要点：共轭梯度方向

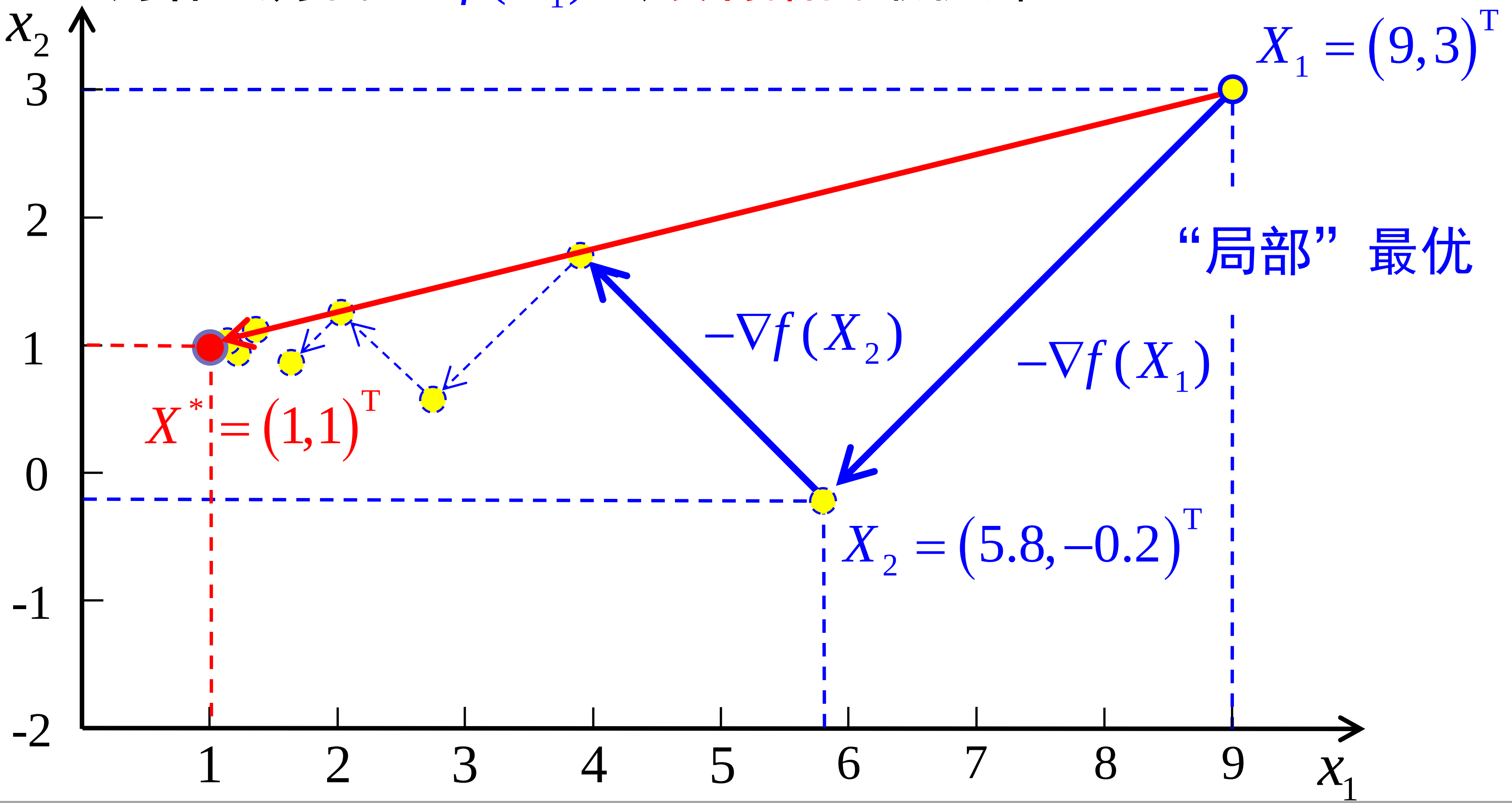
改进梯度下降法的思路

柯西：梯度下降法

$$X_{k+1} = X_k - \lambda_k^* \nabla f(X_k)$$

$$\min f(X) = x_1^2 + 4x_2^2 - 2x_1 - 8x_2 + 5$$

为什么方向 $-\nabla f(X_1)$ 会没有指向最优解 X^*？

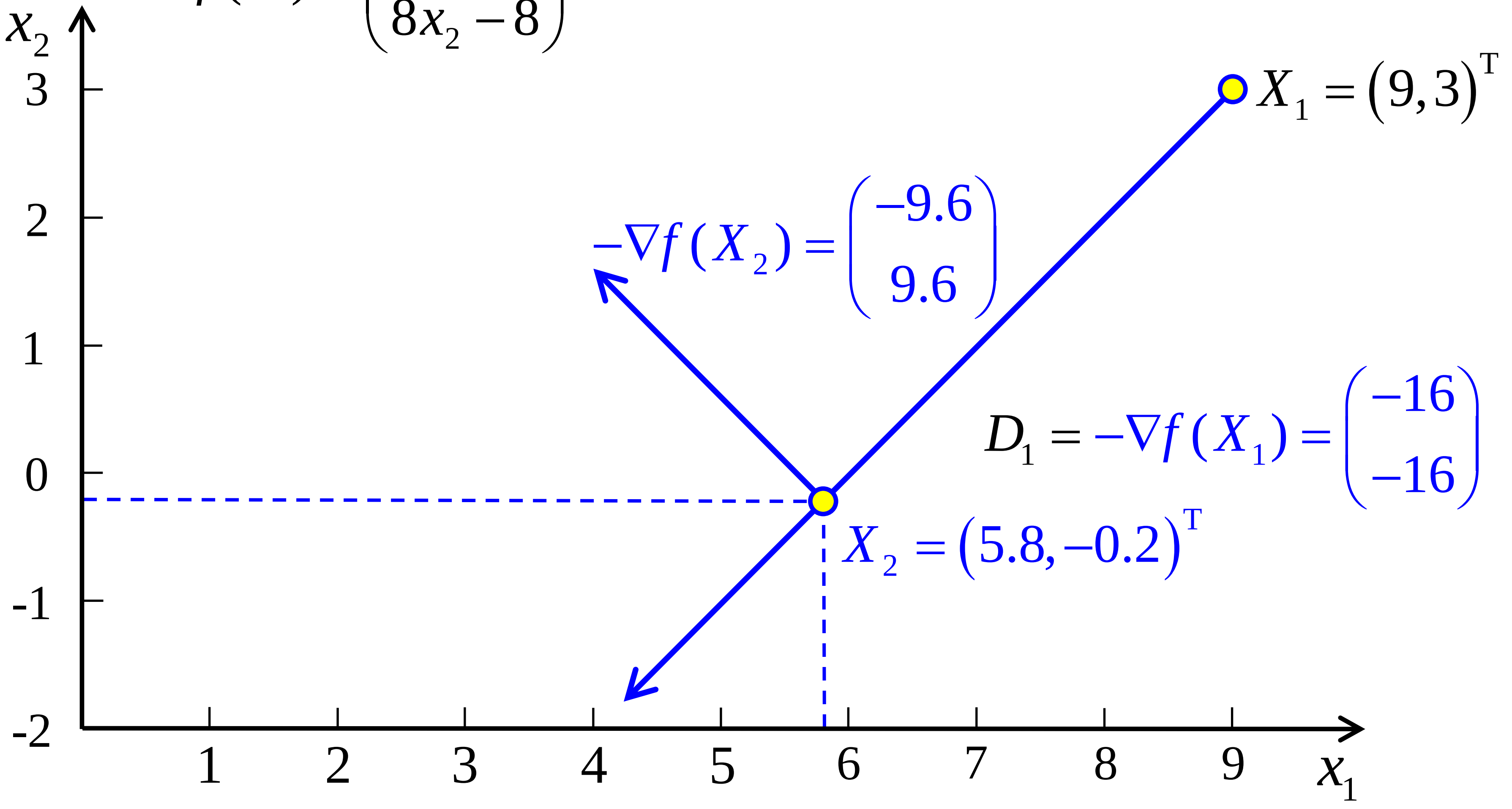


F-R 共轭梯度法 —— 利用共轭梯度寻优

$$\min f(X) = x_1^2 + 4x_2^2 - 2x_1 - 8x_2 + 5$$

第一步沿负梯度寻优

$$\nabla f(X) = \begin{pmatrix} 2x_1 - 2 \\ 8x_2 - 8 \end{pmatrix}$$

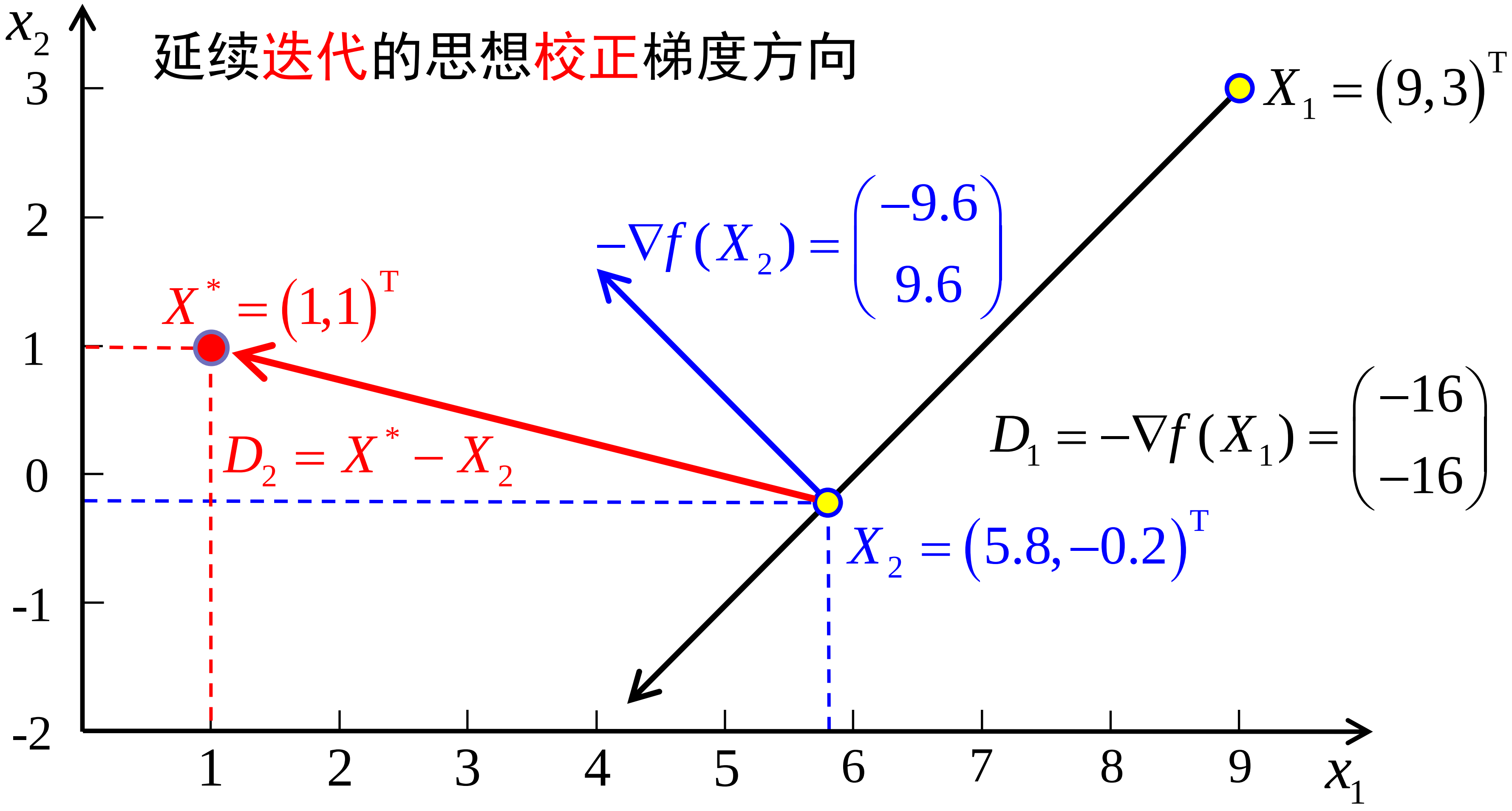


F-R 共轭梯度法 —— 利用共轭梯度寻优

$$\min f(X) = x_1^2 + 4x_2^2 - 2x_1 - 8x_2 + 5$$

$$D_2 = -\nabla f(x_2) + \alpha_1 D_1 \qquad X_{k+1} = X_k + \lambda_k D_k$$

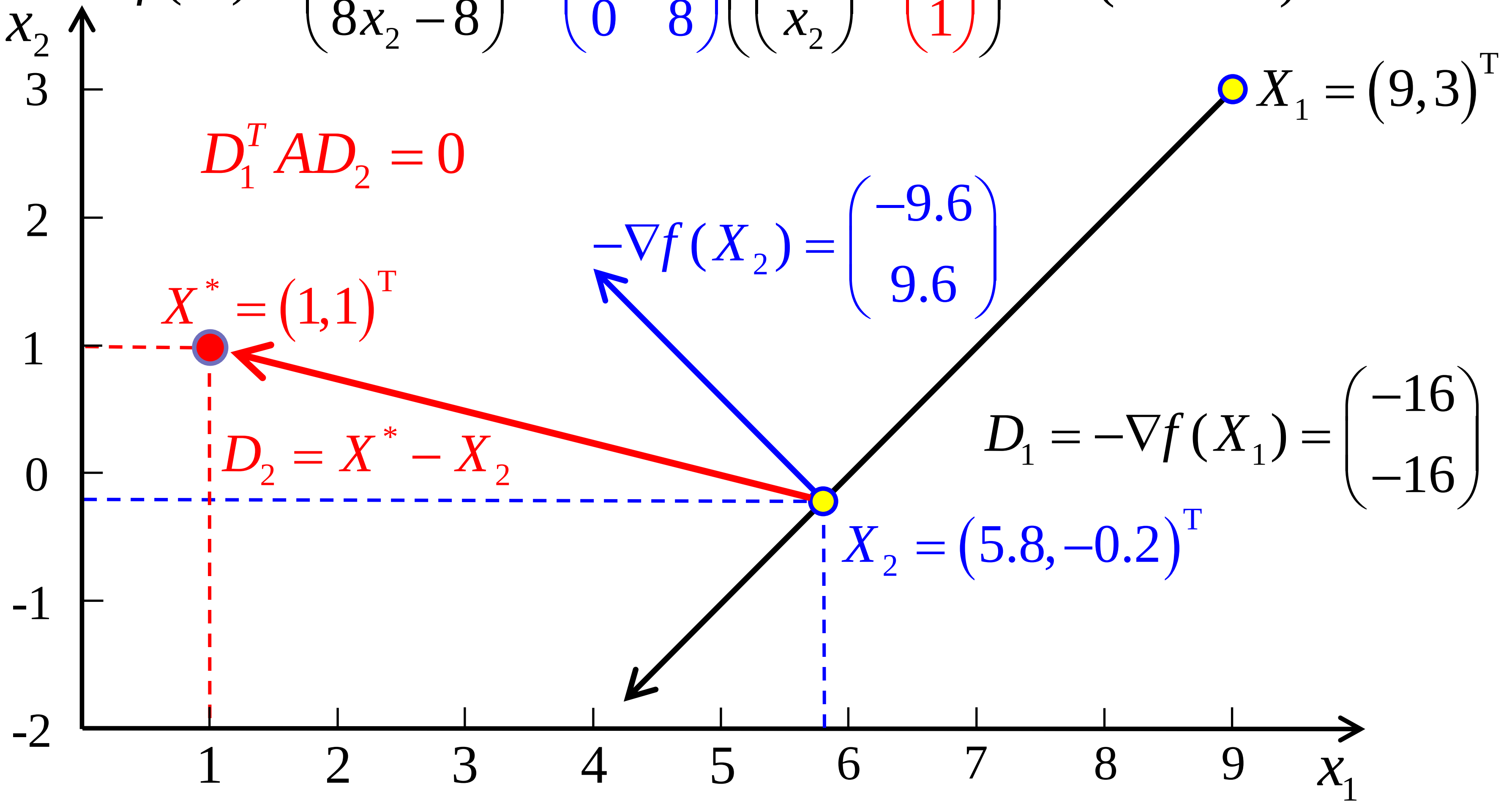
延续迭代的思想校正梯度方向



F-R 共轭梯度法 —— 利用共轭梯度寻优

$$\min f(X)=x_1^2+4x_2^2-2x_1-8x_2+5$$

$$\nabla f(X)=\begin{pmatrix}2x_1-2\\8x_2-8\end{pmatrix}=\begin{pmatrix}2&0\\0&8\end{pmatrix}\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}-\begin{pmatrix}1\\1\end{pmatrix}\right)=A(X-X^*)$$



共轭方向法原理之一

共轭方向定义： $A \in R^{n\times n}$ 对称矩阵， $\vec{p},\vec{q} \in R^n$ 非零向量，若 $\vec{p}^T A\vec{q} = 0$ ，称 $\vec{p},\vec{q}$ 为 A 共轭方向

共轭方向线性无关性

若 $\vec{p}_0, \vec{p}_1, \cdots, \vec{p}_{n-1}$ 互为 $\boxed{A > 0}$ 的共轭方向，则它们线性无关

理由： $\alpha_0 \vec{p}_0 + \alpha_1 \vec{p}_1 + \cdots + \alpha_{n-1} \vec{p}_{n-1} = 0$

$$\Rightarrow \quad \alpha_0 \vec{p}_k^T A\vec{p}_0 + \alpha_1 \vec{p}_k^T A\vec{p}_1 + \cdots + \alpha_{n-1} \vec{p}_k^T A\vec{p}_{n-1} = 0$$

$$\Rightarrow \quad \alpha_k \vec{p}_k^T A\vec{p}_k = 0$$

$$\Rightarrow \quad \alpha_k = 0$$

共轭梯度方向 $D_k = -\nabla f(X_k) + \alpha_{k-1} D_{k-1}$

对于二次正定函数 $f(X) = \frac{1}{2}(X - X^*)^T A(X - X^*)$

$D_1 = -\nabla f(X_1)$

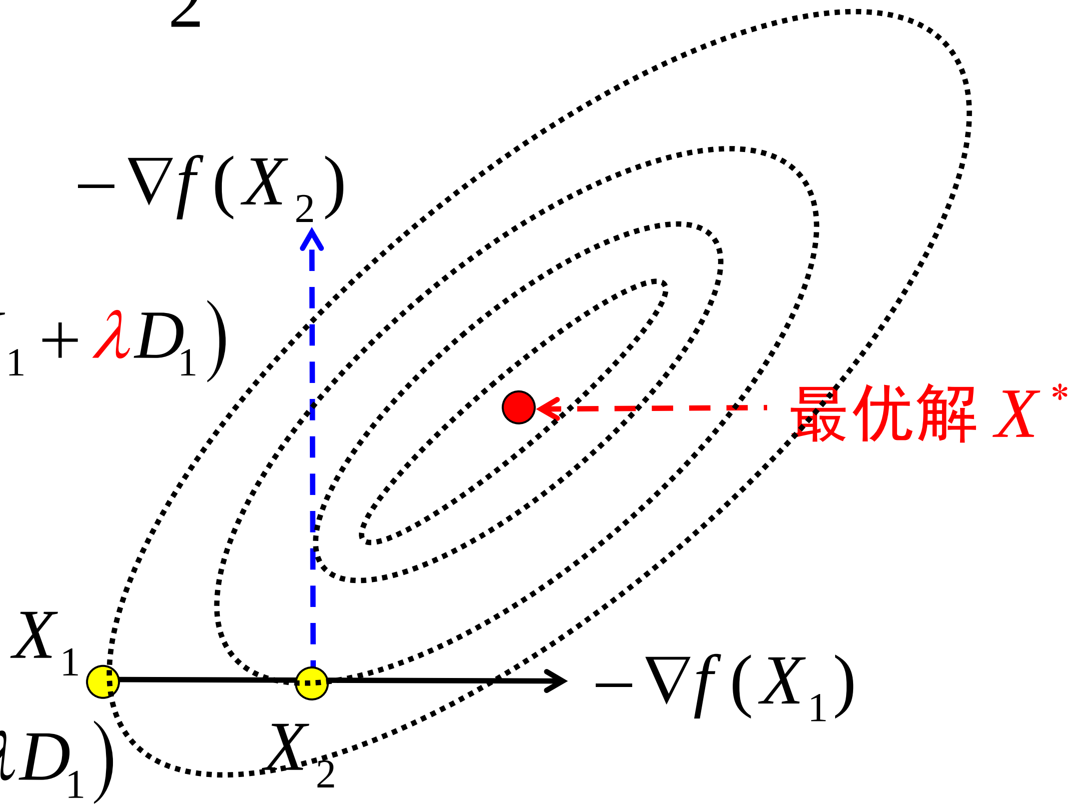
$X_2 = X_1 + \lambda_1^* D_1$

λ_1^* 是优化问题 $\min\limits_{\lambda>0} f(X_1 + \lambda D_1)$ 的最优解

$$\frac{\mathrm{d}\, f(X_1 + \lambda D_1)}{\mathrm{d}\lambda}$$

$$= \frac{\mathrm{d}\, f(X_1 + \lambda D_1)}{\mathrm{d} X^T} \frac{\mathrm{d}(X_1 + \lambda D_1)}{\mathrm{d}\lambda}$$

$$\Rightarrow \nabla^T f(X_1 + \lambda_1^* D_1) D_1 = 0 \quad \Rightarrow \quad \nabla^T f(X_2) D_1 = 0$$



$$f(X)=\frac{1}{2}(X-X^*)^T A(X-X^*) \qquad D_k=-\nabla f(X_k)+\alpha_{k-1}D_{k-1}$$

$$D_1=-\nabla f(X_1)$$

$$D_2=-\nabla f(X_2)+\alpha_1 D_1$$

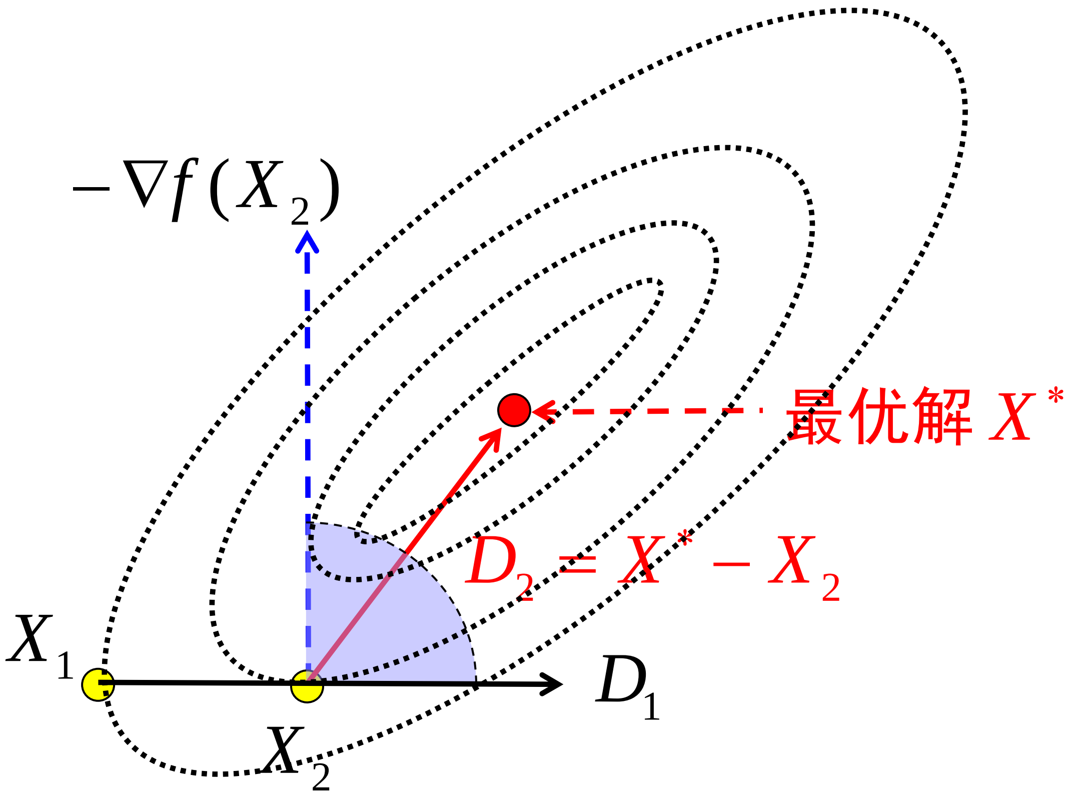
$$\nabla f(X_2)=A(X_2-X^*)$$

$$=-AD_2$$

$$D_1^T\nabla f(X_2)=0$$

$$\Rightarrow \quad -D_1^T AD_2=0$$

$$\Rightarrow \quad D_1^T AD_2=0$$



D_1与 D_2为 A 的共轭方向！

要点：F-R共轭梯度法

F-R 共轭梯度法 —— 利用共轭梯度寻优

$$\min f(X) = x_1^2 + 4x_2^2 - 2x_1 - 8x_2 + 5$$

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}$$

“长记性”寻优方向 D_2 和 D_1 是两个 A 的共轭方向

$$D_2 = -\nabla f(x_2) + \alpha_1 D_1$$

共轭梯度

只需要解决如何计算出合适的参数 α_1

1952年Hestenes和Stiefel提出利用共轭梯度

求解线性方程组 $AX = b,\ X \in \mathrm{R}^n$

$$\min\left(X^{\mathrm{T}}AX - b^{\mathrm{T}}X\right),\ X \in \mathrm{R}^n$$

Fletcher：用“简单”解决“复杂”

R. Fletcher 英国

皇家科学院院士

F-R 共轭梯度法 —— F-R共轭梯度法

1964年，Fletcher 和 Reeves提出了适用于一般无约束最优化问题的求解方法：F-R 共轭梯度法

$$X_{k+1} = X_k + \lambda_k^* D_k$$

$$D_k = \begin{cases} -\nabla f(X_k) & k = 1 \\ -\nabla f(X_k) \boxed{+ \alpha_{k-1} D_{k-1}} & k \geq 2 \end{cases}$$

梯度下降法

$$X_{k+1} = X_k + \lambda_k^* D_k$$

$$D_k = -\nabla f(X_k)$$

相邻两步寻优方向共轭性 $D_k^{\mathrm{T}} A D_{k-1} = 0$ 和精确搜索的特点

$$\alpha_k = \frac{\left\|\nabla f(X_{k+1})\right\|^2}{\left\|\nabla f(X_k)\right\|^2}$$

F-R 共轭梯度法计算简单、寻优速度快，在国际上开启了共轭梯度法求解非线性规划的研究先河！

要点：参数 α 的计算

$$f(X)=\frac{1}{2}(X-X^*)^T A(X-X^*) \qquad D_k=-\nabla f(X_k)+\alpha_{k-1}D_{k-1}$$

$$D_1=-\nabla f(X_1)$$

$$D_2=-\nabla f(X_2)+\alpha_1 D_1$$

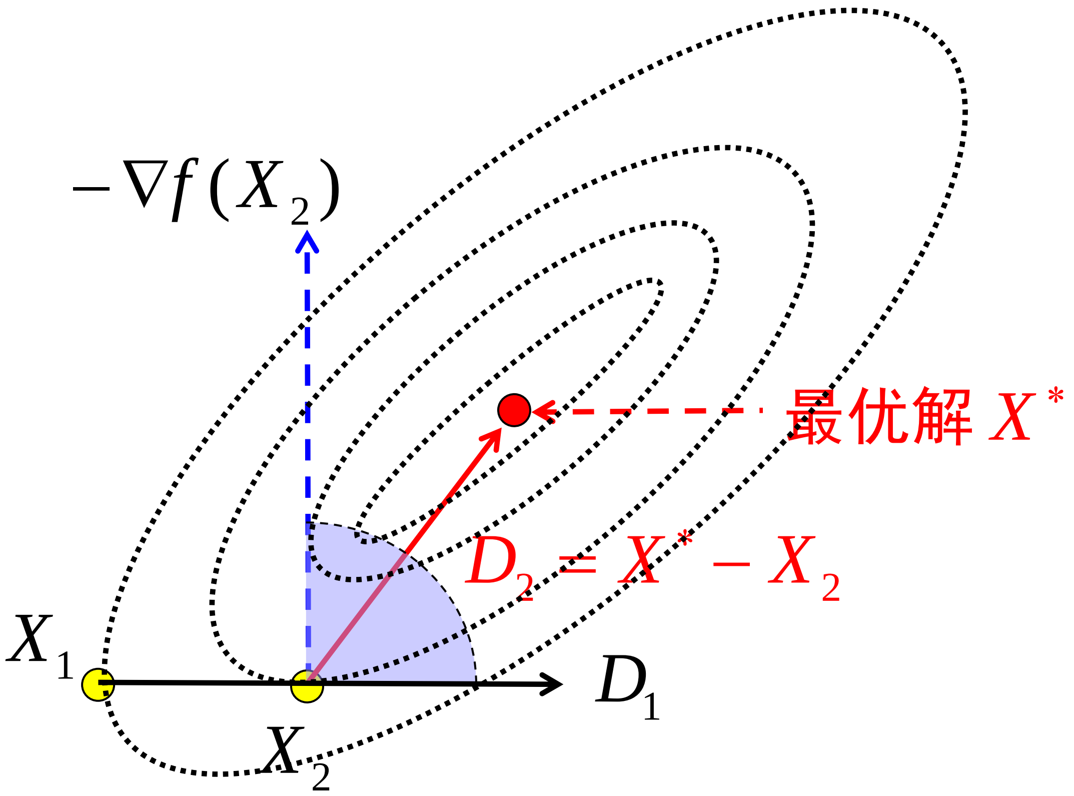
$$\nabla f(X_2)=A(X_2-X^*)$$

$$=-AD_2$$

$$D_1^T\nabla f(X_2)=0$$

$$\Rightarrow \quad -D_1^T AD_2=0$$

$$\Rightarrow \quad D_1^T AD_2=0 \quad \Rightarrow \quad -D_1^T A\nabla f(X_2)+\alpha_1 D_1^T AD_1=0$$



$$\alpha_1=\frac{D_1^T A\nabla f(X_2)}{D_1^T AD_1}$$

D_1与 D_2为 A 的共轭方向！

参数 α 中矩阵 A 的消除方法

由 $X_2 = X_1 + \lambda_1^* D_1 \Rightarrow D_1 = (X_2 - X_1) / \lambda_1^*$

$$\alpha_1 = \frac{D_1^T A \nabla f(X_2)}{D_1^T A D_1} = \frac{\nabla^T f(X_2) A D_1}{D_1^T A D_1}$$

$$= \frac{\nabla^T f(X_2) A (X_2 - X_1) / \lambda_1^*}{D_1^T A (X_2 - X_1) / \lambda_1^*} = \frac{\nabla^T f(X_2) A (X_2 - X_1)}{D_1^T A (X_2 - X_1)}$$

$$= \frac{\nabla^T f(X_2)\left(\nabla f(X_2) - \nabla f(X_1)\right)}{D_1^T \left(\nabla f(X_2) - \nabla f(X_1)\right)} = \frac{\nabla^T f(X_2)\left(\nabla f(X_2) + D_1\right)}{D_1^T \left(\nabla f(X_2) + D_1\right)}$$

$$= \frac{\nabla^T f(X_2) \nabla f(X_2)}{D_1^T D_1} = \frac{\left\|\nabla f(X_2)\right\|^2}{\left\|\nabla f(X_1)\right\|^2}$$

要点：F-R共轭梯度法计算示例

F-R 共轭梯度法 —— 寻优速度对比

$$\boxed{X_{k+1} = X_k + \lambda_k^* D_k}$$

$$\min f(X) = x_1^2 + 4x_2^2 - 2x_1 - 8x_2 + 5$$

$$D_k = \begin{cases} \boxed{-\nabla f(X_k) \quad k=1} \\ \boxed{-\nabla f(X_k) + \alpha_{k-1} D_{k-1} \quad k \ge 2} \end{cases}, \quad \boxed{\alpha_k = \frac{\|\nabla f(X_{k+1})\|^2}{\|\nabla f(X_k)\|^2}}$$

$$X_1 = (9,3)^{\mathrm{T}}$$

F-R 法计算步骤

① $D_1 = -\nabla f(X_1)$

② $\min\limits_{\lambda_1>0} f(X_1 + \lambda_1 D_1)$, $X_2 = X_1 + \lambda_1^* D_1$

③ $\alpha_1 = \dfrac{\|\nabla f(X_2)\|^2}{\|\nabla f(X_1)\|^2}$

④ $D_2 = -\nabla f(X_2) + \alpha_1 D_1$

⑤ $\min\limits_{\lambda_2>0} f(X_2 + \lambda_2 D_2)$, $X_3 = X_2 + \lambda_2^* D_2$

计算结果

$D_1 = -(16,16)^{\mathrm{T}}$

$\lambda_1^* = 0.2$, $X_2 = (5.8, -0.2)^{\mathrm{T}}$

$\nabla f(X_2) = (9.6, -9.6)^{\mathrm{T}}$

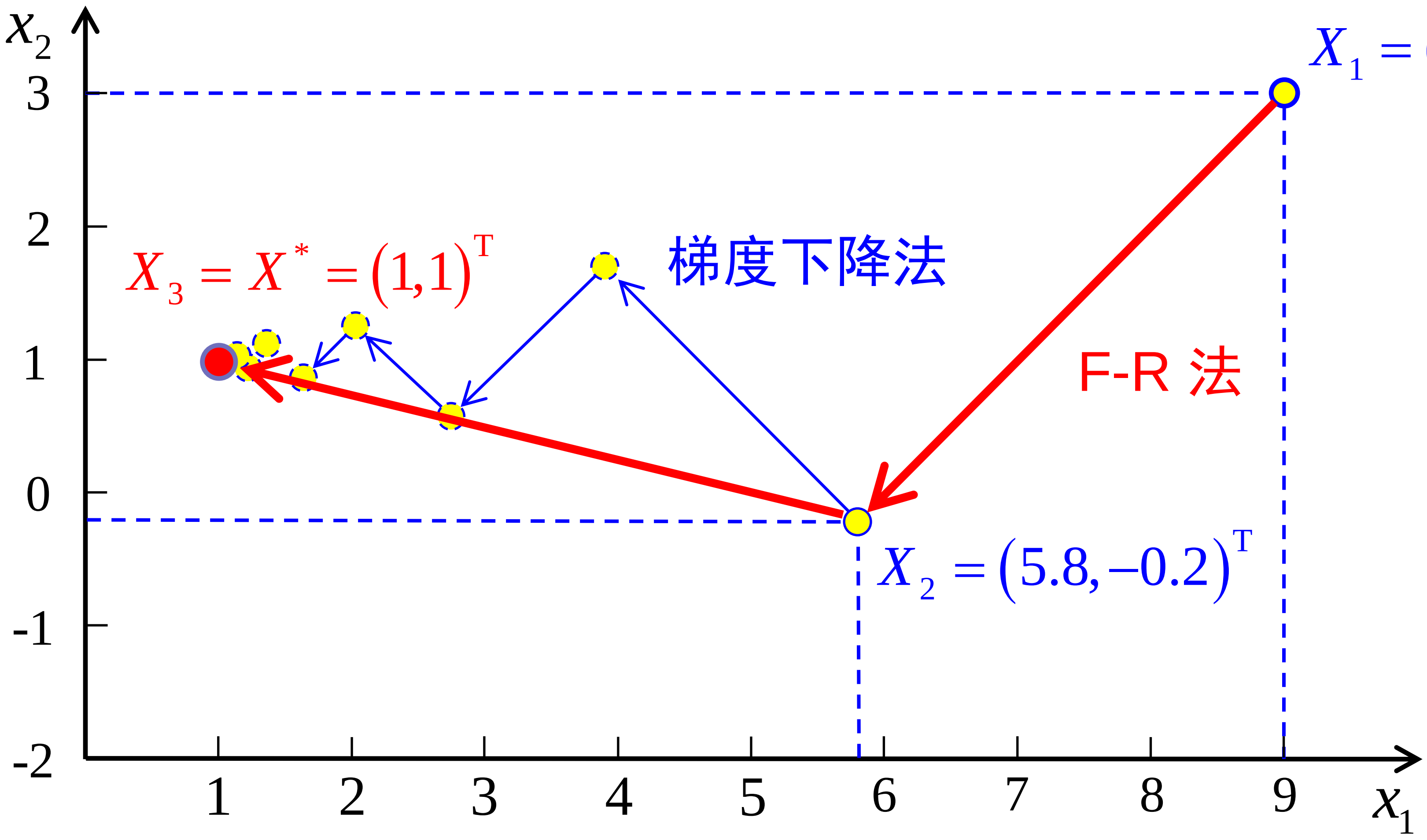
$\alpha_1 = 0.36$

$D_2 = (-15.36, 3.84)^{\mathrm{T}}$

$\lambda_2^* = 0.3125$, $X_3 = (1,1)^{\mathrm{T}}$

F-R 共轭梯度法 —— 寻优轨迹对比

$$\min f(X) = x_1^2 + 4x_2^2 - 2x_1 - 8x_2 + 5$$



要点：与一维最优解的梯度的正交性

共轭方向和一维最优解的梯度的正交性

条件： $f(X)=0.5X^TAX+B^TX+C$， $A>0$

$\vec{p}_0,\vec{p}_1,\cdots,\vec{p}_{n-1}$ 为 A 的共轭方向

$X_0\in R^n$ 是任意的出发点

由下述一维搜索依次确定 $X_1,X_2,\cdots,X_n$

$$f(X_{k+1})=f(X_k+t_k\vec{p}_k)=\min_{t\in R} f(X_k+t\vec{p}_k)$$

$$k=0,1,\cdots,n-1$$

结论： $\vec{p}_j^T\nabla f(X_k)=0,\ \forall 0\le j<k$

理由： $\min_{t>0} f(X_k + t\vec{p}_k) \quad \Rightarrow \quad t_k = -\dfrac{\vec{p}_k^T \nabla f(X_k)}{\vec{p}_k^T A \vec{p}_k}, \forall 0 \le k \le n-1$

$$X_k = X_{k-1} + t_{k-1}\vec{p}_{k-1} \quad \Rightarrow \quad X_k = X_0 + \sum_{i=0}^{k-1} t_i \vec{p}_i$$

$$\Rightarrow \quad \nabla f(X_k) = \nabla f(X_0) + \sum_{i=0}^{k-1} t_i A \vec{p}_i$$

$$\begin{aligned} \vec{p}_j^T \nabla f(X_k) &= \vec{p}_j^T \nabla f(X_0) + t_j \vec{p}_j^T A \vec{p}_j \\ &= \vec{p}_j^T \nabla f(X_0) - \vec{p}_j^T \nabla f(X_j), \quad \forall 0 \le j < k \end{aligned}$$

$$\Rightarrow$$

$$\vec{p}_j^T \nabla f(X_j) = \vec{p}_j^T \nabla f(X_0), \quad \forall j$$

$$\Rightarrow \quad \vec{p}_j^T \nabla f(X_k) = 0, \quad \forall 0 \le j < k$$

推论：沿共轭方向寻优的每个 $X_k, k=1,2,\cdots,n$ 都满足

$$f\left(X_k\right)=\min\left\{ f\left(X\right) \middle| \text{ s.t. } X=X_0+\sum_{j=0}^{k-1}\beta_j\vec{p}_j \right\}$$

理由：$X=X_0+\sum_{j=0}^{k-1}\beta_j\vec{p}_j,\ X_k=X_0+\sum_{j=0}^{k-1}\beta_{kj}\vec{p}_j$

$$\Rightarrow\quad \nabla f\left(X_k\right)^T\left(X-X_k\right)=\sum_{j=0}^{k-1}\nabla f\left(X_k\right)^T\vec{p}_j\left(\beta_j-\beta_{kj}\right)$$

利用 $\nabla f\left(X_k\right)^T\vec{p}_j=0,\ j=0,1,\cdots,k-1$

可得 $\nabla f\left(X_k\right)^T\left(X-X_k\right)=0$

再利用凸函数一阶充要条件可得结论

要点：共轭方向二次函数有限终止性

共轭方向二次函数有限终止性

条件： $f(X)=0.5X^TAX+B^TX+c$，A对称正定

$\vec{p}_0,\vec{p}_1,\cdots,\vec{p}_{n-1}$ 为 A 的共轭方向

$X_0\in R^n$ 是任意的出发点

由下述直线搜索依次确定 $X_1,X_2,\cdots,X_n$

$$f(X_{k+1})=f(X_k+t_k\vec{p}_k)=\min_{t>0} f(X_k+t\vec{p}_k)$$

$$k=0,1,\cdots,n-1$$

结论： $f(X_n)=\min_{X\in R^n} f(X)$

理由：1）由推论可知

$$f\left(X_n\right)=\min\left\{f\left(X\right)\middle|\ \text{s.t. } X=X_0+\sum_{j=0}^{n-1}\vec{p}_j\beta_j\right\}$$

2）由原理之一可知 $R^n=\left\{X\middle|\ X=X_0+\sum_{j=0}^{n-1}\vec{p}_j\beta_j\right\}$

理由：从共轭方向的几个特点出发：

1、共轭方向线性无关 $\Rightarrow$ $\vec{p}_0, \vec{p}_1, \cdots, \vec{p}_{n-1}$ 是 R^n 的一组基

2、$\nabla f(X) = AX + B$ 是 R^n 的列向量，则对于任意的 $\hat{X}$

若 $0 \neq \nabla f(\hat{X}) = \alpha_0 \vec{p}_0 + \alpha_1 \vec{p}_1 + \cdots + \alpha_{n-1} \vec{p}_{n-1}$

3、从 X_i 出发沿 $\vec{p}_i$ 直线搜索，则有 $\nabla^T f(X_k + t_k \vec{p}_k) \vec{p}_k = 0$

即 $\nabla^T f(X_{k+1}) \vec{p}_k = 0$

4、$\nabla f(X_{k+1}) = AX_{k+1} + B = A(X_k + t_k \vec{p}_k) + B = \nabla f(X_k) + t_k A\vec{p}_k$，

则有 $\vec{p}_{k-1}^T \nabla f(X_{k+1}) = \vec{p}_{k-1}^T \nabla f(X_k) + \vec{p}_{k-1}^T t_k A\vec{p}_k$，进而由3

和共轭方向性质有 $\vec{p}_{k-1}^T \nabla f(X_{k+1}) = 0$，依此类推得到

$\vec{p}_i^T \nabla f(X_{k+1}) = 0,\ i = 0, 1, \cdots, k$

如果 $\nabla f(X_n) \neq 0$，则引发如下矛盾

$$\nabla^T f(X_n) \nabla f(X_n) = \nabla^T f(X_n) \left(\alpha_0 \vec{p}_0 + \alpha_1 \vec{p}_1 + \cdots + \alpha_{n-1} \vec{p}_{n-1} \right) = 0$$

要点：共轭方向的生成

共轭方向的生成

用Gram-Schmidt 正交化方法顺序生成 A 共轭方向

利用 A 共轭性确定下面方程组中所有待定系数

$$\vec{p}_0 = -\nabla f(X_0)$$

$$\vec{p}_1 = -\nabla f(X_1) + \alpha_{10}\vec{p}_0$$

$$\vdots$$

$$\vec{p}_{n-1} = -\nabla f(X_{n-1}) + \alpha_{n-10}\vec{p}_0 + \alpha_{n-11}\vec{p}_1 + \cdots + \alpha_{n-1n-2}\vec{p}_{n-2}$$

例如： $\vec{p}_0^T A\vec{p}_1 = 0 \quad \Rightarrow \quad 0 = -\vec{p}_0^T A\nabla f(X_1) + \alpha_{10}\vec{p}_0^T A\vec{p}_0$

$$\Rightarrow \quad \alpha_{10} = \frac{\vec{p}_0^T A\nabla f(X_1)}{\vec{p}_0^T A\vec{p}_0}$$

解前面方程组最终可得

$$\vec{p}_0 = -\nabla f(X_0)$$

$$\vec{p}_1 = -\nabla f(X_1) + \alpha_{10}\vec{p}_0$$

$$\vdots$$

$$\vec{p}_{n-1} = -\nabla f(X_{n-1}) + \alpha_{n-10}\vec{p}_0 + \alpha_{n-11}\vec{p}_1 + \cdots + \alpha_{n-1n-2}\vec{p}_{n-2}$$

其中

$$\alpha_{kj} = \frac{\nabla^T f(X_k) A \vec{p}_j^T}{\vec{p}_j^T A \vec{p}_j}, \quad 1 \le k \le n-1,\ 0 \le j \le k-1$$

为了应用于一般性的非线性函数，需要消除 A

消除 A 的基本途经：

$$X_k = X_{k-1} + t_{k-1}\vec{p}_{k-1} \quad \Rightarrow \quad \nabla f\left(X_k\right) = \nabla f\left(X_{k-1}\right) + t_{k-1}A\vec{p}_{k-1}$$

由推论，$\vec{p}_{k-1}^T \nabla f\left(X_k\right) = \vec{p}_{k-1}^T \nabla f\left(X_{k-1}\right) + t_{k-1}\vec{p}_{k-1}^T A\vec{p}_{k-1} = 0$

$$t_j = -\frac{\vec{p}_j^T \nabla f(X_j)}{\vec{p}_j^T A\vec{p}_j} \qquad \alpha_{kj} = \frac{\nabla^T f(X_k) A\vec{p}_j^T}{\vec{p}_j^T A\vec{p}_j} = \frac{t_j \nabla^T f(X_k) A\vec{p}_j^T}{-\vec{p}_j^T \nabla f(X_j)}$$

$$\Rightarrow \quad = \frac{\nabla^T f(X_k)\left(\nabla f\left(X_{j+1}\right) - \nabla f\left(X_j\right)\right)}{-\vec{p}_j^T \nabla f(X_j)}$$

由于 $j \le k-1$ ，上式已经可以应用于一般性函数，再利用梯度和共轭方向的关系，可进一步简化系数表达式

$$\vec{p}_j = -\nabla f(X_j) + \alpha_{j0}\vec{p}_0 + \alpha_{j1}\vec{p}_1 + \cdots + \alpha_{j\,j-1}\vec{p}_{j-1}$$

$$\Rightarrow \quad \nabla f(X_j) = -\vec{p}_j + \alpha_{j0}\vec{p}_0 + \alpha_{j1}\vec{p}_1 + \cdots + \alpha_{j\,j-1}\vec{p}_{j-1}$$

$$\nabla^T f(X_k)\vec{p}_j = 0, \forall 0 \le j < k$$

$$\Rightarrow \quad \nabla^T f(X_k)\nabla f(X_j)$$

$$= \nabla^T f(X_k)\left(-\vec{p}_j + \alpha_{j0}\vec{p}_0 + \alpha_{j1}\vec{p}_1 + \cdots + \alpha_{j\,j-1}\vec{p}_{j-1}\right)$$

$$= 0, \quad \forall 0 \le j < k$$

$$\nabla^T f(X_k)\nabla f(X_k)$$

$$= \nabla^T f(X_k)\left(-\vec{p}_k + \alpha_{k0}\vec{p}_0 + \alpha_{k1}\vec{p}_1 + \cdots + \alpha_{k\,k-1}\vec{p}_{k-1}\right)$$

$$= -\nabla^T f(X_k)\vec{p}_k$$

$$\nabla^T f(X_k)\nabla f(X_j) = 0,\ \ \forall 0 \le j < k$$

$$\nabla^T f(X_k)\nabla f(X_k) = -\nabla^T f(X_k)\vec{p}_k$$

$$\Rightarrow$$

$$\alpha_{kj} = \frac{\nabla^T f(X_k)\left(\nabla f\left(X_{j+1}\right) - \nabla f\left(X_j\right)\right)}{-\vec{p}_j^T \nabla f(X_j)}$$

$$= \frac{\nabla^T f(X_k)\nabla f\left(X_{j+1}\right) - \nabla^T f(X_k)\nabla f\left(X_j\right)}{-\vec{p}_j^T \nabla f(X_j)}$$

$$\alpha_{kj} = \begin{cases} 0 & \text{if} \quad j < k-1 \\ \dfrac{\nabla^T f(X_k)\nabla f\left(X_k\right)}{\nabla^T f(X_{k-1})\nabla f(X_{k-1})} & \text{if} \quad j = k-1 \end{cases}$$

$$
\begin{aligned}
\alpha_{kk-1} &= \frac{\nabla^T f(X_k)\left(\nabla f\left(X_k\right)-\nabla f\left(X_{k-1}\right)\right)}{-\vec{p}_{k-1}^T \nabla f(X_{k-1})} \\
&= \frac{\nabla^T f(X_k)\nabla f\left(X_k\right)}{\nabla^T f(X_{k-1})\nabla f(X_{k-1})} \\
&= \frac{\left\|\nabla f(X_k)\right\|^2}{\left\|\nabla f(X_{k-1})\right\|^2} \\
&= \frac{\nabla^T f(X_k)\left(\nabla f\left(X_k\right)-\nabla f\left(X_{k-1}\right)\right)}{\left\|\nabla f(X_{k-1})\right\|^2} \\
&= \frac{\nabla^T f(X_k)\left(\nabla f\left(X_k\right)-\nabla f\left(X_{k-1}\right)\right)}{\vec{p}_{k-1}^T\left(\nabla f\left(X_k\right)-\nabla f\left(X_{k-1}\right)\right)}
\end{aligned}
$$

$$
\begin{aligned}
&-\vec{p}_{k-1}^T \nabla f(X_{k-1}) \\
&= 0 - \vec{p}_{k-1}^T \nabla f(X_{k-1}) \\
&= \vec{p}_{k-1}^T \nabla f(X_k) - \vec{p}_{k-1}^T \nabla f(X_{k-1})
\end{aligned}
$$

要点：三种共轭梯度法

共轭梯度法（Fletcher-Reeves）

1）任取 $X_0 \in R^n$，令 $k=0$

2）如果 $\|\nabla f(X_k)\| \le \varepsilon$，停止计算

3）如果 k/n 等于 0 或整数，令 $D_k = -\nabla f(X_k)$

否则令 $D_k = -\nabla f(X_k) + \alpha_{k-1} D_{k-1}$，其中

$$\alpha_{k-1} = \frac{\|\nabla f(X_k)\|^2}{\|\nabla f(X_{k-1})\|^2}$$

4）进行精确搜索获得 $X_{k+1} = X_k + t_k D_k$

5）用 $k+1$ 替换 k，回到2）继续迭代

共轭梯度法（Polak-Ribiere / Polyak）

1）任取 $X_0 \in R^n$，令 $k=0$

2）如果 $\|\nabla f(X_k)\| \le \varepsilon$，停止计算

3）如果 k/n 等于 0 或整数，令 $D_k = -\nabla f(X_k)$

否则令 $D_k = -\nabla f(X_k) + \alpha_{k-1} D_{k-1}$，其中

$$\alpha_{k-1} = \frac{\nabla^T f(X_k)\left(\nabla f(X_k) - \nabla f(X_{k-1})\right)}{\|\nabla f(X_{k-1})\|^2}$$

4）进行精确搜索获得 $X_{k+1} = X_k + t_k D_k$

5）用 $k+1$ 替换 k，回到2）继续迭代

共轭梯度法（Beale-Sorenson / Hestenes-Stiefel）

1）任取 $X_0 \in R^n$，令 $k=0$

2）如果 $\|\nabla f(X_k)\| \le \varepsilon$ ，停止计算

3）如果 k/n 等于 0 或整数，令 $D_k = -\nabla f(X_k)$

否则令 $D_k = -\nabla f(X_k) + \alpha_{k-1} D_{k-1}$ ，其中

$$\alpha_{k-1} = \frac{\nabla^T f(X_k)\left(\nabla f(X_k) - \nabla f(X_{k-1})\right)}{D_{k-1}^T\left(\nabla f(X_k) - \nabla f(X_{k-1})\right)}$$

4）进行精确搜索获得 $X_{k+1} = X_k + t_k D_k$

5）用 $k+1$替换 k ，回到2）继续迭代

关于共轭梯度法的结论

1）共轭梯度法是下降算法

2）对于正定二次目标函数

$$f(X)=\frac{1}{2}X^{T}AX+B^{T}X+C$$

如果从相同的初始点出发，三种共轭梯度法前进的轨迹完全相同，即每一步一维搜索得到的点均相同，并且，经过 n 次精确的一维搜索后一定找到最优解，即 $X_n=-A^{-1}B$

意义：对一般非线性函数在最优解附近快速收敛

要点：几种算法的性能比较

三种基于梯度的搜索方向的比较

	计算量	效率 解附近	 远离解	鲁棒性
负梯度	A	C	A	A
共轭梯度	B	B	B	B
牛顿方向	C	A	C	C