最优化方法 Operations Research

约束优化问题的二阶条件

第八讲

Lagrange对偶

约束优化与Lagrange函数

对于约束优化问题(7.2.2)

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

记由不等式约束组成向量函数为$G(x)$，由等式约束组成的向量值函数记为$H(x)$.

则Lagrange函数可重新写为

$$L(x,u,v) = f(x) - G(x)^T u - H(x)^T v, \qquad x \in \mathbb{R}^n, u \in \mathbb{R}_+^{|I|}, v \in \mathbb{R}^{|\mathcal{E}|}.$$

约束优化与Lagrange函数

$$L(x,u,v) = f(x) - G(x)^T u - H(x)^T v, \qquad x \in \mathbb{R}^n, u \in \mathbb{R}_+^{|I|}, v \in \mathbb{R}^{|\mathcal{E}|}.$$

鞍点 $(u^*, v^*) \in \sup\limits_{u\geq 0, v\in\mathbb{R}^{|\mathcal{E}|}} L(x^*, u, v)$ $\qquad x^* \in \inf\limits_{x\in\mathbb{R}^n} L(x, u^*, v^*)$

$\boxed{x^*\text{改为一般点}x}$

$$\sup_{u\geq 0, v\in\mathbb{R}^{|\mathcal{E}|}} L(x,u,v) = \begin{cases} +\infty, & x \notin \Omega \\ f(x), & x \in \Omega \end{cases}$$

$\Rightarrow$ $\exists c_i(x), i \in \mathcal{E}, c_i(x) \neq 0$或$\exists c_i(x), i \in I, c_i(x) < 0$

取$v_i = -t c_i(x)$，令$t \to +\infty$，
则$-v_i c_i(x) = t c_i(x)^2 \to +\infty$

取$u_i = -t c_i(x) > 0$，令$t \to +\infty$，
则$-u_i c_i(x) = t c_i(x)^2 \to +\infty$

$G(x) \geq 0,$
$H(x) = 0$

$\Leftarrow$

$u = 0$
$H(x)^T v = 0$

约束优化与Lagrange函数

$$\sup_{u\geq 0, v\in\mathbb{R}^{|\mathcal{E}|}} L(x,u,v) = \begin{cases} +\infty, & x\notin\Omega \\ f(x), & x\in\Omega \end{cases}$$

⬇

$$\min_{x\in\mathbb{R}^n}\max_{u\geq 0, v\in\mathbb{R}^{|\mathcal{E}|}} L(x,u,v) = \min\{f(x)\,|\,G(x)\geq 0, H(x)=0\}$$

Lagrange对偶

Lagrange对偶规划

$$\max_{u\geq 0, v\in\mathbb{R}^{|\mathcal{E}|}} \boxed{\min_{x\in\mathbb{R}^n} L(x,u,v)}$$

- 关于x的无约束优化问题
- 关于变量u, v的函数

定义极值函数$\theta(u,v)=\min\{L(x,u,v)\,|\,x\in\mathbb{R}^n\}$，Lagrange对偶规划可改写为

$$\begin{aligned} \max \quad & \theta(u,v) \\ s.t. \quad & u\in\mathbb{R}_+^{|I|}, v\in\mathbb{R}^{|\mathcal{E}|} \end{aligned} \qquad (7.5.1)$$

Lagrange函数关于x和(u,v)的极值未必能达到，故有时将min，max 用inf，sup 替代

Lagrange对偶

定理7.5.1 对偶规划问题目标函数$\theta(u,v)$关于$u\in\mathbb{R}_+^{|\mathcal{I}|}, v\in\mathbb{R}^{|\mathcal{E}|}$为凹函数.

证明： 对任意$u_1,u_2\in\mathbb{R}_+^{|\mathcal{I}|}, v_1,v_2\in\mathbb{R}^{|\mathcal{E}|}$和任意$\lambda\in(0,1)$，利用Lagrange函数$L(x,u,v)$关于乘子$u,v$的线性和最优值函数的性质，有

$$
\begin{aligned}
&\theta(\lambda u_1+(1-\lambda)u_2, \lambda v_1+(1-\lambda)v_2)\\
&=\min_{x\in\mathbb{R}^n} L(x,\lambda u_1+(1-\lambda)u_2,\lambda v_1+(1-\lambda)v_2)\\
&=\min_{x\in\mathbb{R}^n}(\lambda L(x,u_1,v_1)+(1-\lambda)L(x,u_2,v_2))\\
&\geq \min_{x\in\mathbb{R}^n}\lambda L(x,u_1,v_1)+\min_{x\in\mathbb{R}^n}(1-\lambda)L(x,u_2,v_2)\\
&=\lambda\theta(u_1,v_1)+(1-\lambda)\theta(u_2,v_2)
\end{aligned}
$$

Lagrange函数$L(x,u,v)$关于乘子u,v的线性

对偶问题优缺点

原问题

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

对偶问题

$$\begin{aligned} \max \quad & \theta(u, v) \\ s.t. \quad & u \in \mathbb{R}_+^{|I|}, v \in \mathbb{R}^{|\mathcal{E}|} \end{aligned}$$

对偶问题优点

- 约束集结构简单，是部分元素的非负约束，如果原问题约束个数比较少，对偶问题维度较小；
- 目标函数为凹函数，约束集是凸集，可转化为凸规划进行求解.

对偶问题缺点

- $\theta(u, v)$不一定有显式表达式，求解代价大.

弱对偶定理

定理7.5.2(弱对偶定理) 设$x_0, (u_0, v_0)$分别是优化问题(7.2.2)和其对偶问题(7.5.1)的可行解，则$f(x_0) \geq \theta(u_0, v_0)$.

证明： 对原规划任一可行解x_0和对偶规划问题任一可行解(u_0, v_0),

$$\theta(u_0, v_0) = \min\{f(x) - u_0^T G(x) - v_0^T H(x) \mid x \in \mathbb{R}^n\}$$

$$\leq f(x_0) - u_0^T \underbrace{G(x_0)}_{\geq 0} - v_0^T \underbrace{H(x_0)}_{= 0} \leq f(x_0).$$

弱对偶定理的推论

推论1 对于原问题和对偶问题，必有

$$\inf\{f(x)\,|\,G(x)\geq 0, H(x)=0\}\geq \sup\{\theta(u,v)\,|\,u\geq 0\}.$$

推论2 若 $f(\bar{x})\leq\theta(\bar{u},\bar{v})$，其中 $\bar{x}$ 为原问题可行解，$(\bar{u},\bar{v})$ 为对偶问题可行解，则 $\bar{x}$ 和 $(\bar{u},\bar{v})$ 分别为原问题和对偶问题最优解.

推论3 如果 $\inf\{f(x)\,|\,G(x)\geq 0, H(x)=0\}=-\infty$，则对偶问题无可行解.

推论4 如果 $\sup\{\theta(u,v)\,|\,u\geq 0\}=+\infty$，则原问题无可行解.

推论1 $$\min_{x\in\mathbb{R}^n}\max_{u\geq 0, v\in\mathbb{R}^{|\mathcal{E}|}} L(x,u,v) \geq \max_{u\geq 0, v\in\mathbb{R}^{|\mathcal{E}|}}\min_{x\in\mathbb{R}^n} L(x,u,v)$$

一般结论 对定义在集合$\mathbb{X}\times\mathbb{Y}\subset\mathbb{R}^m\times\mathbb{R}^n$上的连续函数$\phi(x,y)$，恒有

$$\min_{x\in\mathbb{X}}\max_{y\in\mathbb{Y}}\phi(x,y)\geq\max_{y\in\mathbb{Y}}\min_{x\in\mathbb{X}}\phi(x,y)$$

- 对偶间隙=原问题最优值−对偶问题最优值≥ 0.
- 完全对偶：对偶间隙为零.
- 强对偶：对偶间隙为零，原问题和对偶问题最优解都存在.

强对偶定理

$$
\begin{aligned}
\min \quad & f(x) \\
s.t. \quad & c_i(x) = 0, \quad i \in \mathcal{E}, \qquad (7.2.2)\\
& c_i(x) \geq 0, \quad i \in I.
\end{aligned}
$$

定理7.5.3 约束优化问题(7.2.2)的Lagrange函数存在鞍点(x^*, u^*, v^*)的充分必要条件是x^*和(u^*, v^*)分别是原规划和对偶规划的最优解，且对偶间隙为零.

强对偶定理证明

定理7.5.3 约束优化问题(7.2.2)的Lagrange函数存在鞍点(x^*,u^*,v^*)的充分必要条件是x^*和(u^*,v^*)分别是原规划和对偶规划的最优解，且对偶间隙为零.

证明： 鞍点 $\Rightarrow$ 强对偶

(x^*,u^*,v^*)为鞍点 $\Rightarrow$ (x^*,u^*,v^*)为K-T对 $\Rightarrow$ x^*和(u^*,v^*)分别是原规划和对偶规划的可行解

鞍点定义 $\Rightarrow$ $$\theta(u^*,v^*)=\min_{x\in\mathbb{R}^n}L(x,u^*,v^*)=L(x^*,u^*,v^*)=\max_{v\in\mathbb{R}^{|\mathcal{E}|},u\in\mathbb{R}_+^{|\mathcal{I}|}}L(x^*,u,v)=f(x^*)$$

对偶间隙为零，根据弱对偶定理推论2，x^*和(u^*,v^*)分别是原规划和对偶规划的最优解.

强对偶定理证明

证明： 强对偶 ➡ 鞍点

反过来，设$\boldsymbol{x}^*$为原规划问题的最优解，则

$$\boldsymbol{G}(\boldsymbol{x}^*) \geqslant \boldsymbol{0}, \quad \boldsymbol{H}(\boldsymbol{x}^*) = \boldsymbol{0}.$$

再根据$(\boldsymbol{u}^*, \boldsymbol{v}^*)$为对偶规划问题的可行解，得

$$\begin{aligned}\theta(\boldsymbol{u}^*, \boldsymbol{v}^*) &= \min\{f(\boldsymbol{x}) - \boldsymbol{G}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{u}^* - \boldsymbol{H}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{v}^* \mid \boldsymbol{x} \in \mathbb{R}^n\} \\ &\leqslant f(\boldsymbol{x}^*) - \boldsymbol{G}(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{u}^* - \boldsymbol{H}(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{v}^* \\ &= L(\boldsymbol{x}^*, \boldsymbol{u}^*, \boldsymbol{v}^*) \leqslant f(\boldsymbol{x}^*).\end{aligned}$$

由于对偶间隙为零，上面的不等式取等号. 从而

$$\theta(\boldsymbol{u}^*, \boldsymbol{v}^*) = L(\boldsymbol{x}^*, \boldsymbol{u}^*, \boldsymbol{v}^*) = f(\boldsymbol{x}^*).$$

结合$\theta(\boldsymbol{u}^*, \boldsymbol{v}^*)$的定义和$f(\boldsymbol{x}^*) = \max\{L(\boldsymbol{x}^*, \boldsymbol{u}, \boldsymbol{v}) \mid \boldsymbol{u} \geqslant \boldsymbol{0}, \ \boldsymbol{v}\}$，对任意的$\boldsymbol{x} \in \mathbb{R}^n$，$\boldsymbol{u} \geqslant \boldsymbol{0}$和$\boldsymbol{v} \in \mathbb{R}^{|\mathcal{E}|}$，

$$\boldsymbol{\theta}(\boldsymbol{u}^*, \boldsymbol{v}^*) = \min_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{u}^*, \boldsymbol{v}^*)$$

$$L(\boldsymbol{x}^*, \boldsymbol{u}, \boldsymbol{v}) \leqslant L(\boldsymbol{x}^*, \boldsymbol{u}^*, \boldsymbol{v}^*) \leqslant L(\boldsymbol{x}, \boldsymbol{u}^*, \boldsymbol{v}^*).$$

充分性得证. 证毕

对偶问题一般结论

定理7.5.5 (von Neumann定理) 设$\mathcal{X},\mathcal{Y}$分别为$\mathbb{R}^n,\mathbb{R}^m$ 中的非空有界闭凸集, $f(\boldsymbol{x},\boldsymbol{y})$为$\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$的连续函数, 且满足

(1) 对任意的$\boldsymbol{y}\in\mathbb{R}^m$, 函数$f(\cdot,\boldsymbol{y})$在$\mathcal{X}$上是凸函数,

(2) 对任意的$\boldsymbol{x}\in\mathbb{R}^n$, 函数$f(\boldsymbol{x},\cdot)$在$\mathcal{Y}$上是凹函数.

则

$$\min_{\boldsymbol{x}\in\mathcal{X}}\max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x},\boldsymbol{y})=\max_{\boldsymbol{y}\in\mathcal{Y}}\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x},\boldsymbol{y}).$$

对偶问题一般结论

定理7.5.6 设$\mathcal{X},\mathcal{Y}$分别为$\mathbb{R}^n,\mathbb{R}^m$中的非空闭凸集，$f(\boldsymbol{x},\boldsymbol{y})$在$\mathcal{X}\times\mathcal{Y}$上连续，且满足

(1) 对任意的$\boldsymbol{y}\in\mathbb{R}^m$，函数$f(\cdot,\boldsymbol{y})$在$\mathcal{X}$上是凸函数，

(2) 对任意的$\boldsymbol{x}\in\mathbb{R}^n$，函数$f(\boldsymbol{x},\cdot)$在$\mathcal{Y}$上是凹函数.

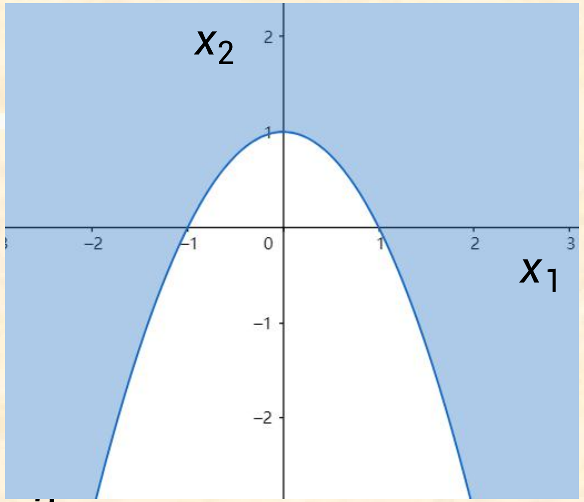
如果$\mathcal{X}$和$\mathcal{Y}$中之一有界，则

$$\min_{\boldsymbol{x}\in\mathcal{X}}\max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x},\boldsymbol{y}) = \max_{\boldsymbol{y}\in\mathcal{Y}}\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x},\boldsymbol{y}).$$

约束优化二阶最优性条件

二阶最优性条件：回顾与思考

- 无约束二阶充分条件：　　$\nabla f(x^*) = 0$，$\nabla^2 f(x^*)$正定
- 有约束二阶充分条件：　　KKT　　？
- 无约束二阶必要条件：　　$\nabla f(x^*) = 0$，$\nabla^2 f(x^*)$半正定
- 有约束二阶必要条件：　　KKT　　？



反例： 考虑如下约束优化问题

$$\min \quad f(x) = x_1^2 + x_2^2$$
$$s.t. \quad x_1^2 + x_2 - 1 \geq 0$$

$$\nabla^2 f(x) = 2I$$

KKT条件

$$\begin{cases} (1-\mu)x_1 = 0 \\ 2x_2 - \mu = 0 \\ \mu \geq 0, x_1^2 + x_2 - 1 \geq 0, \mu(x_1^2 + x_2 - 1) = 0 \end{cases}$$

- $x^* = (0,1)^T$， $\mu = 2$为该问题K-T点对，并且$\nabla^2 f(x^*)$正定.
- 在约束区域边界点$(x_1, 1 - x_1^2)^T$目标函数取值$1 - x_1^2 + x_1^4$.只要$x_1 \in (0,1)$，就有 $f(x) < f(x^*) = 1$.
- $x^* = (0,1)^T$并不是局部最优，$\nabla^2 f(x^*)$未包含约束函数的二阶信息，不能判断x^*为局部最优解.

需要引入$\nabla_{xx} L(x, u, v)$

二阶必要条件

定理7.6.1 设x^*为约束优化问题(7.2.2)

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

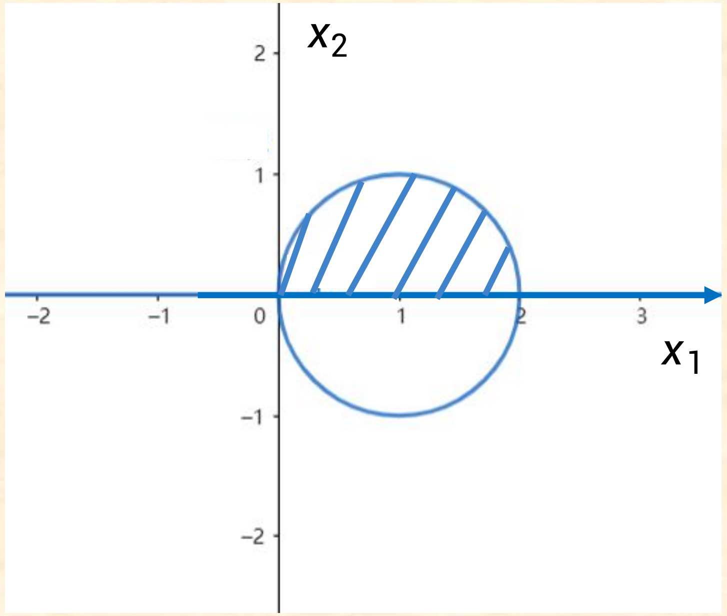
的局部最优解，(x^*,λ^*)为其K-T对.若约束函数全是线性的或者$\nabla c_i(x^*), i\in\mathcal{A}(x^*)$线性无关，则对于满足$s^T\nabla c_i(x^*)=0, i\in\mathcal{A}(x^*)$的任意$s\in\mathbb{R}^n$有 $s^T\nabla_{xx}L(x^*,\lambda^*)s\geq 0.$

例子：二阶必要条件

$$\min \quad x_1$$
$$s.t. \quad x_2 \geq 0,$$
$$1-(x_1-1)^2-x_2^2 \geq 0.$$

解：该问题的最优解为$x^* = (0,0)^T$，积极约束集为$\mathcal{A}(x^*)=\{1,2\}$.

Lagrange函数为$L(x,\lambda) = x_1 - \lambda_1 x_2 - \lambda_2(1-(x_1-1)^2-x_2^2)$.



由KKT条件$\nabla_x L(x^*,\lambda^*)=0$，可知$\begin{cases} \& -2\lambda_2^* = 0 \\ \& -\lambda_1^* = 0 \end{cases}$，从而$\lambda^* = (0,1/2)^T$.

在x^*处，$\nabla c_1(x^*) = (0,1)^T, \nabla c_2(x^*) = (2,0)^T$，满足线性无关.

此时，$\forall s \in \{(s_1,s_2)^T | s^T\nabla c_1(x^*) = 0, s^T\nabla c_2(x^*) = 0\} = (0,0)$，

$\nabla_{xx}L(x^*,\lambda^*) = \begin{bmatrix} 2\lambda_2^* & 0 \\ 0 & 2\lambda_2^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$，满足$s^T\nabla_{xx}L(x^*,\lambda^*)s \geq 0$.

二阶充分条件

定理7.6.2 设(x^*,λ^*)为约束优化问题(7.2.2)的K-T对.若对任意满足

$$\begin{cases} s^T\nabla c_i(x^*) = 0, & 若 i \in \mathcal{E} \\ s^T\nabla c_i(x^*) \geq 0, & 若 \lambda_i^* = 0, i \in I(x^*) \\ s^T\nabla c_i(x^*) = 0, & 若 \lambda_i^* > 0, i \in I(x^*) \end{cases}$$

非零向量$s \in \mathbb{R}^n$都有$s^T\nabla_{xx}L(x^*,\lambda^*)s > 0$，则$x^*$是约束优化问题(7.2.2)的严格局部最优解.进一步，存在$\gamma > 0$和$\delta > 0$使对任意$x \in N(x^*,\delta) \cap \Omega$，成立

$$f(x) \geq f(x^*) + \gamma\|x - x^*\|^2.$$

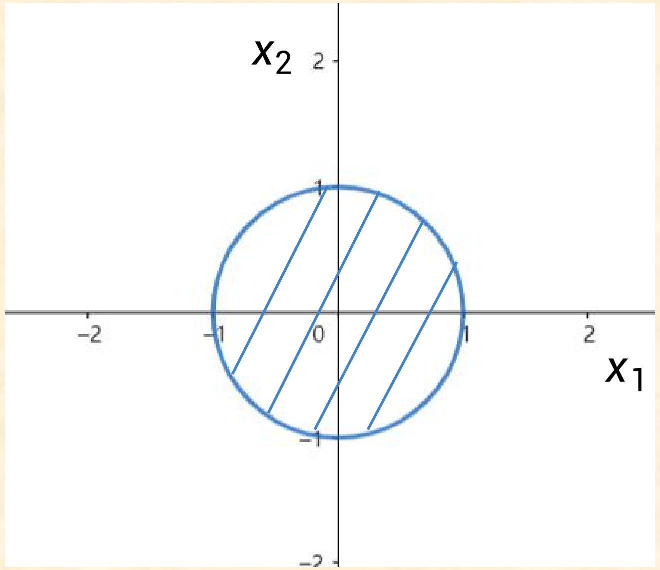
可用于验证KT点为局部最优解

北京交通大學

例子：二阶充分条件

例： 求解如下问题的局部最优解

$$
\begin{aligned}
\min \quad & f(x) = x_1 x_2 \\
s.t. \quad & c(x) = x_1^2 + x_2^2 - 1 = 0.
\end{aligned}
$$



解：该问题的Lagrange函数为$L(x,v) = x_1 x_2 - v(x_1^2 + x_2^2 - 1)$.　　$\nabla_{xx} L(x,v) =- 2vI$

由一阶必要条件(KKT) $\begin{cases} x_2 - 2vx_1 = 0 \\ x_1 - 2vx_2 = 0 \\ x_1^2 + x_2^2 - 1 = 0 \end{cases}$

可知

$v^* = \frac{1}{2}$， $x^* = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T \text{or } x^* = (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T$　　$\nabla_{xx} L(x^*, v^*) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

或

$v^* =- \frac{1}{2}$， $x^* = (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T \text{or } x^* = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T$　　$\nabla_{xx} L(x^*, v^*) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

例子：二阶充分条件

当$\boldsymbol{v}^* = \frac{1}{2}$时，对任意$s \neq 0$满足$s_1 = -s_2, (s^T \nabla c(x^*) = 0)$有

$$s^T \nabla_{xx} L(x^*, v^*) s = -4s_1^2 < 0$$

由二阶充分条件可知，$(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T$与$(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T$是局部极大值点；

当$\boldsymbol{v}^* = -\frac{1}{2}$时，对任意$s \neq 0$满足$s_1 = s_2, (s^T \nabla c(x^*) = 0)$, 有

$$s^T \nabla_{xx} L(x^*, v^*) s = 4s_1^2 > 0$$

由二阶充分条件可知，$(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T$与$(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T$是局部最优解.