le f(X)$,则称$X^*$为$f(X)$在$D$上的全局极小点（非严格）.

定义 2.11　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,若存在点$X^* \in D, \forall X \in D$但$X \ne X^*$,都有$f(X^*) < f(X)$, 则称$X^*$为$f(X)$在$D$上的严格全局极小点.



极小点的判定条件

定理2.3　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,具有连续的一阶偏导数. 若$X^*$是 $f(X)$ 的局部极小点并且是D的内点，则

$$\nabla f(X^*) = 0$$

证　设e是任意单位向量，因为X^* 是$f(X)$ 的局部极小点，所以存在 $\delta > 0$,当$|t| < \delta$或$X^* + te \in N(X^*, \delta)$时总有

$$f(X^* + te) \geq f(X^*)$$

极小点的判定条件

定理2.3　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,具有连续的一阶偏导数. 若$X^*$是 $f(X)$ 的局部极小点并且是D的内点，则

$$\nabla f(X^*) = 0$$

引入辅助一元函数$\varphi(t) = f(X^* + te)$，此时，由$f(X^* + te) \geq f(X^*)$可得$\varphi(t) \geq \varphi(0)$，又因为$X^*$是$D$的内点，所以$t=0$是$\varphi(t)$的局部极小点。根据一元函数极小点的必要条件得到$\varphi'(0) = 0$,

$$\varphi'(t)\big|_{t=0} = f'(X^* + te)^T e\big|_{t=0} = \frac{\partial f(X^*)}{\partial e} = \nabla f(X^*)^T e$$

$$\nabla f(X^*)^T e = 0$$

$$\nabla f(X^*) = 0$$


极小点的判定条件

必要非充分条件

例：$f(x_1, x_2) = x_1 x_2$

在$X^* = [0,0]^T$处的梯度

$\nabla f = [0,0]^T$，但$X^* = [0,0]^T$

是双曲面的鞍点，而不是极小点

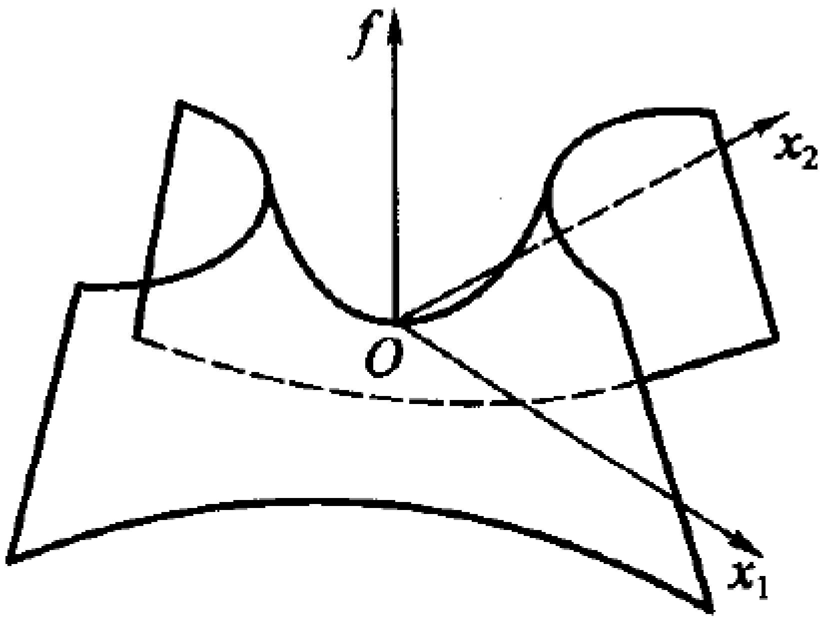
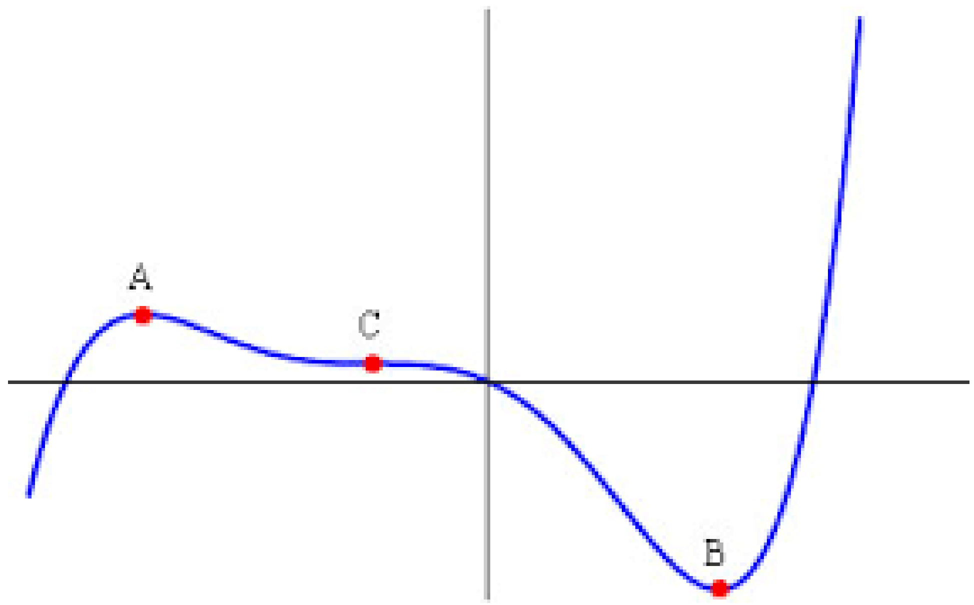


图 2.2

极小点的判定条件

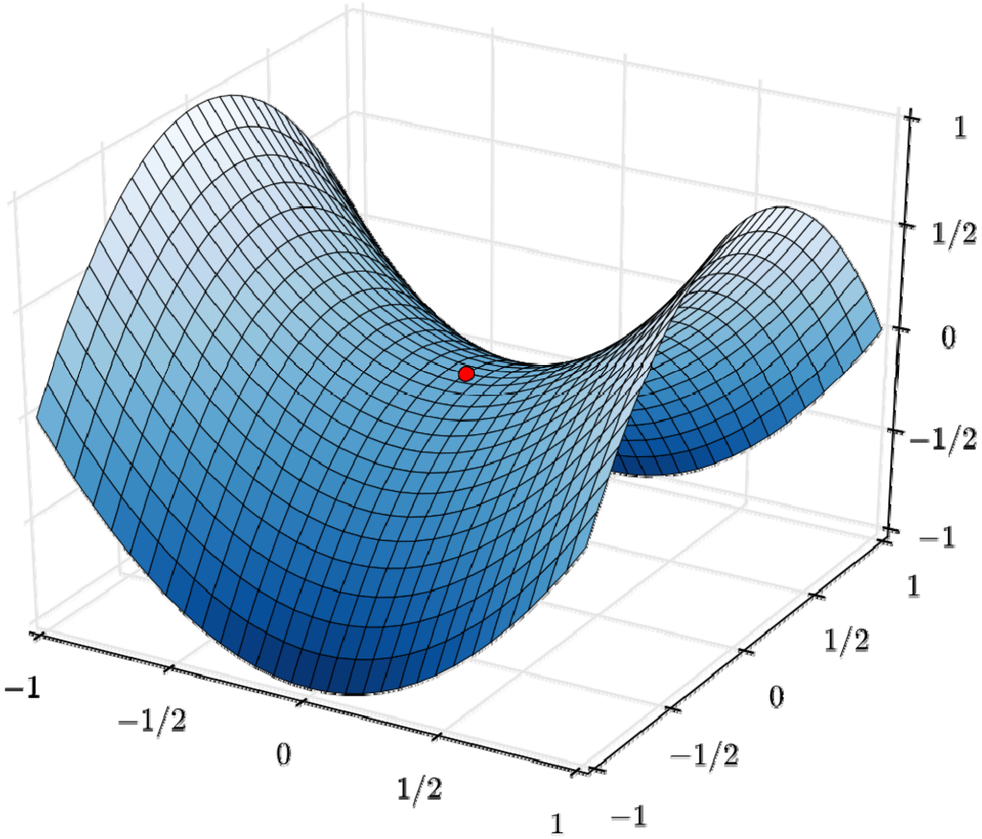
定义2.12 设 $f:D\subset \mathbf{R}^n \to \mathbf{R}^1$, X^* 是D的内点. 若 $\nabla f(X^*)=0$

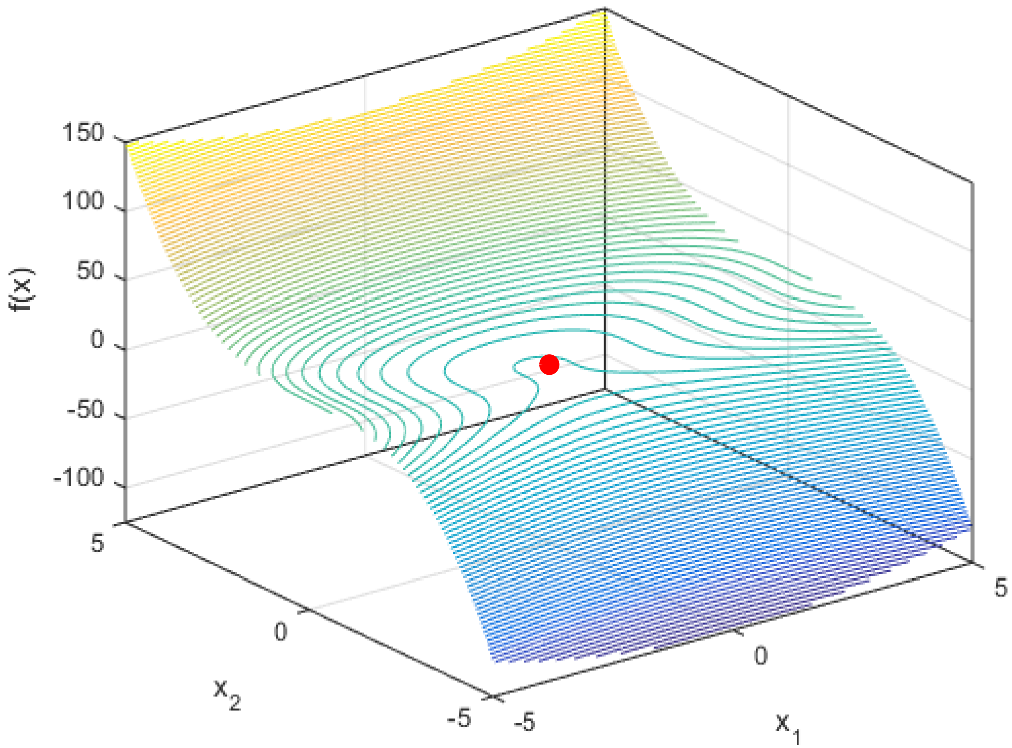
则称 X^* 为 $f(X)$的驻点.



极小点的判定条件

二维非极小点驻点举例





极小点的判定条件

定理2.4　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,具有连续二阶偏导数，$X^*$是$D$的一个内点.若$\nabla f(X^*) = 0$，并且$\nabla^2 f(X^*)$是正定的，则$X^*$是$f(X)$的严格局部极小点.

证　因为$\nabla^2 f(X^*)$是正定矩阵，则对于任意$P \in R^n$有

$$P^T \nabla^2 f(X^*) P > 0$$



极小点的判定条件

定理2.4　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,具有连续二阶偏导数，$X^*$是D的一个内点. 若$\nabla f(X^*)=0$，并且$\nabla^2 f(X^*)$是正定的，则$X^*$是$f(X)$的严格局部极小点.

将$f(X)$在X^*处按泰勒公式展开



Hesse矩阵与泰勒展开

复习泰勒展开

二阶泰勒展开式：

设函数f：$R^n \to R$具有二阶连续偏导数，则

$$f(X)=f(X^*)+\nabla f(X^*)^T(X-X^*)+\frac{1}{2}(X-X^*)^T\nabla^2 f(X^*)^T(X-X^*)+o(\|(X-X^*)\|^2)$$

其中$o(\|(X-X^*)\|^2)$ 当$\|(X-X^*)\|^2 \to 0$时，是关于$(X-X^*)^2$的高阶无穷小量。

极小点的判定条件

定理2.4 设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,具有连续二阶偏导数，$X^*$是D的一个内点. 若$\nabla f(X^*) = 0$，并且$\nabla^2 f(X^*)$是正定的，则$X^*$ 是$f(X)$的严格局部极小点.

将$f(X)$在X^*处按泰勒公式展开

$$f(X) = f(X^*) + \nabla f(X^*)^T (X - X^*) + \frac{1}{2}(X - X^*)^T \nabla^2 f(X^*)^T (X - X^*) + o(\|(X - X^*)\|^2)$$

$$f(X) - f(X^*) = \frac{1}{2}(X - X^*)^T \nabla^2 f(X^*)^T (X - X^*) + o(\|(X - X^*)\|^2)$$

极小点的判定条件

定理2.4　设 $f: D \subseteq \mathbf{R}^n \to \mathbf{R}^1$,具有连续二阶偏导数，$X^*$是D的一个内点. 若$\nabla f(X^*)=0$，并且$\nabla^2 f(X^*)$是正定的，则$X^*$是$f(X)$的严格局部极小点.

$$f(X)-f(X^*)=\frac{1}{2}(X-X^*)^T\nabla^2 f(X^*)^T(X-X^*)\ +o(\|(X-X^*)\|^2)$$

当$X \to X^*$时，上式左端的符号取决于右端第一项，

又 $\frac{1}{2}(X-X^*)^T\nabla^2 f(X^*)^T(X-X^*)>0$,

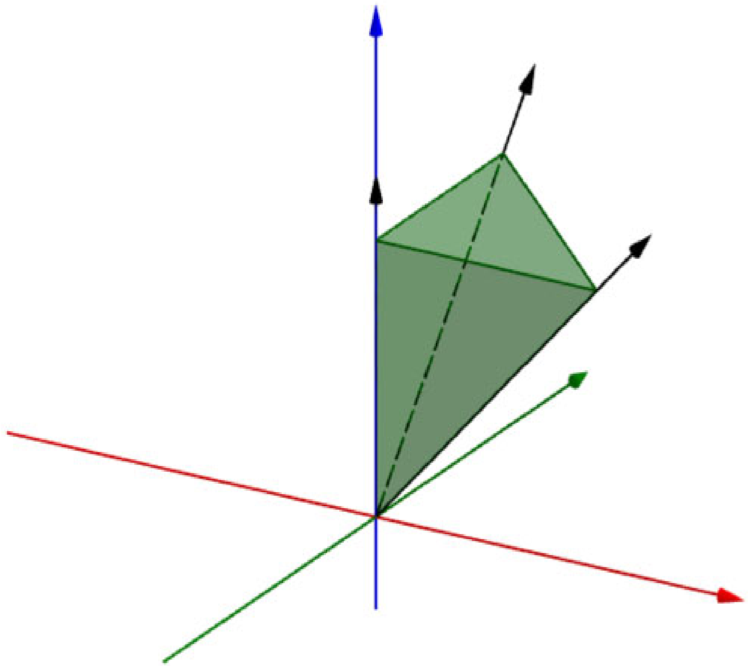
因此

$$f(X)>f(X^*)$$

锥

定义2.13 集合$C \subset \mathbf{R}^n$.若$\forall X \in C$及任意的数$\lambda \geq 0$,

均有$\lambda X \in C$，则称C为锥.

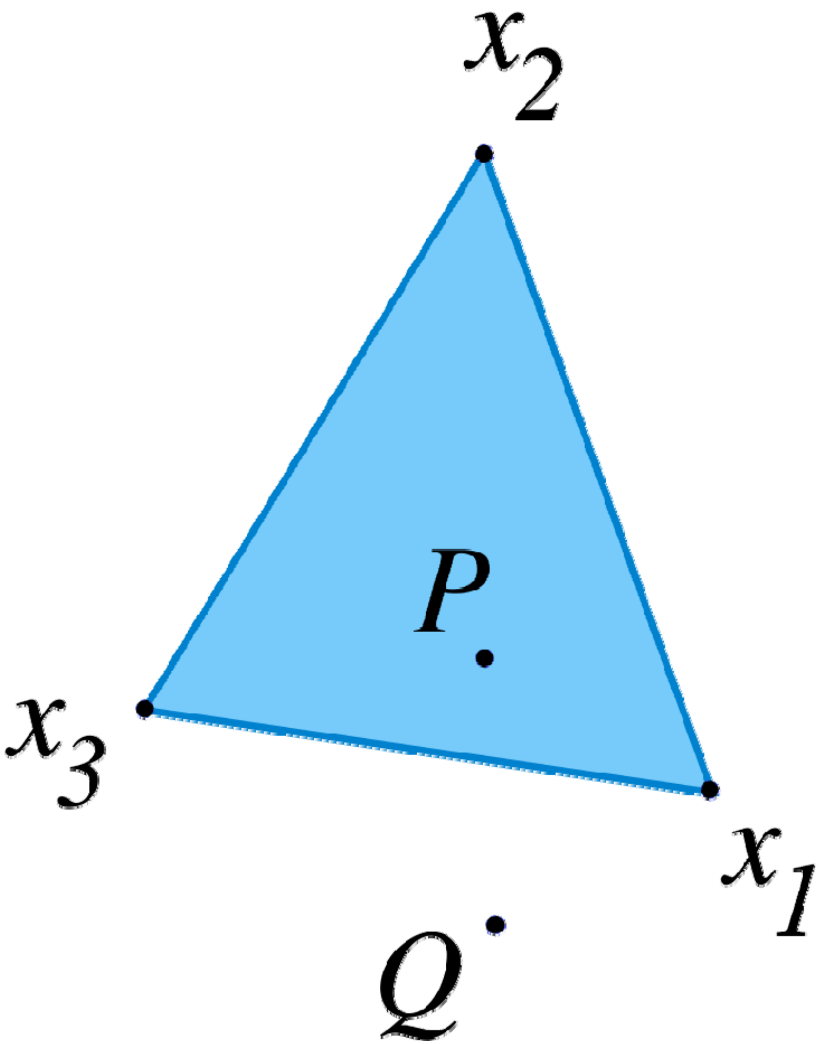


凸组合

定义2.14　设$X_1, X_2, \cdots X_l$是$\mathbf{R}^n$中的l个已知点.若对于某点$X \in \mathbf{R}^n$，存在常数$\lambda_1, \lambda_2, \cdots \lambda_l \geq 0$且$\sum_{i=1}^{l} \lambda_i = 1$使得$X = \sum_{i=1}^{l} \lambda_i X_i$，则称$X$是$X_1, X_2, \cdots X_l$的凸组合. 若$\lambda_1, \lambda_2, \cdots \lambda_l > 0$且$\sum_{i=1}^{l} \lambda_i = 1$，则称$X$是$X_1, X_2, \cdots X_l$的严格凸组合.



凸组合

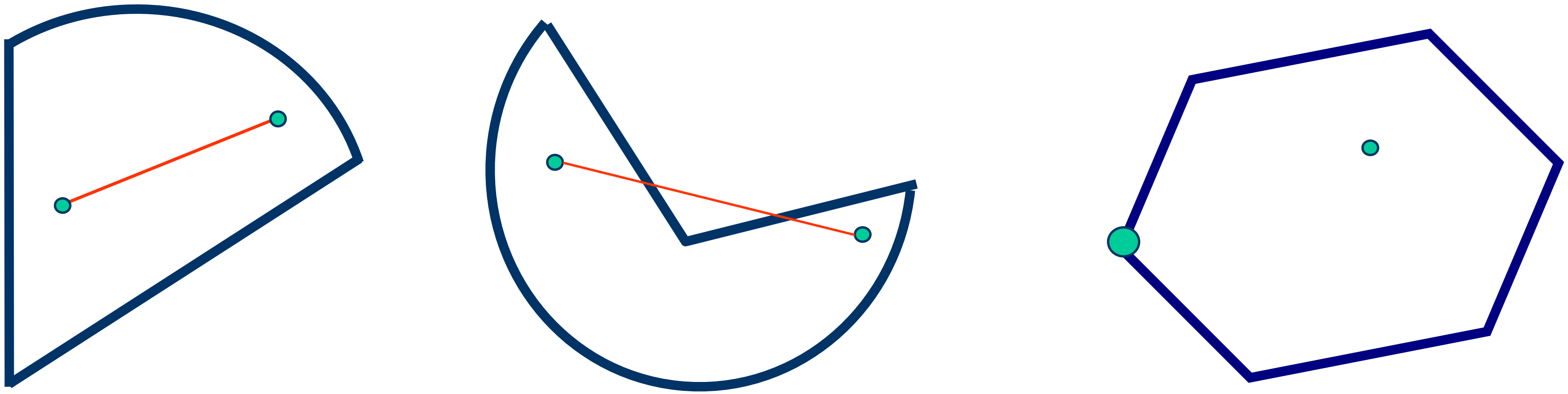


凸集

定义2.15 若集合$C \subset \mathbf{R}^n$. 若$\forall X_1 \in C$和$\forall X_2 \in C$，以及任意的数$\lambda \in [0,1]$, 均有

$$\lambda X_1 + (1-\lambda) X_2 \in C,$$

则称C为凸集.



凸集

定理2.6 任意一组凸集的交仍然是凸集

证 设$C=\bigcap_{i\in I} C_i$，C_i都是凸集，I是其下表集合.任取X_1，$X_2 \in C$，则对于任意$i\in I$都有X_1，$X_2\in C_i$.任取λ_1，$\lambda_2\in[0,1]$且有$\lambda_1+\lambda_2=1$，因C_i是凸集，有

$$\lambda_1 X_1+\lambda_2 X_2\in C_i.$$

凸集

定理2.6 任意一组凸集的交仍然是凸集

证 于是有

$$\lambda_1 X_1 + \lambda_2 X_2 \in \bigcap_{i \in I} C_i = C,$$

即 C 是凸集.



半空间

定义2.16 设$a \in \mathbf{R}^n$且$a \neq 0, b \in \mathbf{R}^1$,则集合

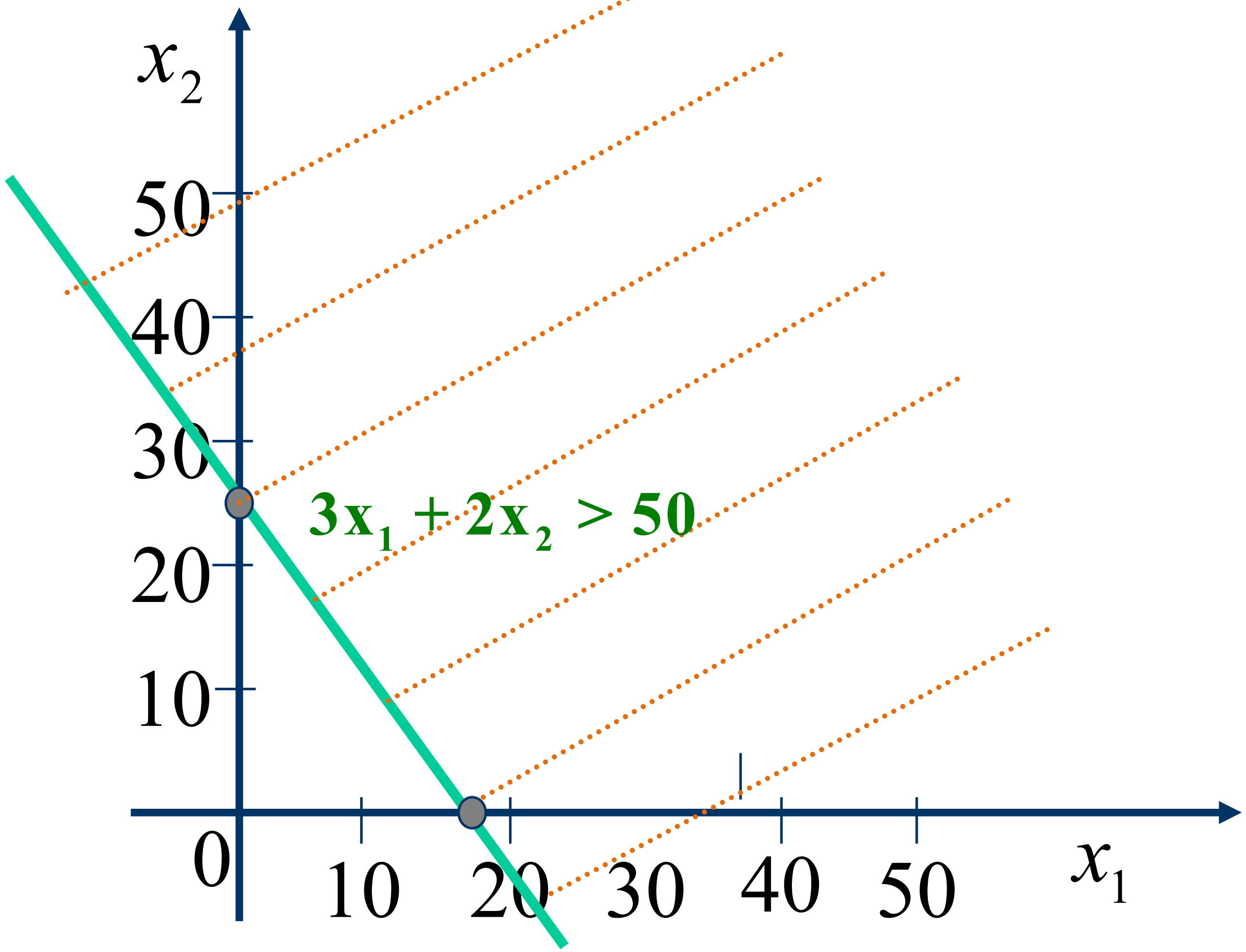
$$\left\{ X \mid a^T X > b, X \in \mathbf{R}^n \right\}$$

称为$\mathbf{R}^n$的半空间.

锥，凸集及其性质

半空间

$$3x_1 + 2x_2 > 50$$

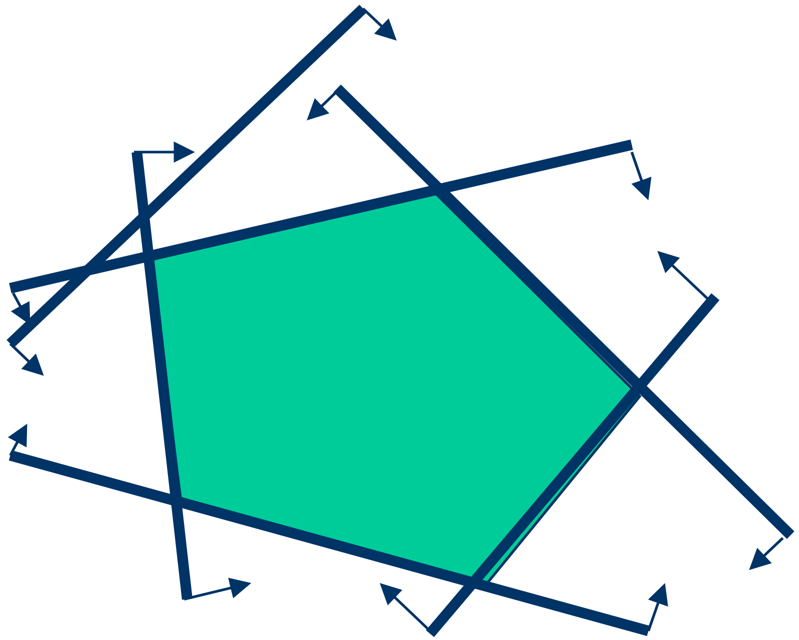


凸集

特别地，规定空集是凸集.容易验证，空间$\mathbf{R}^n$、半空间超平面、直线、点、球都是凸集.

半空间

定义2.17 有限个半空间的交$\{X \mid AX \leq b\}$称为多面集，其中A为$m \times n$矩阵，b为m维向量。

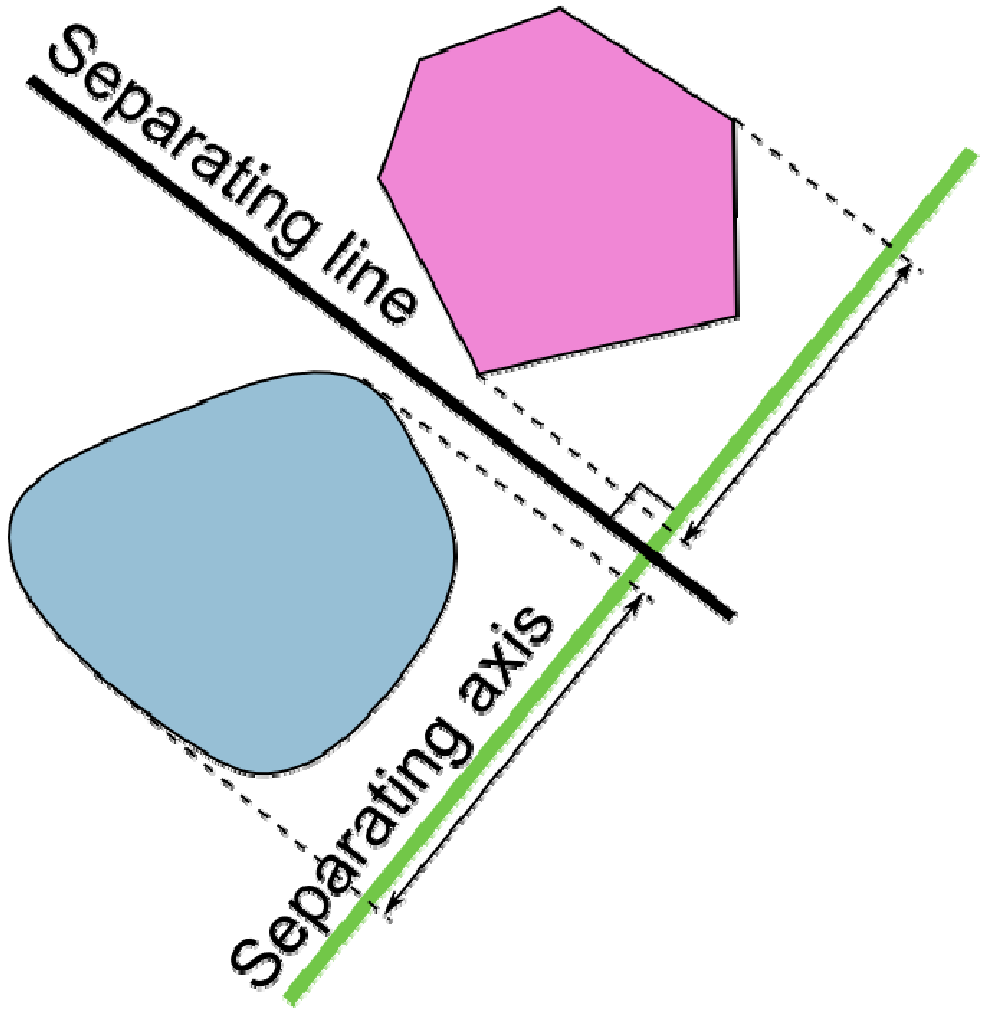


凸集分离定理

定义2.20 设C_1和C_2是R^n中两个非空集合，$H=\{X \mid p^T X=\alpha\}$为超平面，如果对$\forall X \in C_1$都有$p^T X \geq \alpha$，对$\forall X \in C_2$，都有$p^T X \leq \alpha$（或情形恰好相反），则称超平面$H$分离集合$C_1$和$C_2$。



凸集分离定理



凸函数

定义2.21　设C是R^n中的非空凸集，$f(x)$是定义在C上的函数,$\forall X_1, X_2 \in C$,以及$\forall \lambda \in (0,1)$,若有

$$f(\lambda X_1 + (1-\lambda) X_2) \le \lambda f(X_1) + (1-\lambda) f(X_2)$$

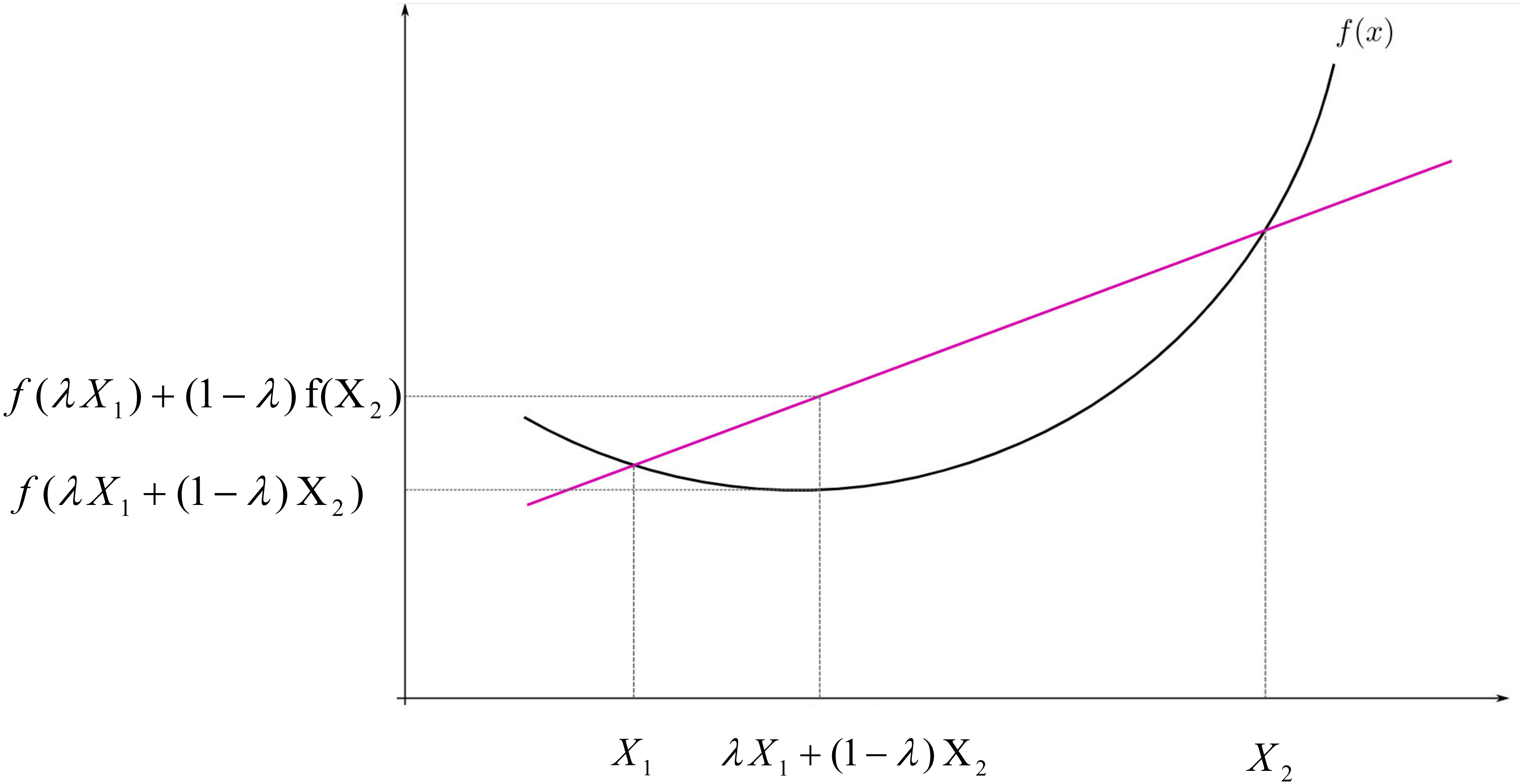
则称$f(X)$是C上的凸函数.若有

$$f(\lambda X_1 + (1-\lambda) X_2) < \lambda f(X_1) + (1-\lambda) f(X_2)$$

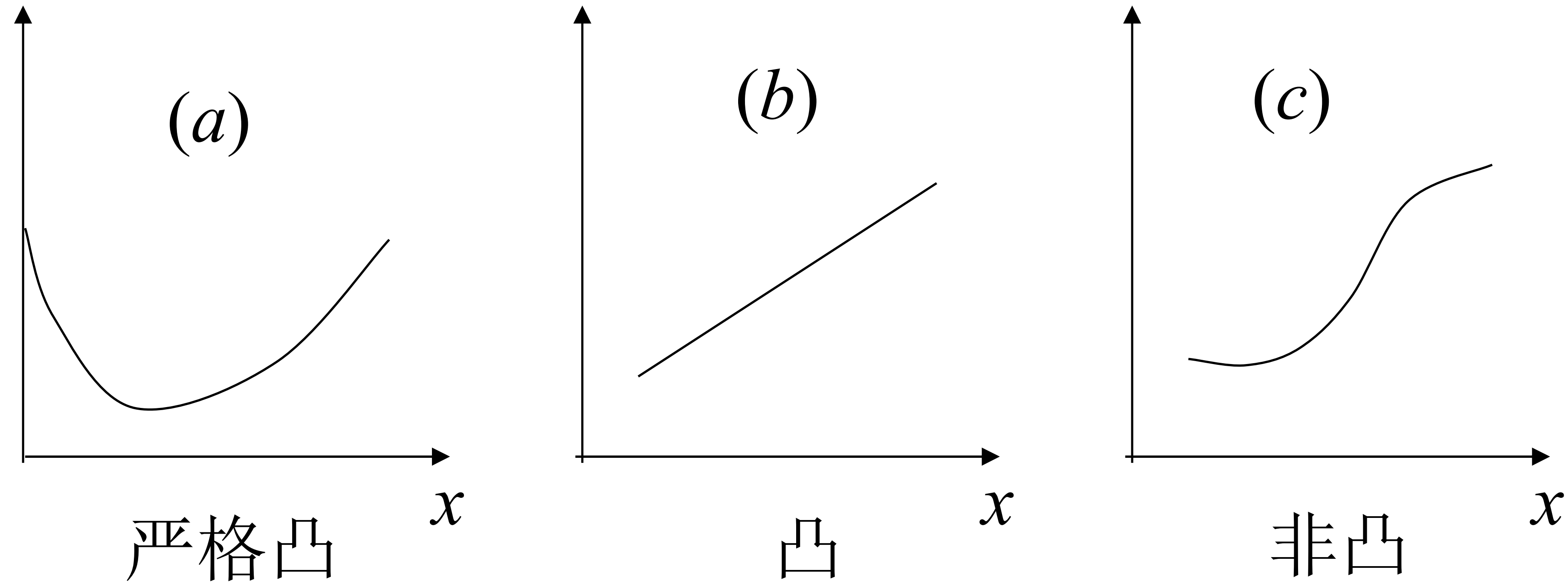
则称$f(X)$是C上的严格凸函数.

凸函数及其性质

凸函数


$f(x)$
$f(\lambda X_1)+(1-\lambda)\,\mathrm{f}(\mathrm{X}_2)$
$f(\lambda X_1+(1-\lambda)\,\mathrm{X}_2)$
X_1
$\lambda X_1+(1-\lambda)\,\mathrm{X}_2$
X_2

凸函数



凸函数的性质

(1) 设$f_1(X)$，$f_2(X)$是凸集C上的凸函数，则函数 $f_1(X)+f_2(X)$在S上也是凸函数。

(2) 设$f(X)$是凸集c上的凸函数，则对任意的$a\geq 0$，函数 $af(x)$是凸的。

推广：设$f_1(X)$，$f_2(\mathrm{X}), \ldots, f_k(X)$是凸集$C$上的凸函数，$a_i\geq 0$, 则$a_1f_1(X)+a_2f_2(X)+\ldots+a_kf_k(X)$也是凸集$C$上的凸函数.

凸函数的判定

定理2.21　设f：$C \subseteq R^n \to R^1$是可微函数，其中C为凸集，则

(1)　f为凸函数的充要条件是$\forall X_1, X_2$,都有

$$f(X_2) \geq f(X_1) + \nabla f(X_1)^T (X_2 - X_1);$$

(2)　f为严格凸函数的充要条件是$\forall X_1, X_2$且$X_1 \neq X_2$都有

$$f(X_2) > f(X_1) + \nabla f(X_1)^T (X_2 - X_1).$$

凸函数的判定

证明:“$\Rightarrow$”

设f是C上的凸函数，则对$\forall X_1, X_2 \in C$及$\lambda \in (0,1)$，有

$$f(\lambda X_2 + (1-\lambda)X_1) \le \lambda f(X_2) + (1-\lambda) f(X_1)$$

即
$$\frac{f(X_1 + \lambda(X_2 - X_1)) - f(X_1)}{\lambda} \le f(X_2) - f(X_1)$$

凸函数的判定

证明:“$\Rightarrow$”

令$\lambda \to 0^+$，由f的可微性，得f在点X_1关于方向$X_2 - X_1$的方向导数

$$\frac{\partial f(X_1)}{\partial (X_2 - X_1)} = \nabla f(X_1)^T (X_2 - X_1) \le f(X_2) - f(X_1),$$

$$\Rightarrow \ f(X_2) \ge f(X_1) + \nabla f(X_1)^T (X_2 - X_1).$$

凸函数的判定

证明：“$\Leftarrow$”

设对$\forall X_1, X_2 \in C$，有

$$f(X_2) \geq f(X_1) + \nabla f(X_1)^T (X_2 - X_2)。$$

$\forall \lambda \in (0,1)$，令$X = \lambda X_1 + (1-\lambda) X_2$，则$X \in C$。

由假设，对X_1, X_2及$X \in C$有

$$f(X_1) \geq f(X) + \nabla f(X)^T (X_1 - X)$$

$$f(X_2) \geq f(X) + \nabla f(X)^T (X_2 - X)$$

凸函数的判定

证明：“$\Leftarrow$”

上式分别乘以λ和（1-λ）并相加得

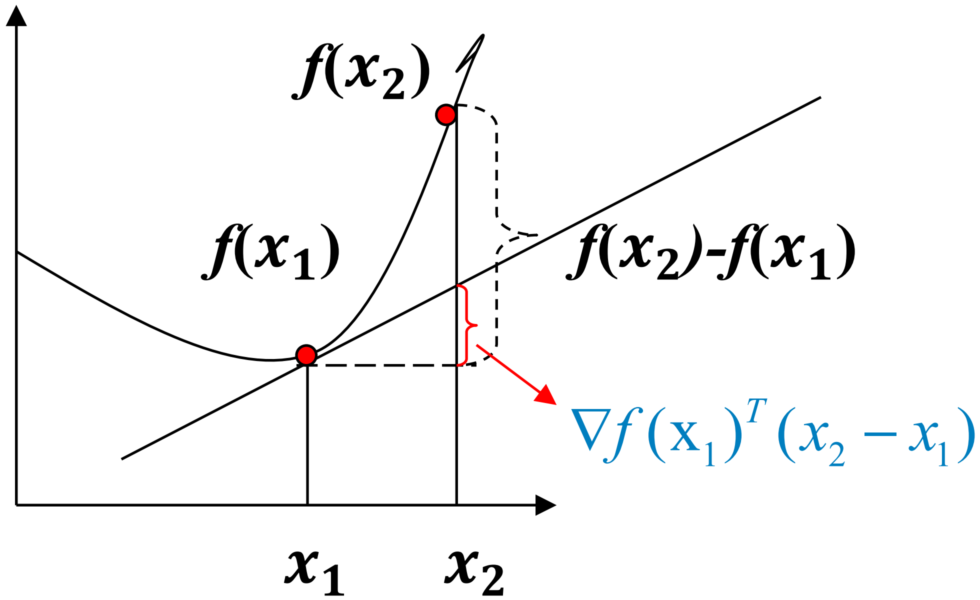
$$\lambda f(X_1)+(1-\lambda)f(X_2)$$

$$\geq f(X)+\nabla f(X)^T(\lambda X_1+(1-\lambda)X_2-X)$$

$$= f(\lambda x^{(1)}+(1-\lambda)x^{(2)})$$

$\Rightarrow f$是凸函数。

物理意义



$f(x)$是凸函数当且仅当任意点处的切线增量不超过函数的增量。

凸函数的判定

定理2.22　设f：$C\subseteq R^{n}\to R^{1}$是二次可微函数，$C$为非空开凸集，则$f$为$C$上凸函数的充要条件是$Hesse$矩阵$\nabla^{2}f(X)$在$C$上任意点均半正定.

证明 "$\Rightarrow$"

设f是C上的凸函数,对任意$\overline{x}\in C$

C是开集，则对$\forall x\in R^{n}$，$\exists\delta>0$使当$\lambda\in(0,\delta)$，有$\overline{x}+\lambda x\in C$。

$$f(\overline{x}+\lambda x)\geq f(\overline{x})+\lambda\nabla f(\overline{x})^{T}x \quad ---(1)$$



凸函数的判定

证明 "$\Rightarrow$"

f在点$\bar{x}\in S$二次可微，

$$f(\bar{x}+\lambda x)=f(\bar{x})+\lambda\nabla f(\bar{x})^T x+\frac{1}{2}\lambda^2 x^T\nabla^2 f(\bar{x})x+o(\lambda^2)\quad ---(2)$$

其中$o(\lambda^2)$是关于λ^2的高阶无穷小。

凸函数的判定

证明 "$\Rightarrow$"

由(1),(2)得， $\frac{1}{2}\lambda^2 x^T \nabla^2 f(\bar{x}) x + o(\lambda^2) \ \geq 0$。

$\frac{1}{2} x^T \nabla^2 f(\bar{x}) x + \frac{1}{\lambda^2} o(\lambda^2) \ \geq 0$。

令$\lambda \to 0$，得 $\frac{1}{2} x^T \nabla^2 f(\bar{x}) x \geq 0$

$\nabla^2 f(\bar{x})$半正定.



凸函数的判定

证明:"$\Leftarrow$"

设$\nabla^2 f(x)$在任意点$x \in C$半正定，对$\forall x, \bar{x} \in C$,

由带$Lagrange$余项的二阶$Taylar$展开式，得

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

其中$\xi = \lambda \bar{x} + (1 - \lambda)x$，$\lambda \in (0, 1)$

凸函数的判定

证明:“$\Leftarrow$”

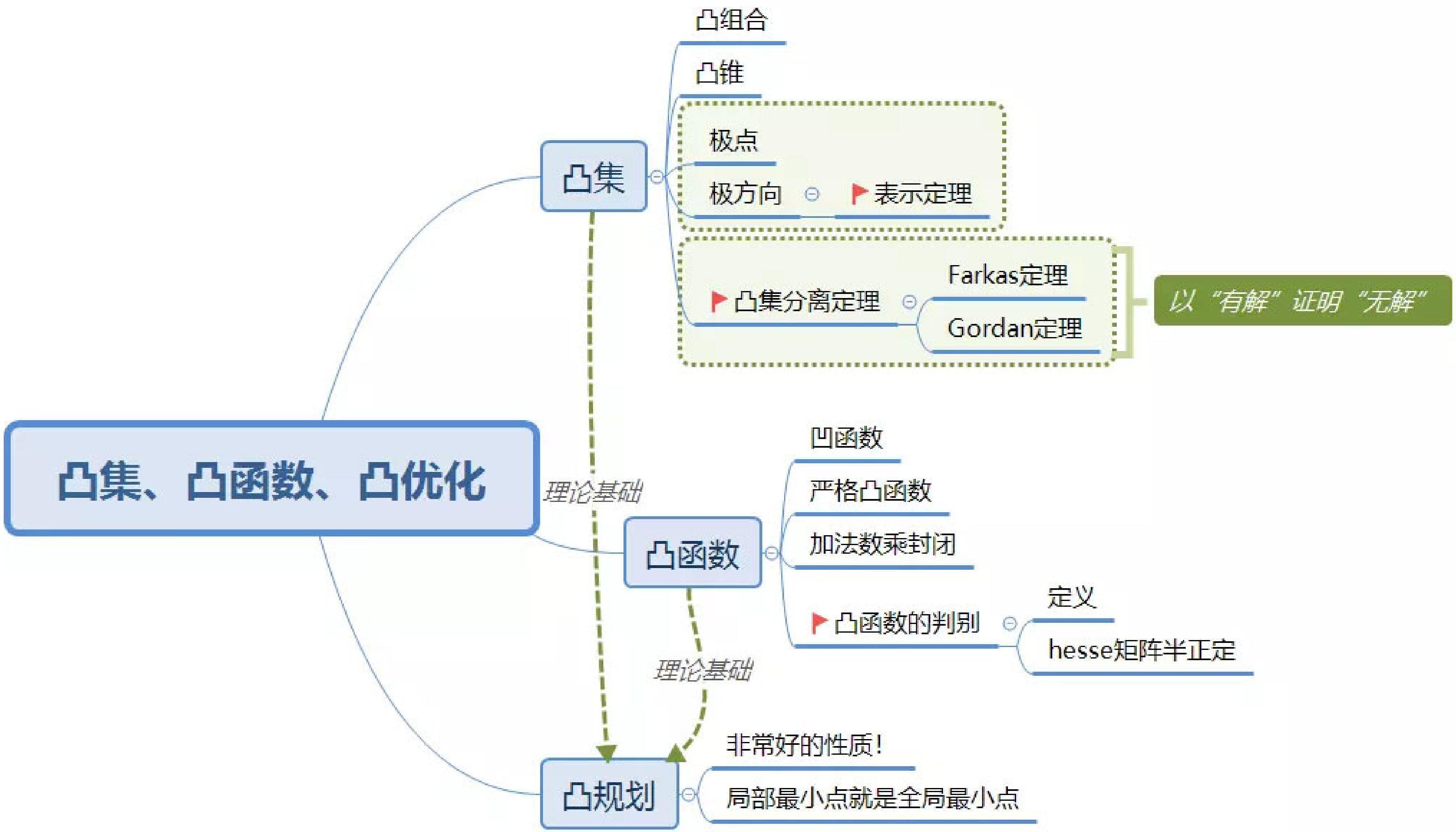
因为C是凸集，所以$\xi \in C$，又$\nabla^2 f(x)$半正定，

$$\frac{1}{2}(x-\overline{x})^T \nabla^2 f(\xi)(x-\overline{x}) \geq 0$$

$$\Rightarrow f(x) \geq f(\overline{x}) + \nabla f(\overline{x})^T (x-\overline{x})$$

f是凸函数。

凸集凸函数凸优化


凸集、凸函数、凸优化
凸集
凸组合
凸锥
极点
极方向
表示定理
凸集分离定理
Farkas定理
Gordan定理
以“有解”证明“无解”
理论基础
凸函数
凹函数
严格凸函数
加法数乘封闭
凸函数的判别
定义
hesse矩阵半正定
理论基础
凸规划
非常好的性质!
局部最小点就是全局最小点

例题

判断函数$f(\mathrm{x})=2x_1^2-x_1x_2+x_2^2$是否为凸函数

解：

$$\nabla \mathrm{f}(\mathrm{x})=\begin{bmatrix}4x_1-x_2\\-x_1+2x_2\end{bmatrix},H(\mathrm{x})=\begin{bmatrix}4 & -1\\-1 & 2\end{bmatrix}$$

$$\det(\mathrm{H}_1)=4\geq 0$$

$$\det(\mathrm{H}_2)=4\times 2-(-1)\times(-1)=7\geq 0$$

$H(\mathrm{x})$半正定

$f(\mathrm{x})$是凸函数

课后练习 #4

证明：凸函数的局部最小值也是全局最小值

谢谢！