Lecture 7 KKT与对偶

1 约束优化一阶最优性条件（KKT）

2 凸优化最优性条件

3 对偶规划与对偶理论

一阶最优性条件

约束优化模型

$$
\begin{aligned}
\min \quad & f(x) \\
s.t. \quad & g_i(x) \geq 0, \quad i = 1,2,\cdots,m \text{----不等式约束} \\
& h_j(x) = 0, \quad j = 1,2,\cdots,l. \text{----等式约束} \\
& x \in R^n
\end{aligned}
$$

可行域（集）： $S = \left\{ x \in R^n \,\middle|\, \begin{matrix} g_i(x) \geq 0, & i = 1,2,\cdots,m \\ h_j(x) = 0, & j = 1,2,\cdots,l \end{matrix} \right\}$

一阶最优性条件

可行方向

定义1 设集合$S \subset \mathbb{R}^n$, $\overline{\boldsymbol{x}} \in cl\ S$以及向量$\boldsymbol{d} \neq 0$, 若存在$\delta > 0$,使得对于任意$\lambda \in (0, \delta)$, 有$\overline{\boldsymbol{x}} + \lambda \boldsymbol{d} \in S$, 则称$d$为**集合$\boldsymbol{S}$**在$\overline{\boldsymbol{x}}$处的一个**可行方向（feasible direction）**.

■ **可行方向锥(集)：** $\overline{\boldsymbol{x}} \in cl\ S$

$$D = \{d \mid d \neq 0, \exists \delta > 0, \forall \lambda \in (0, \delta), 有\overline{\boldsymbol{x}} + \lambda \boldsymbol{d} \in S\}$$

一阶最优性条件

下降方向（回顾）

定义2 设$\bar{\boldsymbol{x}}, \boldsymbol{d} \in \mathbb{R}^n$且$\boldsymbol{d} \neq 0$，若存在$\delta > 0$，对于任意$t \in (0, \delta]$，有

$f(\bar{\boldsymbol{x}} + t\boldsymbol{d}) < f(\bar{\boldsymbol{x}})$，则称$d$是**函数$f$**在$\bar{\boldsymbol{x}}$处的一个**下降方向**.

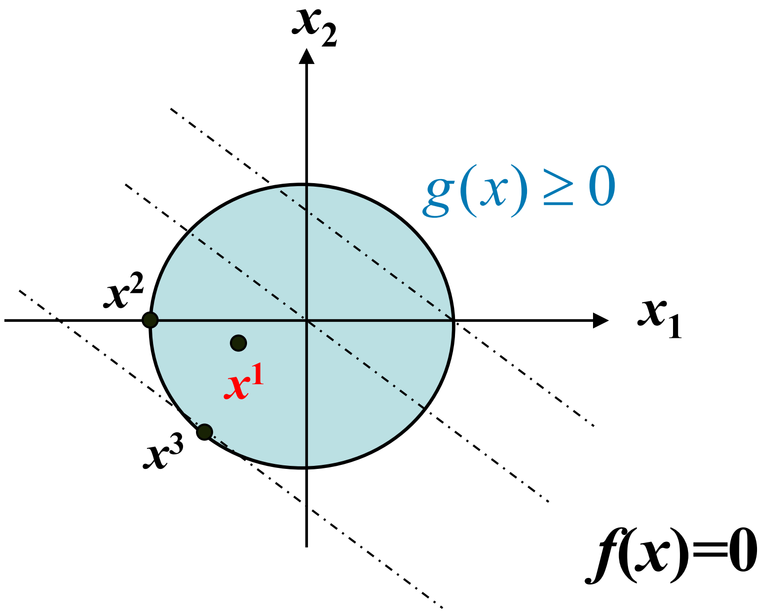
- **下降方向锥（集）：**

$$F_0 = \left\{ d \,\middle|\, \nabla f(\bar{x})^T d < 0 \right\}$$

一阶最优性条件

例1 $\min\ f(x)=x_1+x_2$

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



- 对于任意内点x^1,
  可行方向锥$D=R^2\backslash\{\mathbf{0}\}$
  目标函数下降方向锥$F_0=\{d\in R^2\,|\,d_1+d_2<0\}$
- 对于边界点$x^2=(-1,0)^T$,
  可行方向锥:$D=\{d\in R^2\,|\,d_1>0\}$;
  目标函数下降方向锥$F_0=\{d\in R^2\,|\,d_1+d_2<0\}$
- 最优解:$x^3=\left(-\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2}\right)^T$,
  可行方向锥:$D=\{d\in R^2\,|\,d_1+d_2>0\}$;
  目标函数下降方向锥$F_0=\{d\in R^2\,|\,d_1+d_2<0\}$

最优解处不存在可行且下降的方向!

$D\cap F_0=\emptyset$

一阶最优性条件

几何最优性条件

定理1 设考虑约束优化问题 $\min\limits_{x\in S} f(x)$,其中 $S \subset \mathbb{R}^n$ 为非空集合, $\bar{\boldsymbol{x}} \in S$, f 在 $\bar{\boldsymbol{x}}$ 处可微. 若 $\bar{\boldsymbol{x}}$ 是该问题的一个局部最优解, 则 $\boldsymbol{D} \cap \boldsymbol{F_0} = \emptyset$.

局部最优解处不存在可行且下降的方向!

证明：反证法假设存在 $d \in F_0 \bigcap D$,则 $d \in F_0, d \in D$.

$\because d \in F_0, \therefore \exists \delta_1 > 0$,对 $\forall \lambda \in (0,\delta_1)$,有 $f(\bar{x}+\lambda d) < f(\bar{x})$;

$\because d \in D, \therefore \exists \delta_2 > 0$,对 $\forall \lambda \in (0,\delta_2)$,有 $\bar{x}+\lambda d \in S$.

取 $\delta = \min(\delta_1,\delta_2)$,则当 $\lambda \in (0,\delta)$,有 $\bar{x}+\lambda d \in S$ 且 $f(\bar{x}+\lambda d) < f(\bar{x})$,

与 $\bar{x}$ 为局部最优解矛盾。

一阶最优性条件

几何最优性条件

定理1 设考虑约束优化问题$\min_{x\in S} f(x)$,其中$S\subset\mathbb{R}^n$为非空集合, $\boldsymbol{\bar{x}}\in S$, f在$\boldsymbol{\bar{x}}$处可微. 若$\boldsymbol{\bar{x}}$是该问题的一个局部最优解, 则$\boldsymbol{D\cap F_0=\emptyset}$.

局部最优解处不存在可行且下降的方向!

问题：如何判断$\boldsymbol{D\cap F_0=\emptyset}$

$$F_0=\left\{d\,\middle|\,\nabla f(\bar{x})^T d<0\right\}$$

$$D=\left\{d\,\middle|\,d\neq 0,\bar{x}\in clS,\exists\delta>0,\forall\lambda\in(0,\delta),有\bar{x}+\lambda d\in S\right\}$$

一阶最优性条件

不等式约束优化

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

可行域： $S = \{x | g_i(x) \geq 0, i = 1, \cdots, m\}$

积极约束： 定义：若问题(1)的一个可行点$\bar{x}$(即$\bar{x} \in S$)使某个不等式约束$g_i(x) \geq 0$变成等式，即$g_i(\bar{x}) = 0$，则该不等式约束称为关于可行点$\bar{x}$的起作用约束（积极约束）；否则，若$\bar{x}$使得某个$g_i(\bar{x}) > 0$，则该不等式约束称为关于可行点$\bar{x}$的不起作用约束（非积极约束）。

一阶最优性条件

不等式约束优化

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

可行域： $S = \{x | g_i(x) \geq 0, i = 1, \cdots, m\}$

局部约束方向锥：

积极约束指标集

$\bar{x} \in S$，记 $I := \{i | g_i(\bar{x}) = 0\}, G_0 = \{d | \nabla g_i(\bar{x})^T d > 0, i \in I\}$

称G_0为S在点$\bar{x}$处的局部约束方向锥（或内方向锥）。

一阶最优性条件

几何条件

定理2：设$\bar{x} \in S$，$f(x)$和$g_i(x)(i \in I)$在$\bar{x}$处可微，$g_i(x)(i \notin I)$在$\bar{x}$处连续，如果$\bar{x}$是问题(1)的局部最优解，则$F_0 \cap G_0 = \varnothing$。

一阶最优性条件

问题：如何判断 $G_0 \cap F_0 = \emptyset$

$$F_0 = \left\{ d \,\middle|\, \nabla f(\bar{x})^T d < 0 \right\} \qquad G_0 = \left\{ d \,\middle|\, -\nabla g_i(\bar{x})^T d < 0, i \in I \right\}$$

凸分析基础：Gordan定理

设A为$m \times n$矩阵，那么$Ax < 0$有解的充要条件是

不存在非零向量$y \geq 0$，使得$A^T y = 0$。

一阶最优性条件

代数条件

定理3(*Fritz John*条件) 设$\bar{x}\in S, I=\{i \mid g_i(\bar{x})=0\}, f(x), g_i(x)(i\in I)$在$\bar{x}$处可微，$g_i(x)(i\notin I)$在$\bar{x}$处连续，若$\bar{x}$是问题(1)的局部最优解，则存在不全为零的数$w_0, w_i(i\in I)$,使得

$$\begin{cases} w_0\nabla f(\bar{x})-\sum_{i\in I} w_i\nabla g_i(\bar{x})=0 \\ w_0,\ w_i\geq 0,\quad i\in I. \end{cases}$$

称$\bar{x}$为问题(1)的Fritz John 点(即满足Fritz John条件的点).

一阶最优性条件

Fritz John
(1910 - 1994)

John F., “Extremum Problems with Inequalities as Subsidiary Conditions” ,Studies and Essays, Courant Anniversary Volume, Interscience, New York, 1948.

一阶最优性条件

定理4(*Fritz John*条件) 设$\bar{x} \in S, I = \{i \mid g_i(\bar{x}) = 0\}$, $f(x), g_i(x)$在$\bar{x}$处可微，若$\bar{x}$是问题(1)的局部最优解，则存在不全为零的数$w_0, w_1, \cdots, w_m$,使得

$$\begin{cases} w_0 \nabla f(\bar{x}) - \sum_{i=1}^{m} w_i \nabla g_i(\bar{x}) = 0 \\ \boxed{\begin{array}{l} w_i g_i(\bar{x}) = 0, \quad i = 1, 2, \cdots, m \\ w_0, \; w_i \geq 0, \quad i = 1, 2, \cdots, m \end{array}} \end{cases}$$

取非积极约束对应的$w_i = 0$

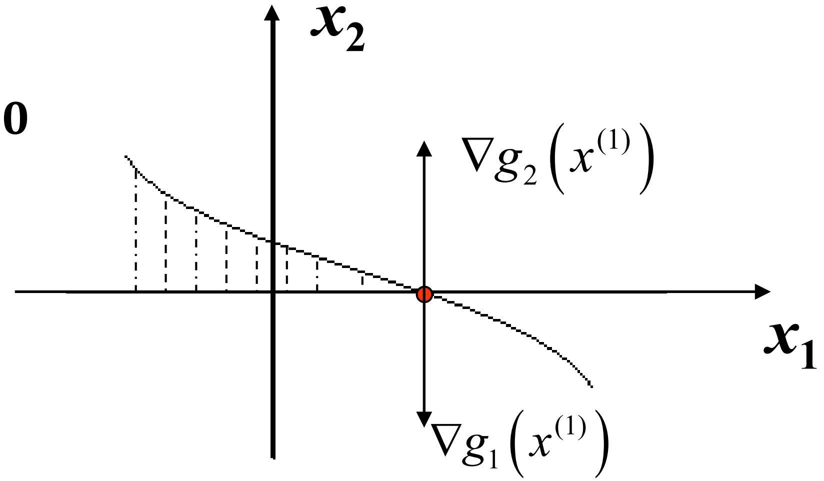
互补松弛条件

一阶最优性条件

例3 考虑非线性优化问题：

$$\min \ -x_1$$
$$s.t.\quad g_1(x) = -x_2 + (1-x_1)^3 \geq 0$$
$$g_2(x) = x_2 \geq 0$$

试判断$x^{(1)} = (1,0)^T$是否为该问题的Fritz John点.



一阶最优性条件

解： $\because \nabla f(x)=(-1,0)^T$

$\nabla g_1(x)=(-3(1-x_1)^2,-1)^T, \nabla g_2(x)=(0,1)^T$

在点$x^{(1)}=(1,0)^T$处, $I=\{1,2\}$

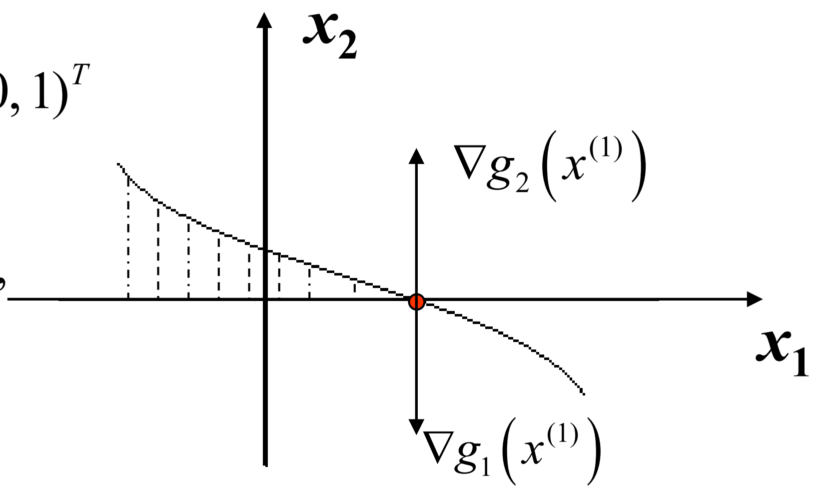
$\nabla f(x^{(1)})=(-1,0)^T, \nabla g_1(x^{(1)})=(0,-1)^T,$

$\nabla g_2(x^{(1)})=(0,1)^T$，设有

$$w_0\begin{pmatrix}-1\\0\end{pmatrix}-w_1\begin{pmatrix}0\\-1\end{pmatrix}-w_2\begin{pmatrix}0\\1\end{pmatrix}=0$$

$\Rightarrow w_0=0$, 取$w_1=w_2>0$

$\Rightarrow x^{(1)}=(1,0)^T$是$Fritz\ John$点。 **该Fritz John点亦是全局最优解！**



一阶最优性条件

例4 考虑非线性优化问题：

Fritz John点不一定是最优解

$$\min\ f(x)=(x_1-1)^2+(x_2-1)^2$$

$$s.t.\quad g_1(x)=(1-x_1-x_2)^3\ge 0$$

$$g_2(x)=x_1\ge 0$$

$$g_3(x)=x_2\ge 0$$

唯一最优解：$\boldsymbol{x}^*=\left(\frac{1}{2},\frac{1}{2}\right)^T$

$\because$ 直线$x_1+x_2=1$上所有可行点$\bar{x}$使$\nabla g_1(\bar{x})=0$，

$\therefore$ 取$w_0=0, w_1=a>0$，$w_2=w_3=0$，总有

$$w_0\nabla f(\bar{x})-w_1\nabla g_1(\bar{x})-w_2\nabla g_2(\bar{x})-w_3\nabla g_3(\bar{x})=0$$

说明在直线$x_1+x_2=1$上每个可行点$\bar{x}$都是$Fritz\ John$点，但除x^*外，都不是最优解。

此时Fritz John条件中实际上不包含目标函数的任何数据

一阶最优性条件

KKT条件

定理5. 考虑问题$\begin{cases}\min\ f(x)\\ s.t.\quad g_i(x)\ge 0, i=1,\cdots,m\end{cases}$设$\bar{x}\in S, f, g_i(i\in I)$在$\bar{x}$处可微，$g_i(i\notin I)$在$\bar{x}$连续，$\{\nabla g_i(\bar{x})\,|\,i\in I\}$线性无关，若$\bar{x}$是局部最优解，则存在非负数$w_i, i\in I$，使得

$$\nabla f(\bar{x})-\sum_{i\in I} w_i\nabla g_i(\bar{x})=0.$$

约束规格（Constraint Qualification）

一阶最优性条件

KKT条件

证明：由定理3，存在不全为零的非负数

$w_0, w_i', i \in I$，使得

$$w_0 \nabla f(\bar{x}) - \sum_{i \in I} w_i' \nabla g_i(\bar{x}) = 0.$$

显然$w_0 \neq 0$，否则$\nabla g_i(\bar{x}) (i \in I)$线性相关，矛盾。

于是，令$w_i = \dfrac{w_i'}{w_0} \geq 0 \ (i \in I)$，得

$$\nabla f(\bar{x}) - \sum_{i \in I} w_i \nabla g_i(\bar{x}) = 0.$$

一阶最优性条件

KKT条件

定理6. 考虑问题$\begin{cases} \min\ f(x) \\ s.t. \quad g_i(x) \geq 0, i=1,\cdots,m \end{cases}$

设$\bar{x} \in S, f, g_i$在$\bar{x}$可微,$\{\nabla g_i(\bar{x}) \mid i \in I\}$线性无关，

若$\bar{x}$是局部最优解，则存在数$w_i, i=1,2,\cdots,m,$使得

$$\nabla f(\bar{x}) - \sum_{i=1}^{m} w_i \nabla g_i(\bar{x}) = 0$$

$$\boxed{\begin{array}{ll} w_i g_i(\bar{x}) = 0 & i=1,2,\cdots,m \\ w_i \geq 0 & i=1,2,\cdots,m. \end{array}}$$

取非积极约束对应的$w_i = 0$

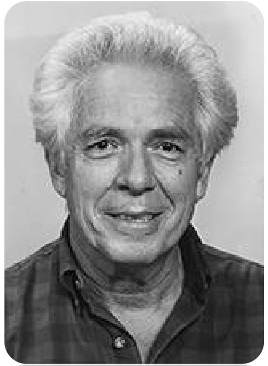
互补松弛条件

一阶最优性条件

KKT条件　　**现代非线性优化兴起的标志性工作!**

Karush 1939, Kuhn&Tucker 1951

William Karush
(1917-1997)



Harold William Kuhn
1925 - 2014

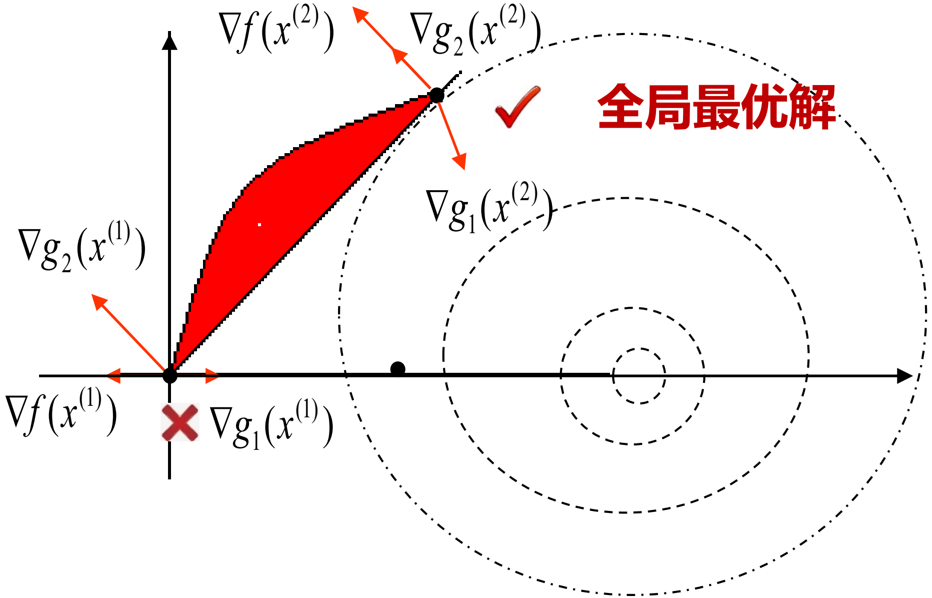
Albert William Tucker
1905 - 1995

一阶最优性条件

例5 考虑非线性优化问题：

$$\min\ (x_1-2)^2+x_2^2$$
$$\text{s.t.}\quad g_1(x)=x_1-x_2^2\geq 0$$
$$g_2(x)=-x_1+x_2\geq 0$$

试判断$x^{(1)}=(0,0)^T$，
$x^{(2)}=(1,1)^T$是否为
该问题的KKT点.



$$\nabla f(x)=\begin{bmatrix}2(x_1-2)\\2x_2\end{bmatrix},\nabla g_1(x)=\begin{bmatrix}1\\-2x_2\end{bmatrix},\nabla g_2(x)=\begin{bmatrix}-1\\1\end{bmatrix}$$

一阶最优性条件

例6 求如下不等式约束优化问题的KKT点：

$$\min\ f(x) = x_1^2 + x_2^2 - 6x_1 - 6x_2 + 10$$

$$\begin{aligned} \text{s.t.}\quad & g_1(x) = 4 - x_1 - x_2 \geq 0 \\ & g_2(x) = x_1 \geq 0 \\ & g_3(x) = x_2 \geq 0 \end{aligned}$$

解： $\nabla f(x) = 2\begin{bmatrix} x_1 - 3 \\ x_2 - 3 \end{bmatrix}$

$$\nabla g_1(x) = \begin{bmatrix} -1 \\ -1 \end{bmatrix},\ \nabla g_2(x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\quad \nabla g_3(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

一阶最优性条件

由KKT条件得 $$\begin{bmatrix} x_1-3 \\ x_2-3 \end{bmatrix} - \lambda_1 \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \lambda_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

由KKT条件及约束条件得 $$\begin{cases} x_1+\lambda_1-\lambda_2=3 \\ x_2+\lambda_1-\lambda_3=3 \\ \lambda_1(4-x_1-x_2)=0 \\ \lambda_2 x_1=0 \\ \lambda_3 x_2=0 \\ x_1+x_2\le 4 \\ \lambda_1,\lambda_2,\lambda_3,x_1,x_2\ge 0 \end{cases}$$

一阶最优性条件

(1) 若 $x_1 = x_2 = 0$：　由$\lambda_1(4 - x_1 - x_2) = 0$可得$\lambda_1 = 0$。

$\therefore \lambda_1 - \lambda_2 = 3 \Rightarrow \lambda_2 = -3$　　这与$\lambda_2 \geq 0$矛盾。

(2) 若 $x_1 = 0, x_2 \neq 0$：$\therefore \lambda_3 = 0$

$$\therefore \begin{cases} x_2 + \lambda_1 = 3 \\ \lambda_1 - \lambda_2 = 3 \end{cases} \Rightarrow x_2 + \lambda_2 = 0 \Rightarrow \lambda_2 = -x_2 < 0$$

这与$\lambda_2 \geq 0$矛盾。

(3) 若 $x_1 \neq 0, x_2 = 0$：$\therefore \lambda_2 = 0$

$$\therefore \begin{cases} x_1 + \lambda_1 = 3 \\ \lambda_1 - \lambda_3 = 3 \end{cases} \Rightarrow x_1 + \lambda_3 = 0 \Rightarrow \lambda_3 = -x_1 < 0$$

这与$\lambda_3 \geq 0$矛盾。

一阶最优性条件

(4) 若 $x_1 \neq 0, x_2 \neq 0$:　$\therefore \lambda_2 = \lambda_3 = 0$

$$\therefore \begin{cases} x_1 + \lambda_1 = 3 \\ x_2 + \lambda_1 = 3 \end{cases} \Rightarrow x_1 = x_2$$

若 $x_1 + x_2 < 4 \ \Rightarrow \lambda_1 = 0 \ \Rightarrow x_1 = x_2 = 3$

$\Rightarrow x_1 + x_2 = 6 > 4 \ \Rightarrow$ 矛盾。

$\therefore x_1 + x_2 = 4 \ \Rightarrow x_1 = x_2 = 2 \ \Rightarrow \lambda_1 = 1$

$\therefore [2,2]^T$ 为KKT点。

一阶最优性条件

一般约束优化

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

$$\Downarrow$$

$$\begin{aligned} &\min\ f(x) \\ &s.t.\quad g(x) \ge 0 \\ &\qquad\quad h(x) = 0 \end{aligned}$$

其中$g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{bmatrix}, h(x) = \begin{bmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_l(x) \end{bmatrix}$

- **难点：等式约束可行移动的描述**

一阶最优性条件

曲面&曲线： 称点集$C=\{x=x(t)\mid t_0\le t\le t_1\}$为曲面$S=\{x\mid h(x)=0\}$上的一条曲线，如果对所有$t\in[t_0,t_1]$均有$h(x(t))=0$.

如果导数$x'(t)=\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}$存在，则称曲线是可微的. 此时的一阶导数$x'(t)$是曲线在点$x(t)$处的切向量.

切平面： 曲面S上在点x处所有可微曲线的切向量组成的集合称为曲面S在点x的切平面，记为$T(x)$.

收集了所有可能的切向量（序列可行方向）！

一阶最优性条件

例9 考虑单位球面$S=\{x\,|\,h(x)=x_1^2+x_2^2+x_3^2-1=0\}$以及

单位圆周曲线$C=\left\{x=x(t)\,\middle|\,x(t)=\begin{pmatrix}\sin t\\ \cos t\\ 0\end{pmatrix},0\le t\le 2\pi\right\}.$

* 单位圆周曲线C为单位球面S上的一条曲线，因为对所有$t\in[0,2\pi]$均有$h(x(t))=0.$
* 由于$x'(t)=\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=(\cos t,-\sin t,0)^T$，因此该曲线是可微的.
* 考虑S上的点$x=(1,0,0)^T$，其切平面$T(x)=\{d\in \mathrm{R}^3\,|\,d_1=0\}$, 即$x_2Ox_3$平面.

  注意到在点$x=(1,0,0)^T$处，$\nabla h(x)=(2,0,0)^T$, 此时有

$$T(x)=\{d\in \mathrm{R}^3\,|\,d_1=0\}=\boxed{\{d\in \mathrm{R}^3\,|\,\nabla h(x)^T d=0\}}.$$

线性子空间，且比切平面更易于计算！

一阶最优性条件

线性化可行方向集(锥)

子空间： $H(x):=\left\{d\mid\nabla h(x)^T d=0\right\}$ 其中$\nabla h(x)=\left(\nabla h_1(x),\nabla h_2(x),\cdots,\nabla h_l(x)\right)$

注意到： □ 记$L_h(\bar{x}):=\left\{x\in R^n\,\middle|\,h(\bar{x})+\nabla h(\bar{x})^T(x-\bar{x})=0\right\}$，它是约束集

$S=\left\{x\in R^n\,\middle|\,h(x)=0\right\}$中的约束函数$h$线性化产生的约束集.

□ $H(\bar{x})$为集合$L_h(\bar{x})$的可行方向集(锥)，因而称为约束集S在

$\bar{x}$处的线性化可行方向集(锥).

结　论： 设$\bar{x}$是约束集合$S=\left\{x\in R^n\,\middle|\,h(x)=0\right\}$上的点，$T(\bar{x})$为$S$在$\bar{x}$

处的切平面，$H(\bar{x})$为相应的线性化可行方向集，则有

$T(\bar{x})\subseteq H(\bar{x})$，即若向量$d\in T(\bar{x})$，则有$\nabla h(\bar{x})^T d=0$.

问题：包含关系是否可以取等号？何时相等？

一阶最优性条件

线性化可行方向集(锥)

定理7：设$\bar{x} \in S = \{x \mid h(x) = 0\}$且$\{\nabla h_1(\bar{x}), \cdots, \nabla h_l(\bar{x})\}$线性无关，

则 $T(\bar{x}) = H(\bar{x})$.

一阶最优性条件

证明：设$d \in H(\bar{x})$,考虑非线性方程组$h(\bar{x}+td+\nabla h(\bar{x})y)=0$

其中$t \in R^1, y \in R^l$.该方程组有解$(y,t)=(0,0)$.

在$t=0$处，h关于y的$Jacobi$矩阵为$\nabla h(\bar{x})^T \nabla h(\bar{x})$

由隐函数定理，在$t=0$的邻域，存在连续可微函数

$y=y(t)(y(0)=0)$,使$h(\bar{x}+td+\nabla h(\bar{x})y(t))=0$成立

令$x(t)=\bar{x}+td+\nabla h(\bar{x})y(t)$,则$x(t)$为曲面$S$上过$x(0)$的一条

曲线。在点$x(0)=\bar{x}$，切向量为

$$x'(0)=d+\nabla h(\bar{x})y'(0)$$

又 $\nabla h(\bar{x})^T d+\nabla h(\bar{x})^T \nabla h(\bar{x})y'(0)=0$

而$d \in H(\bar{x}) \Rightarrow \nabla h(\bar{x})^T d=0, \because \{\nabla h_1(\bar{x}),\cdots,\nabla h_l(\bar{x})\}$线性无关

$\therefore \nabla h(\bar{x})^T \nabla h(\bar{x})$可逆,$\Rightarrow y'(0)=0 \Rightarrow x'(0)=d \Rightarrow d \in T(\bar{x})$.

一阶最优性条件

几何条件

定理8：设$\bar{x} \in S$，$f(x)$和$g_i(x)(i \in I)$在$\bar{x}$处可微，$g_i(x)(i \notin I)$在$\bar{x}$处连续，$h_j(j=1,2,\cdots,l)$在$\bar{x}$处连续可微，且$\{\nabla h_1(\bar{x}),\cdots,\nabla h_l(\bar{x})\}$线性无关。如果$\bar{x}$是问题$(NP)$的局部最优解，则在$\bar{x}$处，有

$$F_0 \cap G_0 \cap H_0 = \varnothing,$$

其中 $F_0 = \{d \mid \nabla f(\bar{x})^T d < 0\}$, 目标函数下降方向

$G_0 = \{d \mid \nabla g_i(\bar{x})^T d > 0, i \in I\}$, 不等式约束局部约束方向

$H_0 = \{d \mid \nabla h_j(\bar{x})^T d = 0, j = 1,2,\cdots,l\}$. 等式约束的切方向

可行方向

一阶最优性条件

定理8：设$\bar{x}\in S$，$f(x)$和$g_i(x)(i\in I)$在$\bar{x}$处可微，$g_i(x)(i\notin I)$在$\bar{x}$处连续，$h_j(j=1,2,\cdots,l)$在$\bar{x}$处连续可微，且$\{\nabla h_1(\bar{x}),\cdots,\nabla h_l(\bar{x})\}$线性无关。如果$\bar{x}$是问题$(NP)$的局部最优解，则在$\bar{x}$处，有

$$F_0\cap G_0\cap H_0=\varnothing,$$

其中

$$F_0=\{d\mid\nabla f(\bar{x})^T d<0\},$$

$$G_0=\{d\mid\nabla g_i(\bar{x})^T d>0,i\in I\},$$

$$H_0=\{d\mid\nabla h_j(\bar{x})^T d=0,j=1,2,\cdots,l\}.$$

$$\begin{cases}Ax<0\\Bx=0\end{cases}\quad\text{无解}$$

一阶最优性条件

引理1 若系统$\boldsymbol{Ax < 0, Bx = 0}$无解,则系统$\boldsymbol{A^T y + B^T z = 0, y \geq 0}$,且$\boldsymbol{(y; z) \neq 0}$ 有解.

证明：假设系统$Ax < 0, Bx = 0$无解，定义

$$S_1 = \left\{ \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \middle| d_1 = Ax, d_2 = Bx, x \in R^n \right\}, \quad S_2 = \left\{ \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \middle| d_1 < 0, d_2 = 0 \right\}$$

则S_1, S_2为非空凸集，且$S_1 \cap S_2 = \varnothing$,由凸集分离定理

存在非零向量$p = \begin{pmatrix} y \\ z \end{pmatrix}$,使得对$\forall x \in R^n, \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \in clS_2$有

$$y^T Ax + z^T Bx \geq y^T d_1 + z^T d_2 = y^T d_1$$

令$x = 0, \because d_1 < 0, \therefore y \geq 0.$

令$d_1 \to 0$，得$y^T Ax + z^T Bx \geq 0.$

取$x = -\left(A^T y + B^T z\right)$,则有$-\left\|A^T y + B^T z\right\|^2 \geq 0$

$\Rightarrow A^T y + B^T z = 0$.由于$p = \begin{pmatrix} y \\ z \end{pmatrix} \neq 0$，因此有$y \neq 0$或者$z \neq 0$.证毕.

一阶最优性条件

Linear Independency Constraint Qualification(LICQ)

定义: 设$\bar{x}$为可行点，不等式约束中在$\bar{x}$起作用约束下标集记为I，如果向量组

$$\left\{\nabla g_i(\bar{x}), \nabla h_j(\bar{x}) \mid i \in I, j = 1, 2, \cdots, l\right\}$$

线性无关，则称约束集$\{x \mid g(x) \geq 0, h(x) = 0\}$在$\bar{x}$处线性无关约束规范（LICQ）成立。

此时，称$\bar{x}$为约束$g(x) \geq 0$和$h(x) = 0$的正则点。

一阶最优性条件

KKT条件

定理11(*KKT*必要条件）考虑问题 $\begin{cases} \min\ f(x) \\ s.t.\quad g_i(x) \geq 0, i = 1,\cdots,m \\ \qquad\quad h_j(x) = 0,\ j = 1,\cdots,l \end{cases}$

设$\bar{x}$为可行点,$I = \{i \mid g_i(\bar{x}) = 0\}$.$f, g_i(i \in I)$在$\bar{x}$处可微，$g_i(i \notin I)$在$\bar{x}$连续，$h_j(j = 1,\cdots,l)$在$\bar{x}$连续可微，向量集

$$\{\nabla g_i(\bar{x}),\ \nabla h_j(\bar{x}) \mid i \in I, j = 1,\cdots,l\}$$

线性无关（即LICQ成立），若$\bar{x}$是局部最优解，则存在数$w_i, i \in I$和$v_j(j = 1,\cdots,l)$，使得

$$\nabla f(\bar{x}) - \sum_{i \in I} w_i \nabla g_i(\bar{x}) - \sum_{j=1}^{l} v_j \nabla h_j(\bar{x}) = 0.$$

$$w_i \geq 0 \quad (i \in I).$$

一阶最优性条件

定理12(KKT必要条件）考虑问题$\begin{cases} \min\ f(x) \\ s.t. \quad g_i(x) \ge 0, i=1,\cdots,m \\ \qquad\quad h_j(x)=0,\ j=1,\cdots,l \end{cases}$

设$\bar{x}$为可行点,$I=\{i \mid g_i(\bar{x})=0\}$.$f,g_i$在$\bar{x}$处可微，$h_j(j=1,\cdots,l)$在$\bar{x}$连续可微，向量集

$$\{\nabla g_i(\bar{x}),\ \nabla h_j(\bar{x}) \mid i\in I, j=1,\cdots,l\}$$

线性无关(LICQ)，若$\bar{x}$是局部最优解，则存在数$w_i, i\in I$和$v_j(j=1,\cdots,l)$，使得

$$\nabla f(\bar{x})-\sum_{i=1}^{m} w_i \nabla g_i(\bar{x})-\sum_{j=1}^{l} v_j \nabla h_j(\bar{x})=0$$

$$w_i g_i(\bar{x})=0, i=1,2,\cdots,m$$

$$w_i \ge 0 \quad i=1,2,\cdots,m.$$

一阶最优性条件

例11 求如下不等式约束优化问题的KKT点：

$$
\begin{cases}
\min\ f=(x_1-1)^2+(x_2-1)^2 \\
s.t. \quad -x_1-x_2+2\ge 0 \\
\qquad\quad x_2-x_1=1 \\
\qquad\quad x_1\ge 0 \\
\qquad\quad x_2\ge 0
\end{cases}
$$

分析： $\nabla f(x)=\left(2(x_1-1),2(x_2-1)\right)^T \quad \nabla g_1(x)=(-1,-1)^T$

$\nabla g_2(x)=(1,0)^T \quad \nabla g_3(x)=(0,1)^T \quad \nabla h(x)=(-1,1)^T$

一阶最优性条件

KKT条件为：

$$\begin{cases} 2(x_1-1)-w_1(-1)-w_2-v(-1)=0 \\ 2(x_2-2)-w_1(-1)-w_3-v=0 \\ -x_1-x_2+2\geq 0 \\ x_2-x_1=1 \\ x_1,x_2\geq 0 \\ w_1(-x_1-x_2+2)=0 \\ w_2x_1=0 \\ w_3x_2=0 \\ w_1,w_2,w_3\geq 0 \end{cases}$$

$x^*=\left(\frac{1}{2},\frac{3}{2}\right)^T$ 为KKT点.

$$\boldsymbol{w_1=1, w_2=w_3=v=0}$$

一阶最优性条件

Lagrange函数

定义 约束优化问题$\min_{x\in\mathbb{R}^n}\{f(x): g_i(x)\geq 0, i=1,\dots,m; h_j(x)=0, j=1,\dots,l\}$的**Lagrange函数**定义为

$$L(x,w,v)=f(x)-\sum_{i=1}^{m}w_i g_i(x)-\sum_{j=1}^{l}v_j h_j(x)$$
$$=f(x)-w^T g(x)-v^T h(x)$$

其中$w=(w_1,w_2,\cdots,w_m)^T\in R_+^m$，$v=(v_1,v_2,\cdots,v_l)^T\in R^l$

称为Lagrange乘子.

- 可行解x处：$f(x)\geq L(x,w,v)$
- 不等式约束的乘子符号由不等号方向决定：$\langle g(x),w\rangle\geq 0$

一阶最优性条件

KKT条件

定理13(KKT必要条件）考虑问题 $\begin{cases} \min\ f(x) \\ s.t.\quad g_i(x) \geq 0, i=1,\cdots,m \\ \qquad\quad h_j(x)=0,\ j=1,\cdots,l \end{cases}$

设$\bar{x}$为可行点,$I=\{i\,|\,g_i(\bar{x})=0\}$.$f,g_i$在$\bar{x}$处可微，$h_j(j=1,\cdots,l)$在$\bar{x}$连续可微，设LICQ在$\bar{x}$处满足，若$\bar{x}$是局部最优解，则存在乘子向量$\bar{w}\geq 0,\bar{v}$,使得 $\nabla_x L(\bar{x},\bar{w},\bar{v})=0.$

一阶最优性条件

一阶最优性条件

$$
\begin{cases}
\nabla_x L(x,w,v)=0 \\
w_i g_i(x)=0,\ i=1,2,\cdots,m \\
g_i(x)\geq 0,\ i=1,2,\cdots,m \\
h_j(x)=0,\ j=1,2,\cdots,l \\
w_i \geq 0,\ i=1,2,\cdots,m
\end{cases}
$$

Lagrange函数梯度为0

互补条件

原始可行性

对偶可行性

（见后续对偶理论部分）

凸优化最优性条件

凸优化模型

$$\min \quad f(x)$$

$$s.t. \quad x \in \Omega := \left\{ x \in R^n \,\middle|\, \begin{array}{l} g_i(x) \geq 0,\ i = 1, \cdots, m \\ h_j(x) = 0,\ j = 1, \cdots, l \end{array} \right\}$$

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.

- $g_i(x)$凹函数 $\Rightarrow -g_i(x)$凸 $\Rightarrow \{x | -g_i(x) \leq 0\}$凸集（水平集性质）
- $h_j(x)$线性函数 $\Rightarrow \{x | -h_j(x) \leq 0\}, \{x | \ h_j(x) \leq 0\}$凸集 $\Rightarrow \{x | \ h_j(x) = 0\}$凸集

基本假设：

- □ **可行域 Ω为非空闭凸集**
- □ **凸函数 f 连续可微**

凸优化最优性条件

常见凸优化

线性规划

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

线性最小二乘

$$\min \quad \frac{1}{2}||Ax - b||^2$$

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

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

凸优化最优性条件

性质1 **凸优化问题的任一局部最优解为全局最优解.**

证明： 设x^*为凸优化局部最优解，则存在$\delta > 0$，对任意$x \in N(x^*, \delta) \cap \Omega$

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

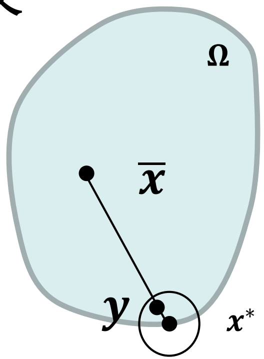
$$f(\bar{x}) < f(x^*).$$

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

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

当λ充分小时， $y \in N(x^*, \delta) \cap \Omega$ 且$\boldsymbol{f(y) < f(x^*)}$,

这与x^*为局部最优解矛盾. 证毕



凸优化最优性条件

性质2 **凸优化问题的最优解集为凸集.**

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

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

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

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

为凸集. 证毕

凸优化最优性条件

性质3 **严格凸优化问题若存在最优解，则最优解唯一.**

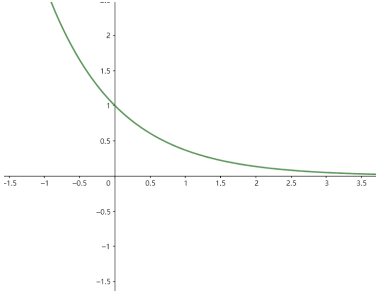
证明： 假设x^*, y^*均为严格凸优化的最优解，且$x^* \neq y^*$, 取$z^* = \frac{1}{2}(x^* + y^*)$，

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

- **严格凸优化问题不一定存在最优解.**



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

$$(e^{-x})'' = e^{-x} > 0$$

凸优化最优性条件

稳定点： 称x^*为约束优化问题 $\min \{f(x): x \in \Omega\}$ 的稳定点，

若满足$\langle \nabla f(x^*), x - x^* \rangle \geq 0, \forall x \in \Omega$.

性质4 凸优化问题的全局最优解与稳定点等价.

证明： “ $\Leftarrow$ ” 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

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

因而 x^*为全局最优解.

凸优化最优性条件

证明：（续）

“ $\Longrightarrow$ ” 设x^*为凸优化问题的全局最优解，由一阶最优性条件可知，

在x^*处不存在可行且使得目标函数下降的方向。注意到, 由Ω的凸性

可知 $\forall x \in \Omega, \forall \lambda \in (0,1)$,

$$x^* + \lambda(x - x^*) = \lambda x + (1-\lambda)x^* \in \Omega$$

因此，$x - x^*$是x^*处的可行方向. 从而$\langle \nabla f(x^*), x - x^* \rangle \geq 0, \forall x \in \Omega$,

即x^*为稳定点. 证毕

当可行域为凸集时，则最优解为稳定点.

凸优化最优性条件

定理1 若约束优化问题的可行域Ω为非空闭凸集，则其任一局部最优解 x^*均为稳定点，即 $\langle \nabla f(x^*), x - x^* \rangle \geq 0, \forall x \in \Omega.$

注意：上述定理与目标函数是否为凸函数无关，但可行域的凸性不能去掉！

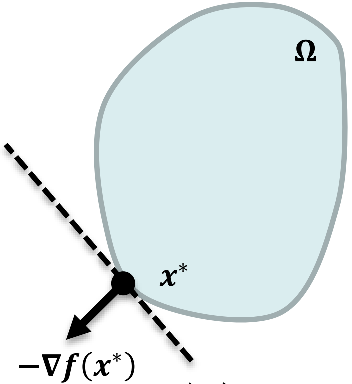


图1

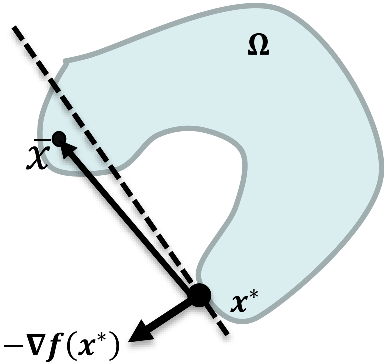


图2

若Ω非凸，上述结论可能不成立.

如图2所示，x^*是局部最优解，但x^*不是稳定点，因为$\langle \nabla f(x^*), \bar{x} - x^* \rangle < 0$.

凸优化最优性条件

鞍点

定义1称$(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|}$为约束优化问题

$$\mathbf{min}_{\boldsymbol{x}} \{f(x): h_i(x) = 0, i \in \mathcal{E}; g_i(x) \geq 0, i \in \mathcal{I}\}$$

的Lagrange函数$L(x, \lambda, \mu)$的**鞍点**,若$\forall (x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}_+^{|\mathcal{I}|}$有

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

$$\Updownarrow$$

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

凸优化最优性条件

定理2 若(x^*,λ^*,μ^*)是凸优化的KKT点对, 则(x^*,λ^*,μ^*)为Lagrange函数的鞍点.

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

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

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

=0 (x^*,λ^*,μ^*)为KKT点对

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

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

$$= -\sum_{i\in I} \mu_i g_i(x^*) \leq 0$$ 对任意$\lambda \in \mathbb{R}^{|\mathcal{E}|}, \mu \in \mathbb{R}_+^{|I|}$

凸优化最优性条件

定理3 若(x^*,λ^*,μ^*)是优化问题Lagrange函数的鞍点, 则(x^*,λ^*,μ^*)为KKT点对.

证明:

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

- 由$(\lambda^*,\mu^*) \in \underset{\substack{\lambda\in\mathbb{R}^{|\mathcal{E}|}\\ \mu\in\mathbb{R}_+^{|\mathcal{I}|}}}{\operatorname{argmax}}\ L(x^*,\lambda,\mu) = f(x^*) - \sum_{i\in\mathcal{E}}\lambda_i h_i(x^*) - \sum_{i\in\mathcal{I}}\mu_i g_i(x^*)$

$$\Downarrow$$

$$\boxed{\begin{aligned} &h_i(x^*) = 0, \quad i\in\mathcal{E};\\ &g_i(x^*) \geq 0,\ \mu_i^* \geq 0,\ \mu_i^* g_i(x^*) = 0, \quad i\in\mathcal{I},\end{aligned}}$$

因此, (x^*,λ^*,μ^*)为KKT点对

注意：该定理无需问题为凸优化!

凸优化最优性条件

全局最优解 $\Longleftrightarrow$ 局部最优解 $\underset{\Longleftarrow}{\overset{约束规格*}{\Longrightarrow}}$ KKT点

$\Updownarrow$ 稳定点　　　　$\Updownarrow$ 鞍点

凸优化

***Slater 条件：** 若存在可行点$\bar{x}$使得 $g_i(\bar{x}) > 0, \ \forall i$

$\Downarrow$

任意可行解处均有可行方向

则称该问题满足Slater约束规格或Slater条件.

- 对凸优化 $\min_x\{f(x) : Ax = b\}$，其中f为适当闭凸函数，若存在$\bar{x} \in \operatorname{int} dom f$，使得$Ax = b$，则称Slater条件成立.

对偶规划与对偶理论

- Lagrange对偶问题
- 弱对偶理论
- 鞍点与强对偶定理
- Lagrange乘子的经济学解释

对偶规划与对偶理论

原始优化问题

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

$\Rightarrow$ 容易处理的集合约束

Lagrange对偶问题

$$
\begin{aligned}
\max\ & \theta(w,v) \\
s.t.\ & w \geq 0
\end{aligned}
\qquad \text{(D)}
$$

其中 $\theta(w,v) = \inf\left\{ \underbrace{\boxed{f(x) - \sum_{i=1}^{m} w_i g_i(x) - \sum_{j=1}^{l} v_j h_j(x)}}_{L(x;w,v)} \,\middle|\, x \in D \right\}$

Lagrange函数

对偶规划与对偶理论

对偶问题必为凸规划

$$
\text{(D)} \quad \begin{aligned} &\max\ \theta(w,v) \\ &s.t. \quad w \ge 0 \end{aligned}
$$

✓ **可行域为凸集合**

✓ **目标函数为凹函数**

对于任意的$x \in D$，*Lagrange*函数$L(x;w,v)$是w,v的线性函数，对偶函数$\theta(w,v)$作为线性函数的逐点下确界，必然是一个凹函数

对偶规划与对偶理论

例1 求如下优化问题的Lagrange对偶问题

$$\min\ x_1^2 + x_2^2$$

$$\text{s.t.}\ \ x_1 + x_2 - 4 \geq 0$$

$$x_1, x_2 \geq 0$$

解. 将容易处理的非负约束看成集合约束，即

$$x \in D \coloneqq \{x = (x_1, x_2)^T : x_1, x_2 \geq 0\}$$

则**Lagrange**对偶问题的目标函数为：

$$\theta(w) \coloneqq \inf_{x\in D}\{x_1^2 + x_2^2 - \mathrm{w}(x_1 + x_2 - 4)\},\ w \geq 0$$

对偶规划与对偶理论

$$
\begin{aligned}
\theta(w) &= \inf\left\{x_1^2 + x_2^2 - w(x_1+x_2-4) \mid x \in D\right\} \\
&= \inf\left\{x_1^2 - wx_1 + x_2^2 - wx_2 + 4w \mid x_1 \geq 0, x_2 \geq 0\right\} \\
&= \inf\left\{x_1^2 - wx_1 \mid x_1 \geq 0\right\} + \inf\left\{x_2^2 - wx_2 \mid x_2 \geq 0\right\} + 4w
\end{aligned}
$$

当$w \geq 0$时,

$$\inf\left\{x_1^2 - wx_1 \mid x_1 \geq 0\right\} = \left(\frac{w}{2}\right)^2 - w \times \frac{w}{2} = -\frac{w^2}{4}.$$

$$\inf\left\{x_2^2 - wx_2 \mid x_2 \geq 0\right\} = \left(\frac{w}{2}\right)^2 - w \times \frac{w}{2} = -\frac{w^2}{4}.$$

$$\therefore \theta(w) = -\frac{w^2}{4} - \frac{w^2}{4} + 4w = -\frac{w^2}{2} + 4w.$$

对偶问题 $\Longrightarrow$

$$
\begin{aligned}
&\max\ -\frac{w^2}{2} + 4w \\
&s.t. \quad w \geq 0
\end{aligned}
\qquad \text{(D)}
$$

对偶规划与对偶理论

习题1 求如下优化问题的Lagrange对偶问题

$$\max\ \boldsymbol{b}^T\boldsymbol{x}$$

$$\text{s.t.}\ \ \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \le \mathbf{1}$$

其中$\boldsymbol{b} \in \mathbb{R}^n \backslash \{\mathbf{0}\}, \boldsymbol{Q} \in \mathbb{R}^{n\times n}$为对称正定矩阵.

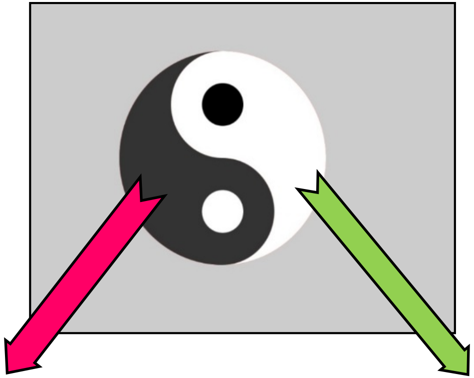
答案. 该问题的**Lagrange**对偶问题为:

$$\min\ \frac{\boldsymbol{b}^T\boldsymbol{Q}^{-1}\boldsymbol{b}}{\mathbf{4}w} + w$$

$$\text{s.t.}\ \ w > 0$$

对偶规划与对偶理论

对偶是一种普遍现象,辩证统一



$$(P)\quad \underbrace{\inf_{x\in D}\sup_{w\geq 0} L(x;w,v)}_{p^*}$$

$$(D)\quad \underbrace{\sup_{w\geq 0}\inf_{x\in D} L(x;w,v)}_{d^*}$$

$$\boxed{p^* \geq d^*}$$

弱对偶性

对偶规划与对偶理论

弱对偶定理

定理 设x和(w,v)分别是原问题和对偶问题的可行解，则

$$f(x)\geq\theta(w,v)。$$

证明：$\because$ x和(w,v)是可行解，

$$\therefore\quad g(x)\geq 0, h(x)=0, w\geq 0$$

$$\therefore \theta(w,v)=\inf_{x'}\left\{f(x')-w^T g(x')-v^T h(x')\,|\,x'\in D\right\}$$

$$\leq f(x)-w^T g(x)-v^T h(x)$$

$$\leq f(x).$$

对偶规划与对偶理论

$$(P)\ \underbrace{\inf_{x\in D}\sup_{w\geq 0} L(x;w,v)}_{p^*} \qquad (D)\ \underbrace{\sup_{w\geq 0}\inf_{x\in D} L(x;w,v)}_{d^*}$$

- 对偶间隙(Duality Gap): $\boldsymbol{\delta = p^* - d^* \geq 0}$

- **问题：对偶间隙 δ 何时为0呢?**

对偶规划与对偶理论

回顾：鞍点

定义 称$(x^*,\lambda^*,\mu^*)\in\mathbb{R}^n\times\mathbb{R}^{|\mathcal{E}|}\times\mathbb{R}_+^{|\mathcal{I}|}$为约束优化问题

$$\mathbf{min}_{\boldsymbol{x}}\ \{f(x): h_i(x)=0, i\in\mathcal{E};\ g_i(x)\geq 0, i\in\mathcal{I}\}$$

的Lagrange函数$L(x,\lambda,\mu)$的**鞍点**,若$\forall(x,\lambda,\mu)\in\mathbb{R}^n\times\mathbb{R}^{|\mathcal{E}|}\times\mathbb{R}_+^{|\mathcal{I}|}$有

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

$$\Updownarrow$$

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

对偶规划与对偶理论

强对偶定理

定理 设 $(\boldsymbol{x}^*, \boldsymbol{w}^*, \boldsymbol{v}^*)$ 为优化问题(P)的Lagrange函数的鞍点, 则$\boldsymbol{x}^*$与$(\boldsymbol{w}^*, \boldsymbol{v}^*)$分别是原问题与对偶问题的最优解且对偶间隙为零.

证明：$(\boldsymbol{w}^*, \boldsymbol{v}^*) \in \underset{\substack{\boldsymbol{v}\in\mathbb{R}^{|\mathcal{E}|} \\ \boldsymbol{w}\in\mathbb{R}_+^{|\mathcal{I}|}}}{\operatorname{argmax}} L(x^*, \boldsymbol{w}, \boldsymbol{v})$ $\Rightarrow$ $h_i(x^*) = 0, i \in \mathcal{E}$; $w_i^* \geq 0, g_i(x^*) \geq 0, w_i^* g_i(x^*) = 0, i \in \mathcal{I}$

$\Downarrow$

$x^* \in \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\ L(x, \boldsymbol{w}^*, \boldsymbol{v}^*)$ $\Leftarrow$ 对(P)的任一可行解x, $f(x) \geq L(x, \boldsymbol{w}^*, \boldsymbol{v}^*) \geq L(x^*, \boldsymbol{w}^*, \boldsymbol{v}^*) = f(x^*)$

最优解x^*

对偶理论

强对偶定理

定理 设 $(\boldsymbol{x}^*, \boldsymbol{w}^*, \boldsymbol{v}^*)$ 为优化问题**(P)**的**Lagrange**函数的鞍点, 则$\boldsymbol{x}^*$与$(\boldsymbol{w}^*, \boldsymbol{v}^*)$分别是原问题与对偶问题的最优解且对偶间隙为零.

证明：**(续)** $x^* \in \operatorname*{argmin}_{x\in\mathbb{R}^n} L(x, \boldsymbol{w}^*, \boldsymbol{v}^*) \Rightarrow \theta(w^*, v^*) = L(x^*, \boldsymbol{w}^*, \boldsymbol{v}^*)$

已证 $L(x^*, \boldsymbol{w}^*, \boldsymbol{v}^*)$为问题**(P)**的最优值$f(x^*)$，由弱对偶定理可知：

对**(D)**的任一可行解 (w, v),有 $\theta(w, v) \le L(x^*, \boldsymbol{w}^*, \boldsymbol{v}^*) = \theta(w^*, v^*)$

则**(D)**的最优解为$(\boldsymbol{w}^*, \boldsymbol{v}^*)$, 且$\boldsymbol{f}(\boldsymbol{x}^*) = \theta(w^*, v^*)$.

最优解 (w^*, v^*)

对偶规划与对偶理论

乘子的经济学解释

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

$\Downarrow$ **线性扰动**

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

$$\varepsilon=(\varepsilon_1,\cdots,\varepsilon_m)^T \quad \lambda=(\lambda_1,\cdots,\lambda_l)^T$$

- 设$\boldsymbol{x}^*$为局部最优解;
  $(\boldsymbol{w}^*,\boldsymbol{v}^*)$为相应的**Lagrange**乘子

$$\boldsymbol{x}^*(0,0)=\boldsymbol{x}^*$$
$$\left(\boldsymbol{w}^*(0),\boldsymbol{v}^*(0)\right)=(\boldsymbol{w}^*,\boldsymbol{v}^*)$$

- 设$\boldsymbol{x}^*(\boldsymbol{\varepsilon},\boldsymbol{\lambda})$为扰动问题的局部最优解;
  $\left(\boldsymbol{w}^*(\boldsymbol{\varepsilon}),\boldsymbol{v}^*(\boldsymbol{\lambda})\right)$为相应的**Lagrange**乘子

对偶规划与对偶理论

乘子的经济学解释

定理：设函数f,g_i,h_j二阶连续可微，x^*是(NP)问题的局部最优解，$\left(w^*,v^*\right)$是相应的Lagrange乘子向量，设$x^*\left(\varepsilon,\lambda\right)$是扰动问题的局部最优解,$\left(w^*(\varepsilon),v^*(\lambda)\right)$是相应的Lagrange乘子向量, 则有：

$$\nabla_\varepsilon f\left(x^*\left(\varepsilon,\lambda\right)\right)\Big|_{\varepsilon=0}=w^*,\nabla_\lambda f\left(x^*\left(\varepsilon,\lambda\right)\right)\Big|_{\lambda=0}=v^*.$$

对偶规划与对偶理论

例： 某企业预算以2千元作为广告费，根据以往的经验，若以x_1千元作广播广告，x_2千元作报纸广告，则销售金额为 $-2x_1^2-10x_2^2-8x_1x_2+18x_1+34x_2$（千元）

试问：(1) 如何分配2千元的广告费？ (2) 广告费预算作微小改变后的影响如何？

解：

$$\begin{aligned} \min\ & f(x)=2x_1^2+10x_2^2+8x_1x_2-18x_1-34x_2 \\ \text{s.t.}\ & x_1+x_2-2=0 \\ & x_1\ge 0,\ x_2\ge 0 \end{aligned}$$

KKT条件为：

$$\begin{cases} 4x_1+8x_2-18-w_1-v=0 \\ 20x_2+8x_1-34-w_2-v=0 \\ x_1w_1=0,\ x_2w_2=0, \\ x_1+x_2-2=0 \\ x_1\ge 0,\ x_2\ge 0 \\ w_1\ge 0,\ w_2\ge 0 \end{cases} \Longrightarrow \begin{aligned} x^* &= (1,1)^T \\ w^* &= (0,0)^T \\ v^* &= -6 \end{aligned}$$

对偶规划与对偶理论

若广告费作微小改动，则相应的扰动问题为

$$
\begin{aligned}
\min\ & f(x) = 2x_1^2 + 10x_2^2 + 8x_1x_2 - 18x_1 - 34x_2 \\
\text{s.t.}\ & x_1 + x_2 - 2 = e \\
& x_1 \geq 0,\ x_2 \geq 0
\end{aligned}
$$

KKT点

$$x^*(e) = \left(1 + \frac{3}{2}e, 1 - \frac{e}{2}\right)^T$$

$$w^*(e) = (0, 0)^T$$

$$v^*(e) = 2e - 6$$

$$f\left(x^*(e)\right) = e^2 - 6e - 32$$

则有 $\frac{df(x^*(e))}{de}\Big|_{e=0} = v^* = -6$

经济学含义：当e增加时，$-f\left(x^*(e)\right)$上升，即当广告费增加后，销售也随着增加，而且销售金额的增加大约为广告费增量的6倍，可见适当增加广告费的预算是有利的。

对偶规划与对偶理论

乘子的经济学解释: 影子价格

- 若把原问题约束条件看成是广义的资源约束, 则右端项的值表示每种资源的可用量.
- 对偶解的经济含义:资源的单位改变量引起目标函数值的增加量.
- 通常称对偶解为**影子价格**.
- 影子价格的大小客观地反映了资源在系统内的稀缺程度. 资源的影子价格越高, 说明资源在系统内越稀缺, 而增加该资源的供应量对系统目标函数值贡献越大.

影子价格的作用

（1）告诉管理者增加何种资源对企业更有利

（2）告诉管理者花多大代价购买进资源或卖出资源是合适的

（3）为新产品定价提供依据