第七讲：

凸规划与鞍点

凸集(convex set)

定义：设x,y为欧式空间E^n中相异的两个点，则点集

$$P=\{\lambda x+(1-\lambda)y \mid \lambda \in R\}$$

称为通过x和y的直线.

定义：设$S \subseteq E^n$,若对$\forall\ x^{(1)},x^{(2)} \in S$及$\forall\ \lambda \in [0,1]$,都有

$$\lambda x^{(1)}+(1-\lambda)x^{(2)} \in S$$

则称S为凸集.

设$x^{(1)},x^{(2)}$，…,$x^{(k)} \in S$,称

$$\lambda_1 x^{(1)}+\lambda_2 x^{(2)}+\ldots+\lambda_k x^{(k)}$$

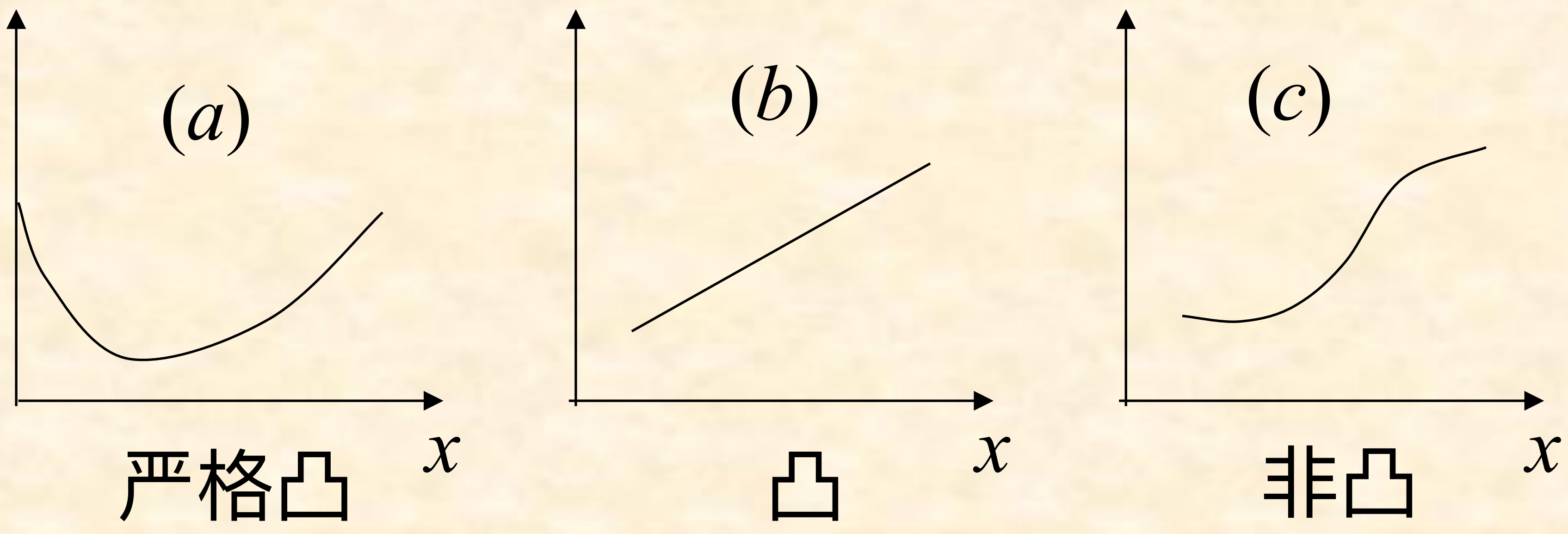
(其中$\lambda_1+\lambda_2+\ldots+\lambda_k=1$)为$x^{(1)},x^{(2)},\ldots,x^{(k)}$的凸组合.

凸函数

凸函数：设S是E^n中的非空凸集，$f(x)$是定义在S上的实函数，如果对于每一对x_1，$x_2 \in S$及每一个a，$0 \le a \le 1$,都有

$$f(ax_1+(1-a)x_2) \le a\,f(x_1)+(1-a)f(x_2)$$

则称函数$f(x)$为S上的凸函数．上式中，若≤变为<，则称为严格凸函数。若-$f(x)$为S的凸函数，则称$f(x)$为S上的凹函数.



北京交通大學

凸函数性质

(1) 设$f_1(x)$，$f_2(x)$是凸集S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S上的凸函数，则对任意的$a\geq 0$，函数$af(x)$是凸的。

推广：设$f_1(x)$，$f_2(x)$，…，$f_k(x)$是凸集S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$S$上的凸函数.

(3) 设$f(x)$是凸集S上的凸函数，对每一个实数c，则集合(level set)

$$S_c=\{x\mid x\in S,\ f(x)\leq c\}\text{是凸集。}$$

(4)设S是E^n中的非空凸集，f是定义在S上的凸函数，则f在S上的局部极小点是整体极小点，且极小点的集合是凸集.

证明:设$\bar{x}$是f在S中的局部极小点，既存在$\bar{x}$的$\varepsilon>0$邻域$N_\varepsilon(\bar{x})$

使得对$\forall x\in S\cap N_\varepsilon(\bar{x})$，有$f(x)\geq f(\bar{x})$。

若$\bar{x}$不是整体极小点，则$\exists x^{(0)}\in S$使$f(\bar{x})>f(x^{(0)})$,

$\because$ S是凸集，$\therefore$ 对$\forall\lambda\in(0,1)$有$\lambda x^{(0)}+(1-\lambda)\bar{x}\in S$,

$\because$ f是S上的凸函数,

$$\therefore\ f(\lambda x^{(0)}+(1-\lambda)\bar{x})\leq\lambda f(x^{(0)})+(1-\lambda)f(\bar{x})$$
$$<\lambda f(\bar{x})+(1-\lambda)f(\bar{x})=f(\bar{x})$$

当λ取得充分小时，可使$\lambda x^{(0)}+(1-\lambda)\bar{x}\in S\cap N_\varepsilon(\bar{x})$,

这与$\bar{x}$为局部极小点矛盾,

$\therefore$ $\bar{x}$是f在S上的整体极小点。

f在S上的极小值也是它在S上的最小值。

极小点集合为$\Gamma_\alpha=\{x\mid x\in S,f(x)\leq\alpha\}$,

则由性质（3），Γ_α为凸集。

凸函数的判别

梯度： $$\nabla f(x)=\left(\frac{\partial f(x)}{\partial x_1},\cdots,\frac{\partial f(x)}{\partial x_n}\right)^T$$

Hesse矩阵：

$$\nabla^2 f(x)=\begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \cdots\cdots\cdots\cdots & & \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

$$\boxed{\begin{aligned} f(x)&=\frac{1}{2}x^T Ax \\ &\quad +b^T x+c \\ \nabla f(x)&=Ax+b \\ \nabla^2 f(x)&=A \end{aligned}}$$

凸性

凸函数的充要条件

定理（一阶充要条件） 设S是E^n中非空开凸集，$f(x)$是定义在S上的可微函数，则$f(x)$为凸函数的充要条件是对任意两点$x^{(1)}, x^{(2)} \in S$，有

$$f(x^{(2)}) \geq f(x^{(1)}) + \nabla f(x^{(1)})^T (x^{(2)} - x^{(1)});$$

$f(x)$为严格凸函数的充要条件是对任意互不相同两点$x^{(1)}, x^{(2)} \in S$，有

$$f(x^{(2)}) > f(x^{(1)}) + \nabla f(x^{(1)})^T (x^{(2)} - x^{(1)}).$$

证 明

"$\Rightarrow$" 设f是S上的凸函数，则对$\forall x^{(1)}, x^{(2)} \in S$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$$\frac{f(x^{(1)} + \lambda(x^{(2)} - x^{(1)})) - f(x^{(1)})}{\lambda} \le f(x^{(2)}) - f(x^{(1)})$$

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

$$Df(x^{(1)}; x^{(2)} - x^{(1)}) = \nabla f(x^{(1)})^T (x^{(2)} - x^{(1)}) \le f(x^{(2)}) - f(x^{(1)}),$$

$$\Rightarrow f(x^{(2)}) \ge f(x^{(1)}) + \nabla f(x^{(1)})^T (x^{(2)} - x^{(1)}).$$

"$\Rightarrow$"当 f是S上的严格凸函数时，对$\forall x^{(1)}, x^{(2)} \in S$及$x^{(1)} \neq x^{(2)}$,

取$\lambda = \frac{1}{2}$，则$y = \frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} \in S$且

$$f(y) = f\left(\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)}\right) < \frac{1}{2}f(x^{(1)}) + \frac{1}{2}f(x^{(2)}).$$

$\because$ f是凸函数，

$$\therefore \quad f(y) \geq f(x^{(1)}) + \nabla f(x^{(1)})^T (y - x^{(1)}).$$

$$\therefore \quad \frac{1}{2}f(x^{(1)}) + \frac{1}{2}f(x^{(2)}) > f(x^{(1)}) + \nabla f(x^{(1)})^T (y - x^{(1)})$$

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

$$\therefore \quad f(x^{(2)}) > f(x^{(1)}) + \nabla f(x^{(1)})^T (x^{(2)} - x^{(1)}).$$

"$\Leftarrow$" 设对$\forall x^{(1)}, x^{(2)} \in S$，有

$$f(x^{(2)}) \ge f(x^{(1)}) + \nabla f(x^{(1)})(x^{(2)} - x^{(1)})。$$

$\forall \lambda \in (0,1)$，令$y = \lambda x^{(1)} + (1-\lambda)x^{(2)}$，则$y \in S$。

由假设，对$x^{(1)}, y \in S$及$x^{(2)}, y \in S$有

$$f(x^{(1)}) \ge f(y) + \nabla f(y)^T (x^{(1)} - y)$$

$$f(x^{(2)}) \ge f(y) + \nabla f(y)^T (x^{(2)} - y)$$

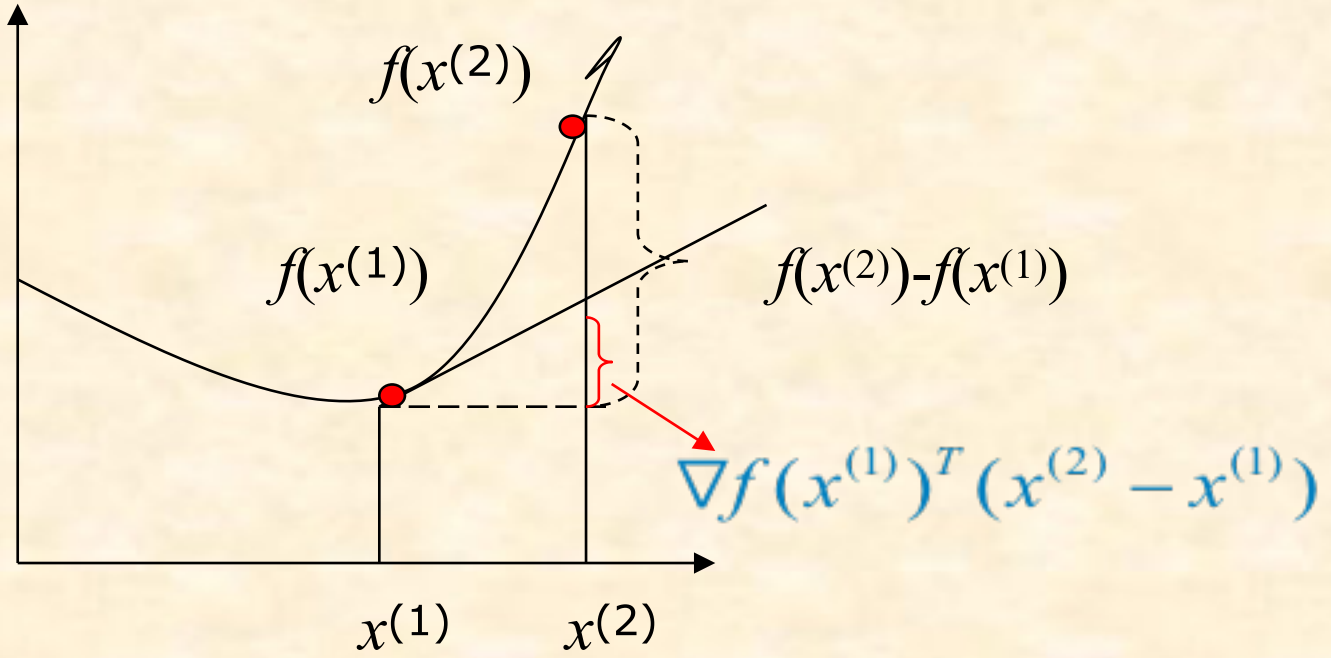
$$\therefore \quad \lambda f(x^{(1)}) + (1-\lambda) f(x^{(2)})$$

$$\ge f(y) + \nabla f(y)^T (\lambda x^{(1)} + (1-\lambda)x^{(2)} - y)$$

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

$\therefore$ f是凸函数。

几何意义



$f(x)$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。

推论：设S是E^n中的凸集，$\bar{x}\in S, f(x)$是定义在E^n上的凸函数，且在点$\bar{x}$可微，则对任意$x\in S$，有

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

定理(二阶充要条件)：设S是E^n中非空开凸集，$f(x)$是定义在S上的二次可微函数，则$f(x)$为凸函数的充要条件是对任意$x \in S$，$f(x)$在x处的$Hessian$矩阵$\nabla^2 f(x)$是半正定的。

证明： "$\Rightarrow$"设f是S上的凸函数, 对任意$\bar{x}\in S$

$\because S$是开集，则对$\forall x\in E^n$，$\exists\delta>0$使当$\lambda\in(0,\delta)$，有$\bar{x}+\lambda x\in S$。

$$\therefore f(\bar{x}+\lambda x)\geq f(\bar{x})+\lambda\nabla f(\bar{x})^T x \qquad (1)$$

$\because f$在点$\bar{x}\in S$二次可微,

$$\therefore f(\bar{x}+\lambda x)=f(\bar{x})+\lambda\nabla f(\bar{x})^T x+\frac{1}{2}\lambda^2 x^T\nabla^2 f(\bar{x})x+\lambda^2\|x\|^2 a \quad (2)$$

其中$\lim\limits_{\lambda\to 0} a=0$。

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

两边除以λ^2后，令$\lambda\to 0$，得，$x^T\nabla^2 f(\bar{x})x\geq 0$。

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

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

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

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

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

$\therefore f$是凸函数。

定理：设S是E^n中非空开凸集，$f(x)$是定义在S上的二次可微函数，如果对任意$x \in S$，有f在x处的$Hessian$矩阵$\nabla^2 f(x)$正定，则$f(x)$为严格凸函数。

对二次函数$f(x) = \dfrac{1}{2}x^T Ax + bx + c$,

若A正定，则$f(x)$为严格凸函数；

若A半正定，则$f(x)$为凸函数。

例：判断下列函数是否为凸函数.

(1) $f(x_1,x_2)=x_1^2-2x_1x_2+x_2^2+x_1+x_2$

(2) $f(x_1,x_2)=x_1e^{-(x_1+x_2)}$.

定理：设$f(x)$是定义在凸集S上的可微凸函数，若$\exists x^* \in S$，使对$\forall x \in S$，都有

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

则x^*是$f(x)$在凸集S上的全局极小点。

证明：对$\forall x \in S$，因为$f(x)$是凸函数，所以有

$$f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*)$$

$$\because \nabla f(x^*)^T (x - x^*) \geq 0$$

$$\therefore f(x) \geq f(x^*)$$

$\Rightarrow$ x^*为全局极小点。

凸规划

- **凸规划：求凸函数在凸集上的极小点。**

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

若$f(x)$是凸函数，$g_i(x)(i=1,\cdots,m)$是凹函数，$h_j(x)(j=1,\cdots,l)$是线性函数，则原问题为凸规划。

性质：凸规划的局部极小点就是整体极小点，
且极小点的集合为凸集。

凸规划的最优性条件

凸规划

考虑

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & x \in \Omega \end{aligned} \qquad (7.4.1)$$

$f: \mathbb{R}^n \to \mathbb{R}$为凸函数，$\Omega$为非空凸集，则称(7.4.1)为凸规划. 若$f(x)$严格凸，则称其为严格凸规划. 本节，我们考虑$f(x)$连续可微，且$\Omega$为非空闭凸集情形.

常见凸规划

线性规划

$$
\begin{aligned}
\min \quad & c^T x \\
s.t. \quad & Ax = b \\
& x \geq 0
\end{aligned}
$$

Lasso

$$
\begin{aligned}
\min \quad & \frac{1}{2} ||Ax - b||^2 \\
s.t. \quad & ||x||_1 \leq s
\end{aligned}
$$

凸二次规划

$$
\begin{aligned}
\min \quad & \frac{1}{2} x^T Q x + g^T x & & Q\text{半正定} \\
s.t. \quad & a_i^T x = b_i, & & i \in \mathcal{E} \\
& a_i^T x \geq b_i, & & i \in I
\end{aligned}
$$

凸规划的性质

性质1 凸规划问题的任一局部最优解为全局最优解.

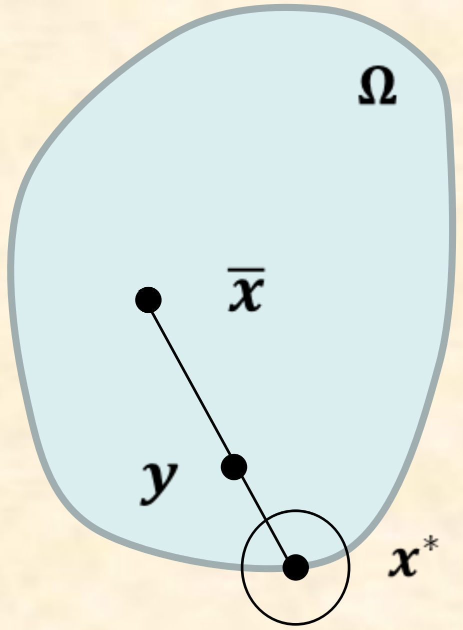
证明：设x^*为凸规划局部最优解，则存在$\delta > 0$，对任意$x \in N(x^*, \delta) \cap \Omega$都有$f(x^*) \le f(x)$.

反证法. 假设x^*不是全局最优解，则存在$\bar{x} \in \Omega$，使得$f(\bar{x}) < f(x^*)$.

考虑线段$[\bar{x}, x^*]$上一点$y := \lambda\bar{x} + (1-\lambda)x^*$，$\lambda \in (0,1)$，由凸函数定义

$$f(y) \le \lambda f(\bar{x}) + (1-\lambda)f(x^*)$$
$$< \lambda f(x^*) + (1-\lambda)f(x^*) = f(x^*)$$

当λ充分小时，$y \in N(x^*, \delta) \cap \Omega$ 且$f(y) < f(x^*)$，这与x^*为局部最优解矛盾. 证毕



凸规划的性质

性质2 凸规划问题（7.4.1）的解集为凸集.

证明： 设最优解集为S，若$S=\emptyset$，结论显然.若$S\neq\emptyset$，设$x^*\in S$，最优解集可表示为

$$S := \{x\in\Omega \mid f(x)\le f(x^*)\} = \Omega\cap\{x\in\mathbb{R}^n \mid f(x)\le f(x^*)\}$$

由于Ω为凸集，凸函数的水平集$\mathcal{L}(x^*)=\{x\in\mathbb{R}^n \mid f(x)\le f(x^*)\}$为凸集，根据两个凸集交集仍为凸集可知

$$S=\Omega\cap\{x\in\mathbb{R}^n \mid f(x)\le f(x^*)\}$$

为凸集. 证毕

凸规划的性质

性质3 问题（7.4.1）为严格凸规划问题时，若存在最优解，则最优解唯一.

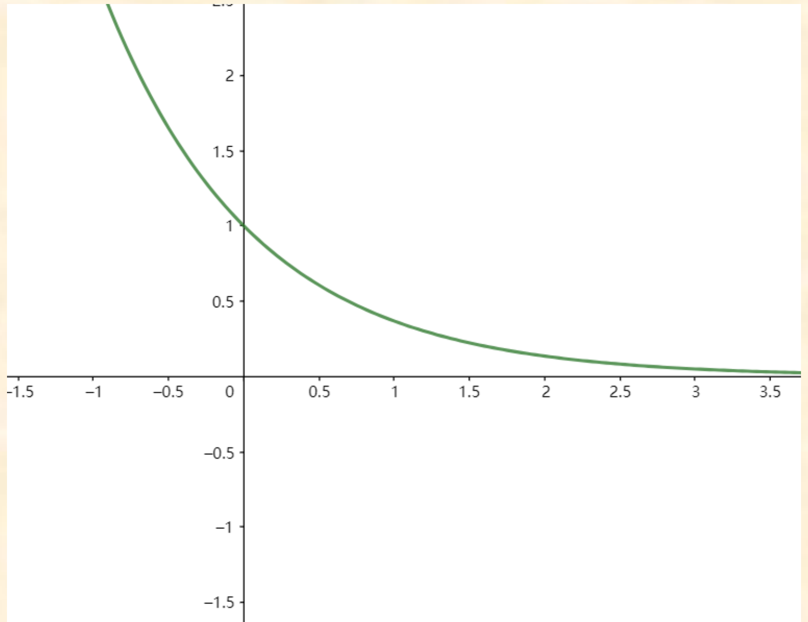
证明： 假设x^*, y^*均为（7.4.1）最优解，且$x^* \neq y^*$

取$z^* = \frac{1}{2}(x^* + y^*)$，则$z \in \Omega$.根据严格凸函数的定义有

$$f(z^*) < \frac{1}{2}f(x^*) + \frac{1}{2}f(y^*) = f(x^*)$$

与x^*为最优解矛盾. **证毕**

严格凸规划并不一定存在最优解.



$$\min_{x \in \mathbb{R}} e^{-x}$$

$$(e^{-x})'' = e^{-x}$$

凸规划的性质

性质4 凸规划问题（7.4.1）的全局最优解与稳定点等价.

证明： 设x^*为凸规划全局最优解，由定理7.1.2， x^*为稳定点

反之若x^*为稳定点，即对任意$x \in \Omega$,

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0.$$

利用凸函数的性质得

$$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle \geq f(x^*),$$

故x^*为全局最优点.　**证毕**

定理 7.1.2 若约束优化问题的可行域为非空闭凸集，则其任一局部最优解为其稳定点。

稳定点 ⟺ 全局最优解 ⟺ 局部最优解

等式与不等式约束下的凸规划

带有等式和不等式约束的凸规划可表示成如下形式

$$\begin{aligned} min \quad & f(x) \\ s.t. \quad & c_i(x) = 0, \quad i \in \mathcal{E}, \\ & c_i(x) \geq 0, \quad i \in I \end{aligned} \qquad (7.4.2)$$

$$i \in I, \Omega_i := \{x \in \mathbb{R}^n | c_i(x) \geq 0\}$$
$$= \{x \in \mathbb{R}^n | -c_i(x) \leq 0\} = \mathcal{L}_{-c_i}(0)$$

$-c_i(x)$为凸函数 $\Rightarrow$ Ω_i为凸集

$$i \in \mathcal{E}, \Omega_i := \{x \in \mathbb{R}^n | c_i(x) = 0\}$$
$$= \{x \in \mathbb{R}^n | -c_i(x) \leq 0\} \cap \{x \in \mathbb{R}^n | c_i(x) \leq 0\}$$
$$= \mathcal{L}_{c_i}(0) \cap \mathcal{L}_{-c_i}(0)$$

$-c_i(x), c_i(x)$为凸函数 $\Rightarrow$ Ω_i为凸集

约束集

$$\Omega = \bigcap_{i \in \mathcal{E} \cup I} \Omega_i$$

为凸集

- 凸规划的等式约束为线性约束，不等式约束函数为凹函数.

Slater约束规格

定义7.4.1 对凸规划问题（7.4.2），若存在可行点$\bar{x}$使得

$$c_i(\bar{x}) > 0, \qquad i \in I$$

则称该规划问题满足Slater约束规格，又称Slater条件.

- 可行点$\bar{x}$是不等式约束集的严格内点.
- 针对凸规划(7.4.2)，利用Slater约束规格，可建立KKT条件.

Slater约束规格

定理7.4.1 对凸规划问题（7.4.2），若Slater约束规格成立，则最优值点为KKT点.

仅证：Slater约束规格 $\Longrightarrow$ M–F约束规格

即证：

存在可行点$\bar{x}$使得$c_i(\bar{x}) > 0$， $i \in I$

存在可行点$\bar{x}$使得$\nabla c_i(\bar{x}), i \in \mathcal{E}$线性无关，且存在非零向量

$s \in R^n$使得$s^T \nabla c_i(\bar{x}) = 0, i \in \mathcal{E}, s^T \nabla c_i(\bar{x}) > 0, i \in I(\bar{x})$

局部极小+M–F约束规格 $\xrightarrow{\text{定理7.2.2}}$ KKT点

Slater与M-F约束规格

引理 7.4.1 对凸规划问题（7.4.2），若Slater约束规格成立，则对任意$x \in \Omega$，必存在非零向量$s \in R^n$使得

$$s^T \nabla c_i(x) = 0, i \in \mathcal{E}, s^T \nabla c_i(x) > 0, i \in I(x).$$

证明 由Slater约束规格，存在$\bar{\boldsymbol{x}} \in \Omega$使

$$c_i(\bar{\boldsymbol{x}}) = 0,\ i \in \mathcal{E};\quad c_i(\bar{\boldsymbol{x}}) > 0,\ i \in \mathcal{I}.$$

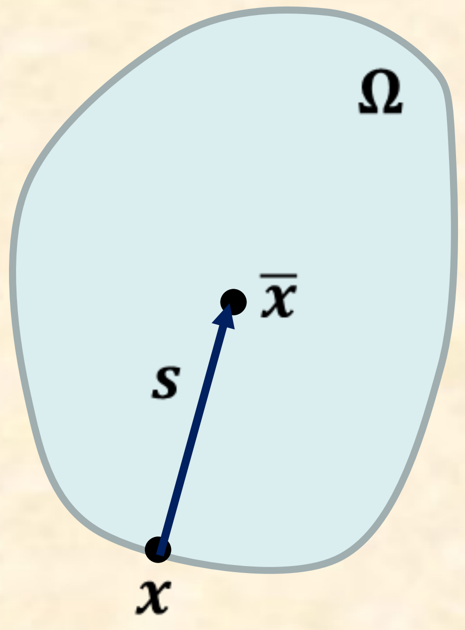
对任意的$\boldsymbol{x} \in \Omega$，令$\boldsymbol{s} = \bar{\boldsymbol{x}} - \boldsymbol{x}$. 对$i \in \mathcal{E}$，由于$c_i(\boldsymbol{x})$为线性函数，所以

$$0 = c_i(\bar{\boldsymbol{x}}) = c_i(\boldsymbol{x} + \boldsymbol{s}) = c_i(\boldsymbol{x}) + \boldsymbol{s}^{\mathrm{T}} \nabla c_i(\boldsymbol{x}) = \boldsymbol{s}^{\mathrm{T}} \nabla c_i(\boldsymbol{x}).$$

对$i \in \mathcal{I}(\boldsymbol{x})$，由于$c_i(\boldsymbol{x})$为凹函数，所以

$$0 < c_i(\bar{\boldsymbol{x}}) = c_i(\boldsymbol{x} + s) \leqslant c_i(\boldsymbol{x}) + \boldsymbol{s}^{\mathrm{T}} \nabla c_i(\boldsymbol{x}) = \boldsymbol{s}^{\mathrm{T}} \nabla c_i(\boldsymbol{x}),$$

即$\boldsymbol{s}$满足命题结论.



弱Slater约束规格

定义7.4.2 对凸规划问题

$$
\begin{aligned}
\min \quad & f(x) \\
s.t. \quad & c_i(x) = 0, \quad i \in \mathcal{E}, \\
& c_i(x) \geq 0, \quad i \in I_1 \cup I_2
\end{aligned} \tag{7.4.3}
$$

其中，$f:\mathbb{R}^n \to \mathbb{R}$为连续可微凸函数，$c_i(x), i \in \mathcal{E} \cup I_1$为线性函数，$c_i(x), i \in I_2$为非线性连续可微凹函数，若存在$\bar{x} \in \Omega$使得

$$c_i(\bar{x}) > 0, i \in I_2,$$

则称该约束优化问题满足弱Slater约束规格.

弱Slater约束规格

定理7.4.2 对凸规划问题(7.4.3)，如果弱Slater约束规格成立，则最优值点为KKT点.

证明：

推论1.4.3 设向量组$\boldsymbol{a}_i \in \mathbb{R}^n$，$i \in \mathcal{E} \cup \mathcal{I}_1 \cup \mathcal{I}_2$和向量$\boldsymbol{b} \in \mathbb{R}^n$满足

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} = 0,\ i \in \mathcal{E};\ \ \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} \geqslant 0,\ i \in \mathcal{I}_1;\ \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} > 0,\ i \in \mathcal{I}_2\}$$

非空. 则集合

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} = 0,\ i \in \mathcal{E};\ \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} \geqslant 0, i \in \mathcal{I}_1;\ \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} > 0,\ i \in \mathcal{I}_2;\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} < 0\} = \varnothing$$

的充分必要条件是存在满足$\lambda_i \geqslant 0, i \in \mathcal{I}_1 \cup \mathcal{I}_2$的数组$\boldsymbol{\lambda}$ 使得

$$\boldsymbol{b} = \sum_{i \in \mathcal{E} \cup \mathcal{I}_1 \cup \mathcal{I}_2} \lambda_i \boldsymbol{a}_i.$$

x^*为凸规划问题（7.4.3）的最优值点

⬇ 仿照引理7.4.1

$$\exists s \in \mathbb{R}^n, s^T\nabla c_i(x^*) = 0, i \in \mathcal{E};\ \ s^T\nabla c_i(x^*) \geq 0, i \in I_1(x^*);\ \ s^T\nabla c_i(x^*) > 0, i \in I_2(x^*)$$

⟶ s为可行方向

⬇

$$\{s \in \mathbb{R}^n | s^T\nabla c_i(x^*) = 0, i \in \mathcal{E};\ \ s^T\nabla c_i(x^*) \geq 0, i \in I_1(x^*);\ \ s^T\nabla c_i(x^*) > 0, i \in I_2(x^*);\ \ s^T\nabla f(x^*) < 0\} = \emptyset$$

⬇ 推论1.4.3

x^*为KKT点

常见约束规格

- 约束均为线性函数

$$c_i(x) = a_i^T x + b_i$$

- 线性无关

$\nabla c_i(x^*), i \in \mathcal{E}$ 线性无关

- M–F约束规格

$\nabla c_i(x^*), i \in \mathcal{E}$ 线性无关，且存在非零向量 $s \in R^n$ 使得

$$s^T \nabla c_i(x^*) = 0, i \in \mathcal{E}, s^T \nabla c_i(x^*) > 0, i \in I(x^*)$$

- Slater约束规格

$$\exists \hat{x}, s.t. \begin{cases} c_i(\hat{x}) = 0, i \in \mathcal{E} \\ c_i(\hat{x}) > 0, i \in I \end{cases}$$

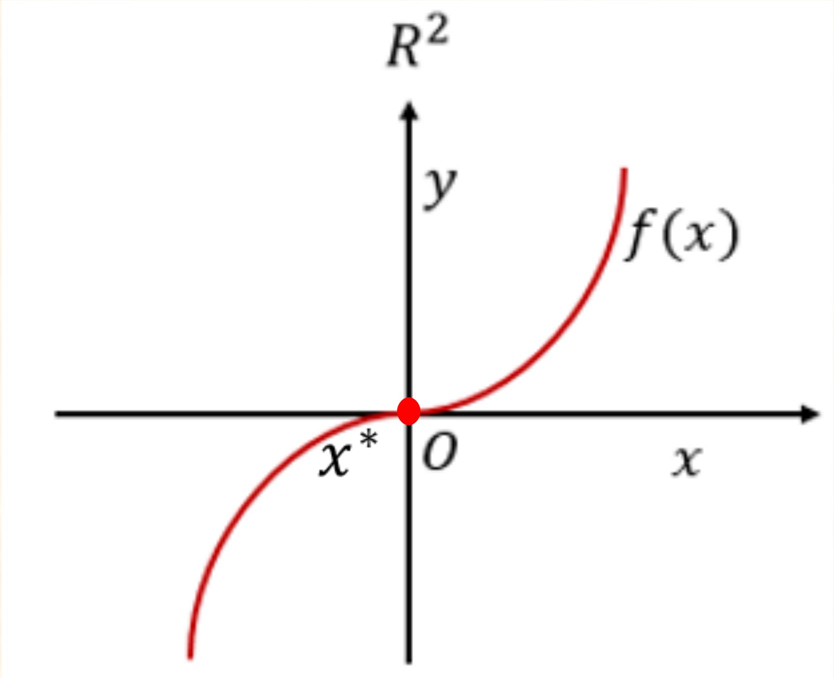
- 弱Slater约束规格

$$\exists \hat{x}, s.t. \begin{cases} c_i(\hat{x}) = 0, i \in \mathcal{E} \\ c_i(\hat{x}) \geq 0, i \in I_1 \\ c_i(\hat{x}) > 0, i \in I_2 \end{cases}$$

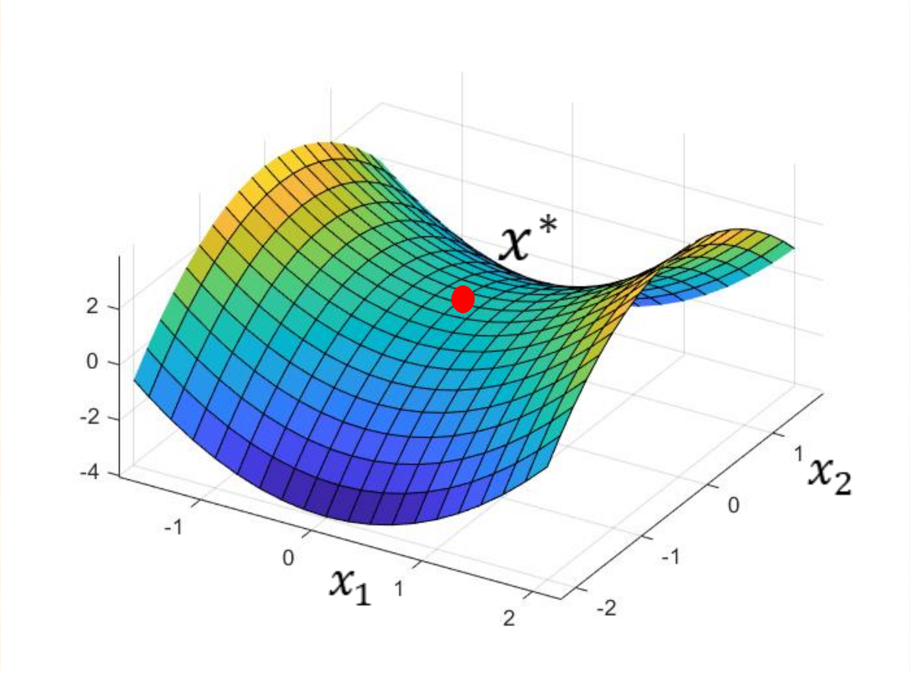
Lagrange函数的鞍点

回顾：鞍点（saddle point）

定义： 对于无约束优化问题 $\min_{x\in\mathbb{R}^n} f(x)$，若 $\nabla f(x^*) = 0$，则 x^* 称为 $f(x)$ 的一个驻点（稳定点）. 既不是极小值点，也不是极大值点的驻点称为鞍点. **但是，…**



$f(x) = x^3$



$f(x) = x_1^2 - x_2^2$

回顾：等式约束优化问题最优性条件

在正则性条件下，等式约束优化问题的一阶最优性条件可以用Lagrange函数的梯度描述.

$$\begin{cases} \nabla_x L(x^*,\lambda^*) = 0, \\ \nabla_\lambda L(x^*,\lambda^*) = c(x^*) = 0. \end{cases}$$

(x^*,λ^*)构成Lagrange函数的稳定点

其中x^*为最优解，λ^*为最优Lagrange乘子.

(x^*,λ^*)是哪种类型稳定点?

局部最优?

鞍点?

约束优化与Lagrange函数

考虑约束优化问题

$$\min\{f(x)|c_i(x)=0, i\in\mathcal{E}; c_i(x)\geq 0, i\in I\} \qquad (7.2.2)$$

则相应的Lagrange函数为

$$L(x,\lambda,\mu)=f(x)-\sum_{i\in\mathcal{E}}\lambda_i c_i(x)-\sum_{i\in I}\mu_i c_i(x) \qquad \forall\lambda\in\mathbb{R}^{|\mathcal{E}|}, \forall\mu\in\mathbb{R}_+^{|I|}, \forall x\in\mathbb{R}^n$$

Lagrange函数的鞍点

定义7.3.1 对约束优化问题

$$\min\{f(x)|c_i(x)=0, i\in\mathcal{E}; c_i(x)\geq 0, i\in I\} \qquad (7.2.2)$$

若存在$x^*\in\mathbb{R}^n, \lambda^*\in\mathbb{R}^{|\mathcal{E}|}$, 以及$\mu^*\in\mathbb{R}_+^{|I|}, (\mu_i^*\geq 0, i\in I)$满足

$$L(x^*,\lambda,\mu)\leq L(x^*,\lambda^*,\mu^*)\leq L(x,\lambda^*,\mu^*), \quad \forall\mu^*\in\mathbb{R}_+^{|I|}, \forall\lambda\in\mathbb{R}^{|\mathcal{E}|}, \forall x\in\mathbb{R}^n$$

则(x^*,λ^*,μ^*)称为该约束优化问题Lagrange函数的鞍点.

- $L(x^*,\lambda^*,\mu^*)\leq L(x,\lambda^*,\mu^*), \forall x\in\mathbb{R}^n$
- $L(x^*,\lambda,\mu)\leq L(x^*,\lambda^*,\mu^*), \forall\lambda\in\mathbb{R}^{|\mathcal{E}|}, \boxed{\mu_i\geq 0, i\in I}$

$$x^*\in\underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\ L(x,\lambda^*,\mu^*)$$

$$(\lambda^*,\mu^*)\in\underset{\substack{\lambda\in\mathbb{R}^{|\mathcal{E}|}\\ \mu\in\mathbb{R}_+^{|I|}}}{\operatorname{argmax}} L(x^*,\lambda,\mu)$$

对于不等式约束的Lagrange乘子有非负约束

鞍点与KKT点对

定理7.3.1 设(x^*,λ^*,μ^*)为约束优化问题(7.2.2)的Lagrange函数的鞍点，则(x^*,λ^*,μ^*)是该约束优化问题的一个KKT点对.

鞍点

$$x^* \in \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\ L(x,\lambda^*,\mu^*)$$

$$(\lambda^*,\mu^*) \in \underset{\substack{\lambda\in\mathbb{R}^{|\mathcal{E}|}\\ \mu\in\mathbb{R}_+^{|\mathcal{I}|}}}{\operatorname{argmax}}\ L(x^*,\lambda,\mu)$$

$$\Rightarrow$$

KKT对

$$\nabla_x L(x^*,\lambda^*,\mu^*) = 0,$$

$$\mu_i^* \geq 0, c_i(x^*) \geq 0, \mu_i^* c_i(x^*) = 0, \qquad \forall i \in \mathcal{I},$$

$$c_i(x^*) = 0, \qquad \forall i \in \mathcal{E}$$

鞍点与KKT点对

证明：

- $x^* \in \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\ L(x,\lambda^*,\mu^*)$ $\Rightarrow$ $\nabla_x L(x^*,\lambda^*,\mu^*) = 0,$

- $\lambda^* \in \underset{\lambda\in\mathbb{R}^{|\mathcal{E}|}}{\operatorname{argmax}}\ L(x^*,\lambda,\mu^*)$ $\Rightarrow$ $\nabla_\lambda L(x^*,\lambda^*,\mu^*) = -c_i(x^*) = 0, \forall i \in \mathcal{E}$ $\Leftrightarrow$ $c_i(x^*) = 0, \quad \forall i \in \mathcal{E}$

- $\mu^* \in \underset{\mu\in\mathbb{R}_+^{|I|}}{\operatorname{argmax}}\ L(x^*,\lambda^*,\mu)$

$\Updownarrow$

$\mu^* \in \underset{\mu\in\mathbb{R}_+^{|I|}}{\operatorname{argmin}}\ -L(x^*,\lambda^*,\mu)$

$\Rightarrow$

$$\boxed{\begin{array}{l} -\nabla_\mu L(x^*,\lambda^*,\mu^*) - \gamma = 0 \\ \gamma_i \geq 0, \mu_i^* \geq 0, \gamma_i \mu_i^* = 0, \qquad i \in I \end{array}} \Rightarrow \gamma_i = c_i(x^*), \qquad i \in I$$

$\Updownarrow$

$$\mu_i^* \geq 0, c_i(x^*) \geq 0, \mu_i^* c_i(x^*) = 0, \qquad \forall i \in I.$$

鞍点与全局最优解

定理7.3.2 设(x^*,λ^*,μ^*)为约束优化问题(7.2.2)的Lagrange函数的鞍点，则x^*是该约束优化问题的全局最优解.

证明： 由定理7.3.1，x^*为优化问题(7.2.2)的K-T点，利用鞍点定义，对任意$x \in \Omega$，

$$L(x^*,\lambda^*,\mu^*) \le L(x,\lambda^*,\mu^*)$$

即

$$f(x^*) - \sum_{i\in\mathcal{E}} \underbrace{\lambda_i^* c_i(x^*)}_{=0} - \sum_{i\in I} \underbrace{\mu_i^* c_i(x^*)}_{=0} \le f(x) - \sum_{i\in\mathcal{E}} \underbrace{\lambda_i^* c_i(x)}_{=0} - \sum_{i\in I} \mu_i^* c_i(x)$$

KKT条件互补性

$\boldsymbol{x \in \Omega, c_i(x) = 0, \forall i \in \mathcal{E}}$

$$f(x^*) \le f(x) - \sum_{i\in I} \mu_i^* c_i(x) \le f(x)$$

$\boldsymbol{c_i(x) \ge 0, \mu_i^* \ge 0}$

从而x^*为约束优化问题(7.2.2)全局最优解.

鞍点的存在性

例： 考虑非线性规划 $\min\limits_{x\in\mathbb{R}} \{ x^3 | -x^2 \geq 0\}$

该问题最优解$x^* = 0$.相应的Lagrange函数为

$$L(x,\lambda) = x^3 + \lambda x^2$$

根据鞍点定义，要寻找$\lambda^* \geq 0$对任意$x \in \mathbb{R}, \lambda \geq 0$都有

$$L(x^*,\lambda) \leq L(x^*,\lambda^*) \leq L(x,\lambda^*),$$

即满足

$${x^*}^3 + \lambda {x^*}^2 \leq {x^*}^3 + \lambda^* {x^*}^2 \leq x^3 + \lambda^* x^2,$$

等价于满足

$$x^3 + \lambda^* x^2 \geq 0, \forall x \in \mathbb{R}$$

可以验证任意给定$\lambda^* \geq 0$，均不满足上式. 因为若$\lambda^* = 0$，则$x = -1$时，上式不成立；

若$\lambda^* > 0$，则$x = -2\lambda^*$时，上式不成立. 因此该优化问题Lagrange函数不存在鞍点.

鞍点要比全局最优解更强，但它不一定存在，并且难以求解，所以通常不希望求解鞍点.

最优性条件关系

K–T点

线性约束
M-F约束规范

鞍点

局部最优值点

Ω为闭凸集

稳定点

定理7.1.2

全局最优值点

凸规划：KKT点对与鞍点

定理7.4.3 若(x^*,λ^*,μ^*)凸规划问题（7.4.2）的KKT点对，则(x^*,λ^*,μ^*)为Lagrange函数的鞍点.

证明：

- $x^* \in \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\, L(x,\lambda^*,\mu^*)$

$$L(x,\lambda^*,\mu^*) = f(x) - \sum_{i\in\mathcal{E}} \lambda_i^* c_i(x) - \sum_{i\in I} \mu_i^* c_i(x)$$ 关于$x \in \mathbb{R}^n$为凸函数

$$\Downarrow$$

$$L(x,\lambda^*,\mu^*) \geq L(x^*,\lambda^*,\mu^*) + (x-x^*)^T \overset{=0}{\nabla_x L(x^*,\lambda^*,\mu^*)} = L(x^*,\lambda^*,\mu^*)$$

$\boxed{(x^*,\lambda^*,\mu^*)\text{为K−T对}}$

- $(\lambda^*,\mu^*) \in \underset{\substack{\lambda\in\mathbb{R}^{|\mathcal{E}|}\\ \mu\in\mathbb{R}_+^{|I|}}}{\operatorname{argmax}}\, L(x^*,\lambda,\mu)$

对任意$\lambda \in \mathbb{R}^{|\mathcal{E}|}, \mu \in \mathbb{R}_+^{|I|}$

$$L(x^*,\lambda,\mu) - L(x^*,\lambda^*,\mu^*) = -\sum_{i\in\mathcal{E}} \lambda_i \overset{=0}{c_i(x^*)} - \sum_{i\in I} \mu_i c_i(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* \overset{=0}{c_i(x^*)} + \sum_{i\in I} \overset{=0}{\mu_i^* c_i(x^*)}$$

$\boxed{\text{互补}}$

$$= -\sum_{i\in I} \mu_i c_i(x^*) \leq 0$$

凸规划：KKT点与全局最优解

对于凸规划问题（7.4.2）

全局最优值点 $\xleftarrow{\text{定理7.3.2}}$ 鞍点 $\underset{\text{定理7.4.3}}{\overset{\text{定理7.3.1}}{\rightleftarrows}}$ KKT点

定理7.4.4 凸规划问题（7.4.2）的KKT点为其全局最优解.

总结：凸规划最优性条件

对于凸规划问题（7.4.2）

全局最优值点 $\Longleftrightarrow$ 局部最优值点 $\overset{\text{弱Slater条件}}{\rightleftarrows}$ KKT点

全局最优值点 $\Updownarrow$ 稳定点

KKT点 $\Updownarrow$ 鞍点