最优化理论与方法

研究生学位课

陈军华（副教授、主任）

- **基本思路**
  - 对于不等式约束，可行区域在内部或边界。找到一个在可行域内的出发点，如果步长过长，会使其到达边界，或离开可行域。故，在优化过程中，考虑适当步长使其停留在可行域内。
  - 最简单的实现方式是：若沿搜索方向前进，不出可行域则正常搜索，如沿搜索方向离开了可行域，则沿破坏了那个约束条件的梯度方向前进，直到重新进入可行域。
  - 此法称“刺绣法”(hemstitching)

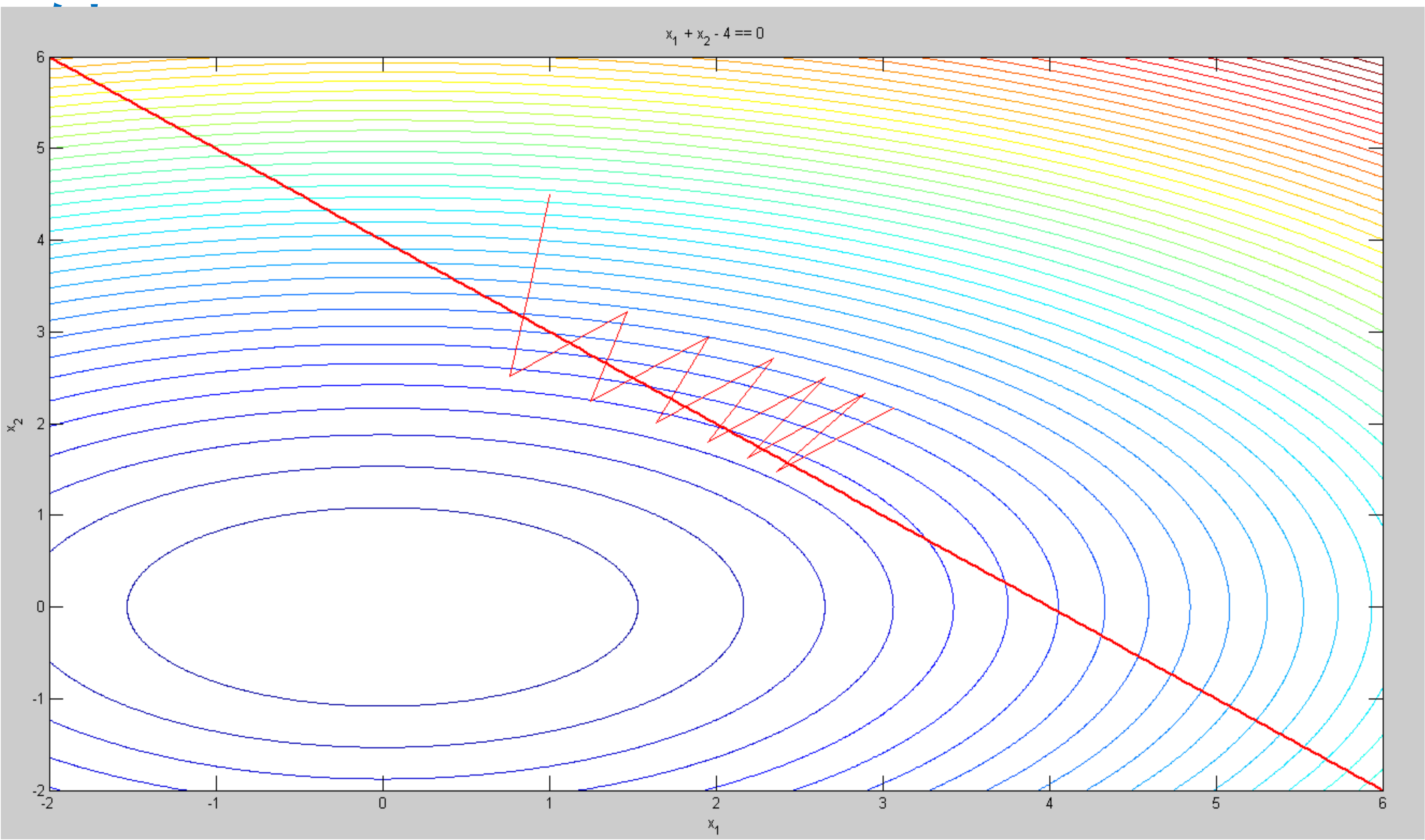
- **“刺绣法”（hemstitching）算法**
  - $\boldsymbol{min f(x)}, \quad \boldsymbol{s.t.} \quad \boldsymbol{g(x) \geq 0}.$　已知：$\boldsymbol{x^0, h, \delta}$，求 $\boldsymbol{x^*}$
  - **step1.**求归一化的 $\boldsymbol{\nabla f(x), \nabla g(x)}$
  - **step2.** ***while***$\boldsymbol{(\|x^{p+1} - x^p\| > \delta)}$

    if $\boldsymbol{x^p}$ 在可行域内部，则：$\boldsymbol{x^{p+1} = x^p - h\nabla f(x^p)}$

    if $\boldsymbol{x^p}$ 在可行域外部，则：$\boldsymbol{x^{p+1} = x^p + h\nabla g(x^p)}$

- 例 7_1
  - $\boldsymbol{minf}(\boldsymbol{x}) = x_1^2 + \mathbf{2}x_2^2$
  - $\boldsymbol{s.t.} \quad \boldsymbol{x_1} + \boldsymbol{x_2} \geq \mathbf{4}$
  - 设初始点为（**1,4.5**），固定步长为**1.0**

- **例 7_1**
  - $\min f(x) = x_1^2 + 2x_2^2$
  - $s.t. \quad x_1 + x_2 \geq 4$
  - 设初始点为（1,4.5），固定步长为1.0
  - 归一化后的梯度为：
  - $\nabla f(x) = \frac{1}{\sqrt{4x_1^2+16x_2^2}}\begin{bmatrix}2x_1\\4x_2\end{bmatrix}, \nabla g(x) = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$
  - $if\ \ x^p$ 在可行域内部，则： $x^{p+1} = x^p - 1.0\nabla f(x^p)$
  - $if\ \ x^p$ 在可行域外部，则： $x^{p+1} = x^p + 1.0\nabla g(x^p)$
  - (1,4.5)→(0.89,3.5)→(0.76,2.50)→(1.47,3.21)→(1.05,2.30)→(1.76,3.01)→(1.37,2.10)→(2.08,2.81)



- 刺绣法的效率并不高。目标函数值不是每一步都向优。

- **改进的刺绣法**
  - 算法在处理出可行域时，同时考虑目标函数优化和可行域偏差。具体实现是：
  - $x^{p+1} = x^p + k_p d^p$
  - $if\ \ x^p$ 在可行域内部，则： $d^p = -\frac{\nabla f(x^p)}{\|\nabla f(x^p)\|}$
  - $if\ \ x^p$ 在可行域外部，则：

$$d^p = -\frac{\nabla f(x^p)}{\|\nabla f(x^p)\|} + \sum_{g_i\ (x)\ violate\ at\ x^p} \frac{\nabla g_i(x^p)}{\|\nabla g_i(x^p)\|}$$

  - 当两个梯度方向正好相反时，效率会低。

- **例 7_2**
  - 写出下面规划问题的刺绣法的递推迭代公式：
  - $\boldsymbol{minf}(\boldsymbol{x}) = x_1^2 + x_{\mathbf{2}}^2$
  - $\boldsymbol{s.t.} \quad \mathbf{3}\boldsymbol{x}_{\mathbf{1}} + \mathbf{2}\boldsymbol{x}_{\mathbf{2}} \geq \mathbf{6}$
  - $\quad \boldsymbol{x}_{\mathbf{1}}, \boldsymbol{x}_{\mathbf{2}} \geq \mathbf{0}$
  - 起始点：(2,3), 步长为$\boldsymbol{k}_{\boldsymbol{p}}$

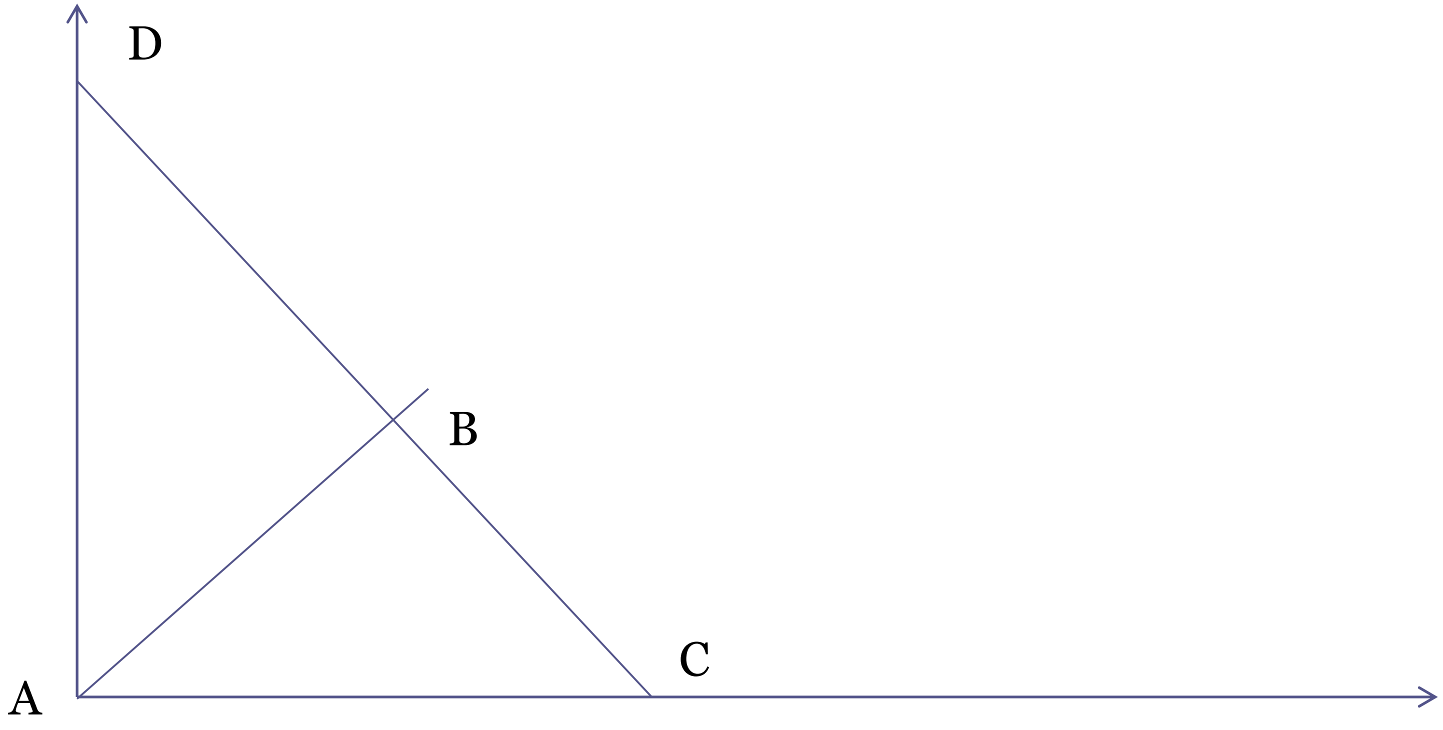
- 例 7_2
  - $\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}, \nabla g(x) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$
  - $x^{p+1} = \begin{cases} x^p - \dfrac{k_p}{\sqrt{4x_1^2+4x_2^2}} \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} \\ x^p + \dfrac{k_p}{\sqrt{13}} \begin{bmatrix} 3 \\ 2 \end{bmatrix} \end{cases}$
  - $x^* = \left(\dfrac{18}{13}, \dfrac{12}{13}\right)$

- 例 7_3
  - $\max f(x) = x_1^2 + x_2^2$

  $s.t. \quad 6 - 3x_1 - 2x_2 \geq 0$

  $x_1, x_2 \geq 0$
  - 起始点：(0,0)

- **例 7_3**
  - $\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}, \rightarrow \nabla f(x_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
  - 若出发点在$x_2 = \frac{2}{3}x_1 \rightarrow \left(\frac{18}{13}, \frac{12}{13}\right)$; 若在ABC内→(2,0); 在ABD内→(0,3)



- **坐标轮换法**

坐标轮换法–算法步骤

- 1.给定初始点x^0，初始步长δ，缩放因子$u>1$及精度$\varepsilon_1,\varepsilon_2$，置$m=0$
- 2.令$y=x^k$
- 3.从y出发，按步长$t=\delta_1^0$沿坐标轴e^1的正方向搜索，令$y^1=y+te^1$，如果y^1在可行域内，即满足$g(y^1)\geq 0$，且有$f(y^1)<f(y)$,则取$t=2t$，加速向前搜索直到不满足可行性条件或函数值下降的条件，转5.

  如果沿坐标轴 e^1的正方向找不到同时满足两条件的点，则转向负方向搜索，再转5.
- 4.如果沿坐标轴 e^1的正方向和负方向都找不到同时满足可行性条件和函数值下降的点，则减小搜索步长 $\delta^0=\delta^0/u$

- **坐标轮换法–算法步骤**
  - 5.依次沿其他坐标轴进行同样的搜索，最终求得此轮搜索的最优点$\boldsymbol{y}^{*}$
  - 6.如果$\left\|\boldsymbol{y}^{*}-\boldsymbol{x}^{k}\right\| \leq \boldsymbol{\varepsilon}_{2}$，转7，否则$\boldsymbol{x}^{k+1}=\boldsymbol{y}^{*}, \boldsymbol{k}=\boldsymbol{k}+\mathbf{1}$,转2.
  - 7.如果步长$\boldsymbol{max}(\boldsymbol{\delta}^{0}) \leq \boldsymbol{\varepsilon}_{1}$,则$\boldsymbol{x}^{*}=\boldsymbol{x}^{k}$,否则$\boldsymbol{\delta}^{0}=\boldsymbol{\delta}^{0}/\boldsymbol{u}$，转2.

- **例 7_4**
  - 用坐标轮换法求解约束优化问题
  - $\min f(s,t) = s^2 + 2t^2 - 4s - 8t + 15$
  - $s.t. \quad 9 - s^2 - t^2 \geq 0$
  - $\qquad s \geq 0, t \geq 0$
  - $(s,t) = (1,1), \delta = 0.1$,缩放因子1.5，步长精度取0.001

- **例 7_4**
  - 步骤1.1 沿$\boldsymbol{e}_1 = (1;0)$方向探测：
  - $\boldsymbol{y}_0 = \boldsymbol{x}_0 = (1;1), f(\boldsymbol{y}_0) = 6$

    $\boldsymbol{y}_1 = \boldsymbol{y}_0 + \delta \boldsymbol{e}_1 = (1.1;1)$

$$\because\ f(\boldsymbol{y}_1) = 5.81 < f(\boldsymbol{y}_0)\ \&\ g(\boldsymbol{y}_1) = \begin{pmatrix} 6.79 \\ 1.1 \\ 1 \end{pmatrix} \geq 0$$

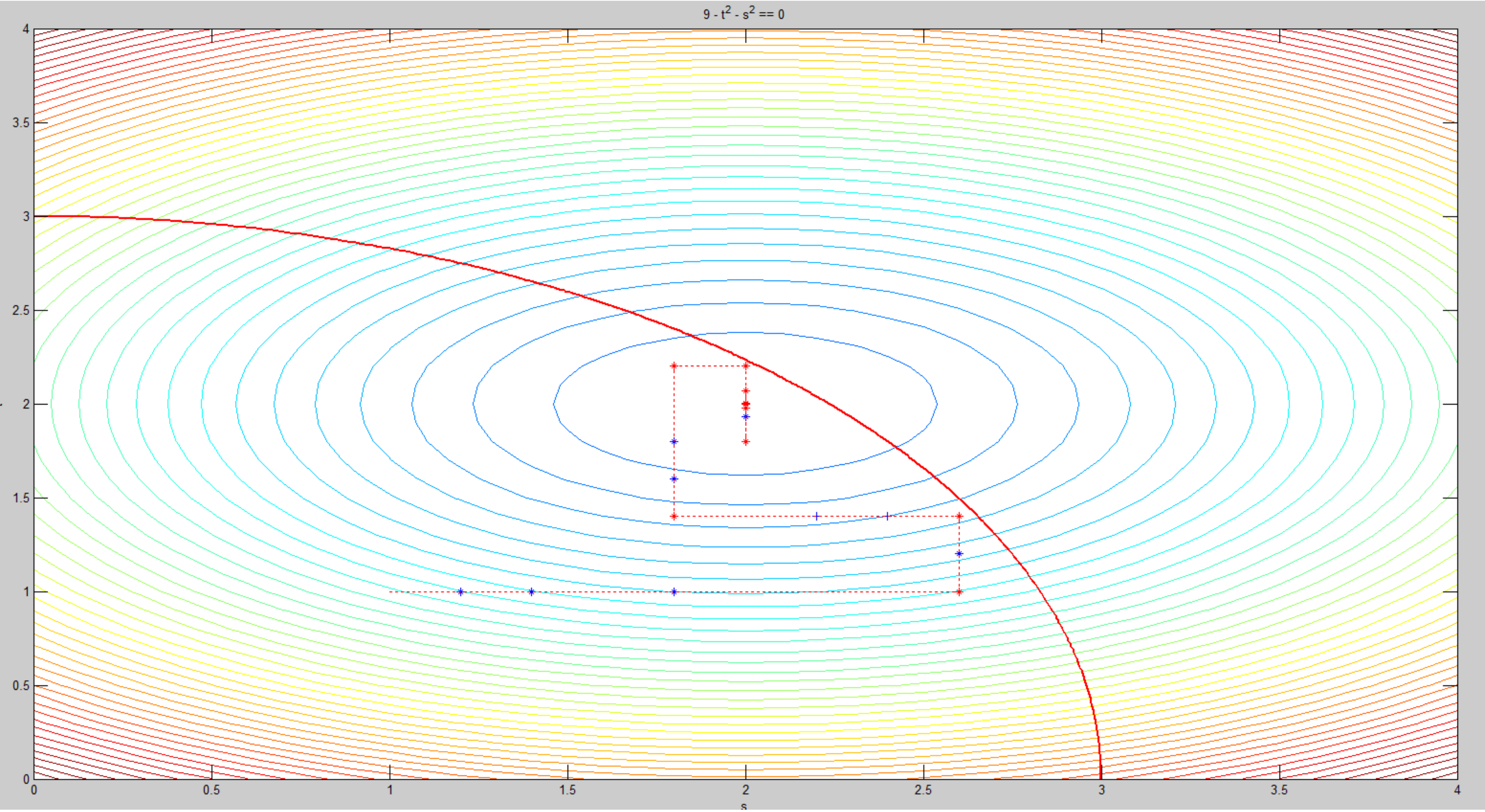
$\therefore$搜索到可行点，可加大步长 $\delta = 2\delta = 0.2$

$$\boldsymbol{y}_2 = \boldsymbol{y}_0 + \delta \boldsymbol{e}_1 = (1.2;1)$$

$$\because\ f(\boldsymbol{y}_2) = 5.64 < f(\boldsymbol{y}_0)\ \&\ g(\boldsymbol{y}_1) = \begin{pmatrix} 6.56 \\ 1.2 \\ 1 \end{pmatrix} \geq 0$$

继续加大步长 $\delta = 2\delta = 0.4$ ……

- **例 7_4**



- **有约束问题梯度算法（可行方向法）**
  - $\min f(x)$

    $s.t.\quad g_i(x) \le 0 \qquad i = 1,2,\ldots,m$
  - 对两式进行一阶Taylor展开：

$$f(x^p + d^p) \approx f(x^p) + (\nabla f(x^p), d^p)$$

$$g_i(x^p + d^p) \approx g_i(x^p) + (\nabla g_i(x^p), d^p)$$

  - 要使解优化，则：$f(x^p + d^p) - f(x^p) \approx (\nabla f(x^p), d^p) < 0$
  - 若x^p为内点（严格在可行域内），那么$\exists k_p > 0$,使

$x^{p+1} = x^p + k_p d^p$ 为可行解，d^p 为任意方向。

  - 若x^p在边界 $g_i(x^p) = 0$，那么x^{p+1}对于部分 $k_p > 0$，并满足：

    $g_i(x^p) + (\nabla g_i(x^p), d^p) < 0$,可在可行域内。

- **有约束问题梯度算法（可行方向法）**
  - 为寻求合适 d^p，使得点 x^{p+1}为可行解且向最优方向。引入松弛变量$x_0 \geq 0$，并得到如下线性规划：

$$\max_{d^p} x_0$$

$$s.t. \qquad x_0 + (\nabla f(x^p), d^p) \leq 0$$

$$x_0 + (\nabla g_i(x^p), d^p) + g_i(x^p) \leq 0 \quad i = 1, \ldots, m$$

$$\left| d_j^p \right| \leq 1$$

若 $x_0^* = 0$，约束问题d^p达到最优。

- **可行方向法算法步骤**
  - 1：计算$\nabla f(x^p), \nabla g_i(x^p), i = 1, \ldots, m$
  - 2：解线性规划问题$\max\limits_{d^p} x_0$

$$s.t. \quad x_0 + (\nabla f(x^p), d^p) \le 0$$

$$x_0 + (\nabla g_i(x^p), d^p) \le -g_i(x^p) \quad i = 1, \ldots, m$$

$$(\nabla g_i(x^p), d^p) \le 0 \quad i = 1, \ldots, m \text{ (可如此简化)}$$

$$\left| d_j^p \right| \le 1$$

  - 3：当$x_0 = 0$时停止
  - 4：求得合适步长$k_p > 0$,并保证 $\min f\left(x^p + k_p d^p\right)$

例 7_5

- $\boldsymbol{minf}(\boldsymbol{x}) = x_1^2 + \mathbf{2}x_2^2$
- $\boldsymbol{s.t.} \qquad \boldsymbol{x_1} + \boldsymbol{x_2} \geq \mathbf{4}$
- $\overline{\boldsymbol{x}} = \begin{pmatrix} \mathbf{0.85} \\ \mathbf{3.15} \end{pmatrix}$

- **例 7_5**
  - $\boldsymbol{min f(x)} = x_1^2 + 2x_2^2$
  - $\boldsymbol{s.t.} \quad \boldsymbol{x_1 + x_2 \geq 4}$

$$\max_{dp} x_0$$

$$s.t. \quad x_0 + (\nabla f(\overline{x}), d) \leq 0$$

$$(\nabla g(\overline{x}), d) \leq 0$$

$$|d_1| \leq 1, |d_2| \leq 1$$

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 2x_2 \end{pmatrix}, \nabla f(\overline{x}) = \begin{pmatrix} 0.85 \\ 6.30 \end{pmatrix}$$,归一化为: $$\begin{pmatrix} 0.134 \\ 0.991 \end{pmatrix}$$

  - $\nabla g(\overline{x}) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$,归一化为: $\begin{pmatrix} -0.707 \\ -0.707 \end{pmatrix}$

- **例 7_5**
  - $minf(x) = x_1^2 + 2x_2^2$
  - $s.t. \quad x_1 + x_2 \geq 4$

$$\max_{d^p} x_0$$

$$s.t. \quad x_0 + 0.134d_1 + 0.991d_2 \leq 0$$

$$-0.707d_1 - 0.707d_2 \leq 0$$

$$|d_1| \leq 1, |d_2| \leq 1$$

可得：$d_2 = -0.5$，$d_1 = 1.\ x_0 = 0.357$

- **例 7_5**
  - 再考查一个点：$\bar{x} = \begin{pmatrix} 2.667 \\ 1.333 \end{pmatrix}$

$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 2x_2 \end{pmatrix}, \nabla f(\bar{x}) = \begin{pmatrix} 2.667 \\ 2.666 \end{pmatrix}$,归一化为$\begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix}$

  - $\nabla g(\bar{x}) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$,归一化为：$\begin{pmatrix} -0.707 \\ -0.707 \end{pmatrix}$
  - $\max\limits_{d^p} x_0$

$$s.t.\quad x_0 + 0.707d_1 + 0.707d_2 \le 0$$

$$-0.707d_1 - 0.707d_2 \le 0$$

$$|d_1| \le 1, |d_2| \le 1$$

$$d_1 = 1，\ d_2 = -1, x_0 = 0$$