# 最优化理论
# 第五章：无约束最优化方法

人工智能学院
智能感知与图像理解教育部重点实验室

# 无约束最优化方法

1 共轭方向法

2 共轭梯度法

3 FR共轭梯度法

4 算法总结

# 无约束最优化方法

1 共轭方向法

2 共轭梯度法

3 FR共轭梯度法

4 算法总结

# 共轭方向法

## 一、共轭方向法引入动机

构成各种不同最优化方法，往往取决于如何从基本迭代公式 $X_{k+1}=X_k+t_kP_k$ 中确定搜索方向$P_k$。

**最速下降法：计算步骤简单**，但 $P_k=-\nabla f(X_k)$ 导致搜索路线出现锯齿状，**收敛速度慢**。

**Newton法和修正Newton法**：$P_k=-[\nabla^2 f(X_k)]^{-1}\nabla f(X_k)$使得**收敛速度快**，但**计算计算量大且要求Hesse矩阵正定，导致算法对初始点选择要求严格。**

因此需要寻找一种好的算法，它的收敛速度介于最速下降法和牛顿法之间，对于正定二次函数只需迭代有限次就可达到极小点，收敛速度快同时计算简单的算法——**共轭方向法**。

# 共轭方向法

## 一、共轭方向法引入动机

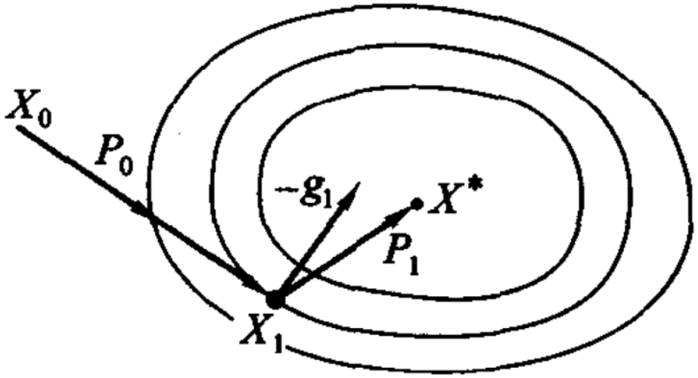
共轭方向法涉及共轭方向的概念和性质。共轭方向的概念是在研究正定二次函数

$$f(\vec{x}) = \frac{1}{2}\vec{x}^T A\vec{x} + b^T\vec{x} + c \qquad (5.13)$$

其中 $A \in R^{n\times n}$ 是正定矩阵，$b \in R^n$，$c \in R^1$。

任选取初始点 $X_0$，沿某个下降方向 $P_0$ 作精确一维搜索得

$$X_1 = X_0 + t_0 P_0$$




由精确一维搜索的性质，可知

$$(\nabla f(X_1))^T P_0 = 0$$

下一次迭代如果按最速下降法选择负梯度为搜索方向，将会发生锯齿现象。

## 一、共轭方向法引入动机

为了克服锯齿现象，假设能够选定下一次迭代的搜索方向直指极小点 $X^*$，那么对于二元二次函数只需依次沿搜索方向 $P_0, P_1$ 进行两次精确一维搜索就可以求到极小点 $X^*$，即

$$X^* = X_1 + \mathbf{t}_1 P_1$$

$X^*$是$\boldsymbol{f}(\boldsymbol{x})$的极小点，故$X^*$是$\boldsymbol{f}(\boldsymbol{x})$的驻点。对(5.13)函数求导数得

$$\nabla f(X) = AX + \boldsymbol{b}$$

# 共轭方向法

## 一、共轭方向法引入动机

$$\nabla f(X^*) = AX^* + \boldsymbol{b}$$
$$= A(X_1 + \mathbf{t}_1 P_1) + \boldsymbol{b}$$
$$= (AX_1 + \boldsymbol{b}) + \mathbf{t}_1 AP_1$$
$$= \nabla f(X_1) + + \mathbf{t}_1 AP_1 = 0$$

$$P_0^T(\nabla f(X_1) + \mathbf{t}_1 AP_1) = 0$$
$$P_0^T \nabla f(X_1) + \mathbf{t}_1 P_0^T AP_1 = 0$$
$$\text{而} P_0^T \nabla f(X_1) = 0 \text{且} \mathbf{t}_1 \neq 0,$$
$$\text{即} P_0^T AP_1 = 0$$

2次迭代要得到二元二次函数的极小点，$P_1$ 必须满足的条件是：$\boxed{(P_0)^T AP_1 = 0}$

搜索方向 $P_0, P_1$ 是A共轭的。

$P_0$是某个下降方向

## 二、共轭方向法基本原理

**定义5.1**

设 $A\in R^{n\times n}$ 是对称正定矩阵，$P_1,P_2\in R^n$，如果 $P_1^T AP_2=0$，则称向量 $P_1$和$P_2$ 是**A共轭**(或A正交)的。

如果对有限个向量 $P_1,P_2,\ldots,P_m$，有

$$(P_i)^T AP_j=0\ (i\neq j,\ j=1,2,\ldots,m)$$

则称这个向量组是*A共轭方向组*(或*A*正交方向组），也称它们是一组*A*的共轭方向。

若 $A=I$，$P_i^T P_j=0$（$i\neq j$），则 $P_i$与$P_j$ 是正交的。共轭是正交的推广

共轭向量组是正交向量组的推广。

## 二、共轭方向法基本原理

**定理5.1：在$n$维空间中与$n$个线性无关的向量都正交的一定是零向量。**

证明：假设 $P_1,P_2,\ldots,P_n \in R^n$ 线性无关，向量 $d \in R^n$与 $P_1,P_2,\ldots,P_n$ 正交，证 $d=0$.

因为 $P_1,P_2,\ldots,P_n$ 线性无关，所以 $P_1,P_2,\ldots,P_n$ 可作为 $R^n$ 中的一组基，故 $d$可由 $P_1,P_2,\ldots,P_n$ 线性表示，即存在一组实数 $\alpha_1,\alpha_2,\ldots,\alpha_n \in R$, s.t.

$$d=\alpha_1 P_1+\alpha_2 P_2+\ldots+\alpha_n P_n.$$

$$\begin{aligned} d^T d &= d^T\left(\alpha_1 P_1+\alpha_2 P_2+\ldots+\alpha_n P_n\right) \\ &=\alpha_1 d^T P_1+\alpha_2 d^T P_2+\ldots+\alpha_n d^T P_n \\ &=\alpha_1 \cdot 0+\alpha_2 \cdot 0+\ldots+\alpha_n \cdot 0=0 \end{aligned}$$

故 $d=0$.

## 二、共轭方向法基本原理

**定理5.2：若$R^n$中的非零向量 $P_1, P_2, \ldots, P_m\ (m \le n)$是A共轭向量组，则这m个向量是线性无关的。**

证明：假设 $\alpha_1 P_1 + \alpha_2 P_2 + \ldots + \alpha_m P_m = 0$，要证 $P_1, P_2, \ldots, P_m$ 线性无关，只需证明 $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$.

因为 $P_1, P_2, \ldots, P_m$ 是A共轭向量组，两边同乘 $(P_k)^T A, \forall k = 1, \ldots, m$,

$$
\begin{aligned}
0 &= (P_k)^T A(\alpha_1 P_1 + \alpha_2 P_2 + \ldots + \alpha_m P_m) \\
&= 0 + 0 + \ldots + \alpha_k (P_k)^T A P_k + \ldots + 0 \\
&= \alpha_k (P_k)^T A P_k,
\end{aligned}
$$

$(P_i)^T A P_j = 0\ (i \ne j, j = 1, 2, \ldots, m)$

因为A正定，$p^k \ne 0$，所以 $(P_k)^T A P_k > 0$，故 $\alpha_k = 0, \forall k = 1, \ldots, m$，得证。

## 二、共轭方向法基本原理

**定理5.3：**

设**n**元函数 $f(x)=\dfrac{1}{2}X^TAX+b^TX+c,\ A=A^T$ 正定，又设**n**维非零向量组 $P_1,P_2,\ldots,P_n$ 是**A**共轭向量组，从任意点 $X_1$ 出发，依次以该向量组为搜索方向进行精确一维搜索，则

(1) $\nabla f(X_{k+1})$ 与 $P_1,P_2,\ldots,P_k\ (k=1,2,\ldots,n)$ 正交；

(2) 最多**n**次迭代必达到二次函数***f*(*x*)**的极小点。

## 二、共轭方向法基本原理

**(1)** $\nabla f(X_{k+1})$ 与 $P_1, P_2, \ldots, P_k (k=1,2,\ldots,n)$ 正交；

证明：$f(X)=\frac{1}{2}X^T AX+b^T X+c$,即$\nabla f(X)=AX+b$

$$\nabla f(X_{k+1})=AX_{k+1}+b=A(X_k+t_k P_k)+b=(AX_k+b)+t_k AP_k$$
$$=\nabla f(X_k)+t_k AP_k$$

$t_k$ 是按照精确一维搜索得到的，所以有 $\nabla f(X_{k+1})^T P_k=0$

下证 $\nabla f(X_{k+1})^T P_i=0, \quad i=1,2,\ldots,k-1.$

$$\nabla f(X_{k+1})=\nabla f(X_k)+t_k AP_k$$
$$=\nabla f(X_{k-1})+t_{k-1}AP_{k-1}+t_k AP_k=\ldots$$
$$=\nabla f(X_{i+1})+t_{i+1}AP_{i+1}+\ldots+t_{k-1}AP_{k-1}+t_k AP_k, i=1,2,\ldots,k-1.$$

## 二、共轭方向法基本原理

**(1)** $\nabla f(X_{k+1})$ 与 $P_1,P_2,\ldots,P_k(k=1,2,\ldots,n)$ 正交；

$$\nabla f(X_{k+1})=\nabla f(X_{i+1})+t_{i+1}AP_{i+1}+\ldots+t_{k-1}AP_{k-1}+t_kAP_k, i=1,2,\ldots,k-1.$$

$$\left(\nabla f(X_{k+1})\right)^T P_i=\left(\nabla f(X_{i+1})\right)^T P_i+t_{i+1}(AP_{i+1})^T P_i$$

$$+\ldots+t_{k-1}(AP_{k-1})^T P_i+t_k(AP_k)^T P_i$$

$$=\nabla f(X_{i+1})^T P_i+t_{i+1}(P_{i+1})^T AP_i$$

$$\boxed{\nabla f(X_{k+1})^T P_k=0}$$

$$+\ldots+t_{k-1}(P_{k-1})^T AP_i+t_k(P_k)^T AP_i$$

$$=0$$

故 $\nabla f(X_{k+1})$ 与 $P_1,P_2,\ldots,P_k(k=1,2,\ldots,n)$ 正交。

$P_1,P_2,\ldots,P_k$ 是$A$共轭向量组

## 二、共轭方向法基本原理

**(2) 最多*n*次迭代必达到二次函数*f*(*x*)的极小点。**

证明：假设 $\nabla f(X_1),\nabla f(X_2),\ldots,\nabla f(X_n)$ 都不是0，下证

$$\nabla f(X_{n+1})=0.$$

利用(1)的结果，$\nabla f(X_{n+1})$ 与A共轭向量组 $P_1,P_2,\ldots,P_k$ 都正交，由定理5.1可知，

$$\nabla f(x^{n+1})=0.$$

定理5.1：在*n*维空间中与*n*个线性无关的向量都正交的一定是零向量。

## 二、共轭方向法基本原理

在下降迭代法中，若取下降方向是共轭方向，所得到的方法我们称为共轭方向法。

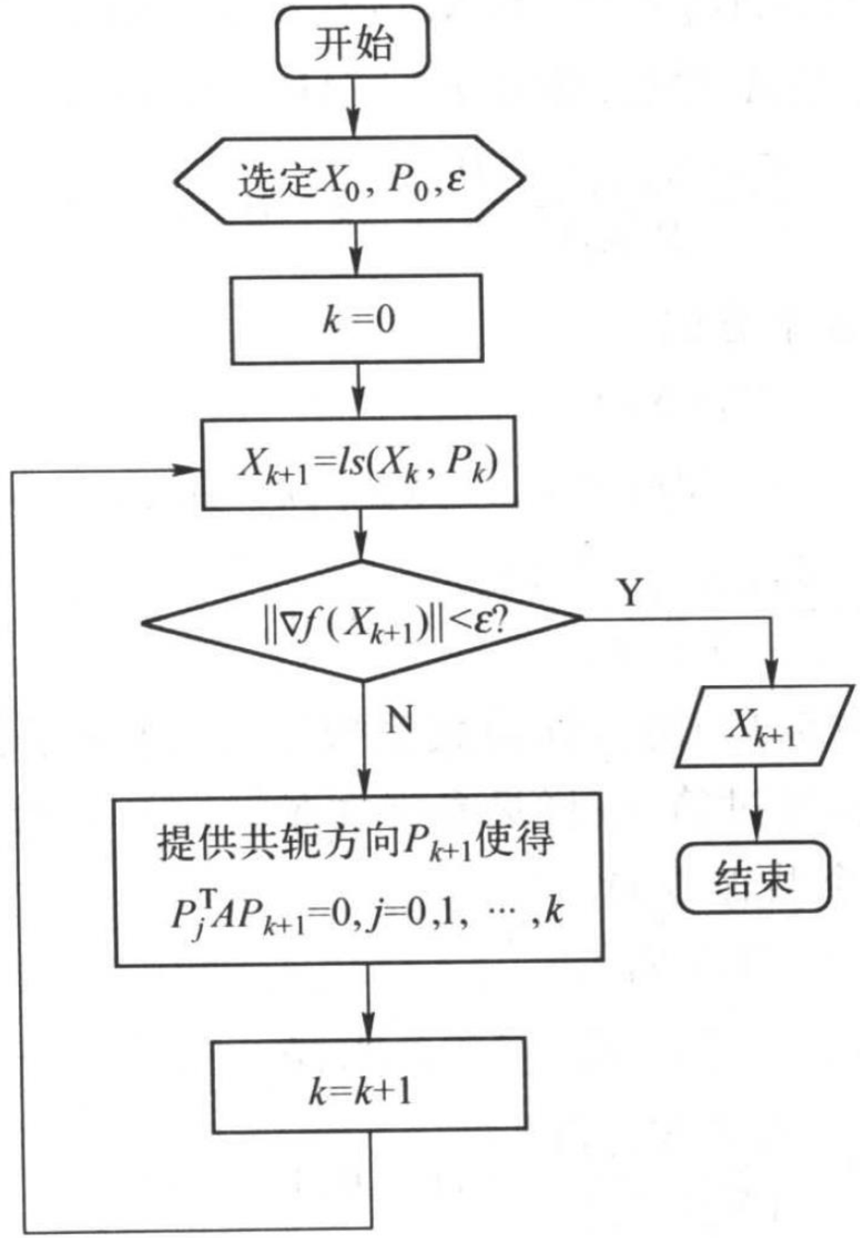
由定理5.3可知，若能找到一组$A$共轭向量，以这组共轭向量为下降方向进行搜索，则在有限步可得到正定二次函数的极小点，即共轭方向法具有二次收敛性。

定义5.2：一个算法若能在有限步内求得正定二次函数的极小点，则称该算法具有二次收敛性(又称二次终止性)。

# 共轭方向法

## 三、共轭方向法流程图

怎么选取共轭方向？



# 无约束最优化方法

1 共轭方向法

2 共轭梯度法

3 FR共轭梯度法

4 算法总结

## 一、共轭方向与共轭梯度法关系

共轭方向的选取有很大任意性，而一组不同的共轭方向就对应着不同的共轭方向法。作为一种迭代算法，我们自然希望共轭方向能在迭代过程中逐次生成。

下面先以正定二次函数为例，介绍一种生成共轭方向的方法，再将这种方法推广到非二次函数上。这种方法中每一个共轭向量都是依赖于迭代点处的负梯度而构造出来，所以称为共轭梯度法，它是共轭方向法中的一种。

## 二、共轭梯度法

令 $f(\mathrm{X})=\frac{1}{2}X^{T}AX+b^{T}X+c$，$A$对称正定

(1) 从任取初始点$X_1$出发，沿负梯度方向进行精确一维搜索:

$$P_1=-\nabla f(\mathrm{X}_1) \qquad X_2=X_1+t_1P_1$$

(2) 若 $\nabla f(X_2)=0$则停止，否则在$-\nabla f(\mathrm{X}_2)$ 和 $P_1$ 形成的正交锥中找一个向量$P_2$，即令$P_2=-\nabla f(X_2)+\alpha_1 P_1$ 使得 $P_1$与$P_2$共轭，即

$$(P_2)^T AP_1=0$$

$$\alpha_1=\frac{\nabla f(\mathrm{X}_2)^T AP_1}{(P_1)^T AP_1}$$

## 二、共轭梯度法

(3) 在$X_2$处沿$P_2$方向进行精确一维搜索，$X_3 = X_2 + t_2 P_2$

(4) 以此类推，$X_4, X_5, \ldots$

(5) 若$\nabla f(X_{k+1}) = 0$ ， 停止; 否则在$-\nabla f(X_{k+1})$ 和$P_k$ 形成的正交锥中找一个向量 $P_{k+1}$，即令$P_{k+1} = -\nabla f(X_{k+1}) + \alpha_k P_k$ 使得$P_{k+1}$与$P_k$共轭，即

$$(P_{k+1})^T A P_k = 0.$$

$$\alpha_k = \frac{\nabla f(X_{k+1})^T A P_k}{(P_k)^T A P_k}$$

## 二、共轭梯度法

这样便构造了一组向量$P_1, P_2, \ldots, P_n$，相邻两个向量是共轭的。

$$P_1 = -\nabla f(X_1)$$

$$P_{k+1} = -\nabla f(X_{k+1}) + \alpha_k P_k,\ k = 1, 2, \ldots, n-1,$$

$$\alpha_k = \frac{\nabla f(X_{k+1})^T A P_k}{(P_k)^T A P_k}$$

实际上，这组向量 $P_1, P_2, \ldots, P_n$ 是一个A共轭向量组。

每一个搜索方向都依赖迭代点处的负梯度构造出来，对应的算法称为共轭梯度法。

## 二、共轭梯度法

**定理5.4：** 设向量组$P_1, P_2, \ldots, P_n$ 是由上述方法产生的向量组, 向量 $g_1, g_2, \ldots, g_n$ 是由各点的梯度生成的向量组$\left(g_k = \nabla f(X_k)\right)$ ，则$g_1, g_2, \ldots, g_n$ 是正交向量组，且 $P_1, P_2, \ldots, P_n$ 是**A共轭向量组**。

简单证明：$g_{k+1}$与$p_k, p_{k-1}, \cdots, p_1$正交

$$p_k = g_k + \alpha_k p_{k-1}$$

$$g_{k+1}^T p_k = g_{k+1}^T g_k + \alpha_k g_{k+1}^T p_{k-1} = 0$$

$\rightarrow g_{k+1}, g_k$正交

注：为保证方向的共轭性，初始方向取负梯度方向。

## 二、共轭梯度法

注：初始方向为下降方向，但不是负梯度方向

**例** 用共轭梯度法求下列问题的极值

$$f(\boldsymbol{x}) = x_1^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2 \text{ ，已知初始 } X_1 = \begin{pmatrix}1\\1\\1\end{pmatrix},\ P_1 = \begin{pmatrix}-1\\-2\\0\end{pmatrix}.$$

解：第一次迭代：沿 $P_1$ 方向进行精确一维搜索，得 $t_1 = 2/3$

$$X_2 = X_1 + t_1 P_1 = \left(\frac{1}{3}, -\frac{1}{3}, 1\right)^T, g_2 = \nabla f(X_2) = \left(\frac{2}{3}, -\frac{1}{3}, 1\right)^T,$$

$$\alpha_1 = \frac{g_2^T A P_1}{(P_1)^T A P_1} = -\left(\frac{2}{3}\right)\Big/ 6 = -\frac{1}{9},$$

$$P_2 = -g_2 + \alpha_1 P_1 = \left(-\frac{5}{9}, \frac{5}{9}, -1\right)^T.$$

# 二、共轭梯度法

第二次迭代：沿$P_2$方向进行精确一维搜索，得 $t_2 = 21/26$

$$X_3 = X_2 + t_2 P_2 = \left(-\frac{9}{78}, \frac{9}{78}, \frac{5}{26}\right)^T, g_3 = \left(-\frac{18}{78}, \frac{9}{78}, \frac{5}{26}\right)^T,$$

$$\alpha_2 = \frac{g_3^T A P_2}{(P_2)^T A P_2} = \frac{45}{676},$$

$$P_3 = -g_3 + \alpha_2 P_2 = \frac{1}{676}(131, -53, -175)^T.$$

容易验证，$P_1$与$P_2$关于A共轭，$P_2$与$P_3$也关于A共轭，但是$P_1$与$P_3$不共轭。因此，$P_1, P_2$与$P_3$不是A共轭向量组。

初始方向选择负梯度方向特别重要。

# 二、共轭梯度法

针对 $f(\mathrm{X})=\frac{1}{2}X^{T}AX+b^{T}X+c$，$A$对称正定，最多$n$次迭代达到极小点找到了一组共轭方向：

$$
\begin{cases}
P_1=-\nabla f(\mathrm{X}_1) \\
P_{k+1}=-\nabla f(\mathrm{X}_{k+1})+\alpha_k P_k,\ k=1,2,\ldots,n-1, \\
\alpha_k=\dfrac{\nabla f(\mathrm{X}_{k+1})^T AP_k}{(P_k)^T AP_k}
\end{cases}
$$

**存在问题：计算量、存储量都很大，对任意目标函数，无法得到A**

在正定二次函数的前提下，将 $\alpha_k$变形？

# 共轭梯度法

## 三、$\alpha_k$ 不同的五种共轭梯度法

(1) Daniel共轭梯度法 $$\alpha_k = \frac{\nabla f(x^{k+1})^T Ap^k}{(p^k)^T Ap^k} = \frac{(g_{k+1})^T Ap^k}{(p^k)^T Ap^k} \quad (Daniel, 1967)$$

(2) SW共轭梯度法 $$\alpha_k = \frac{(g_{k+1})^T (g_{k+1} - g_k)}{(p^k)^T (g_{k+1} - g_k)} \quad (Sorenson - Wolfe, 1972)$$

(3) DM共轭梯度法 $$\alpha_k = -\frac{(g_{k+1})^T g_{k+1}}{(p^k)^T g_k} \quad (Dixon - Myers, 1972)$$

**(4) FR共轭梯度法**
（简洁，应用广泛） $$\alpha_k = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \quad (Flecher - Reeves, 1964)$$

(5)PPR共轭梯度法
（更好用） $$\alpha_k = \frac{(g_{k+1})^T (g_{k+1} - g_k)}{(g_k)^T g_k} \quad (Polyak - Polak - Ribiere, 1969)$$

对于正定二次函数，上面得到的5个计算公式是等价的。

# 无约束最优化方法

# 一、FR共轭梯度法

设$f(x)=\frac{1}{2}x^T Ax+bx+c, x_0$为初始点，$p_0=-\nabla f(x_0)$为初始方向

求解$\mathrm{t}_0=\min\limits_t f(x_0+tp_0)$

$x_1=x_0+t_0 p_0,$(有$p_0=\frac{1}{t_0}(x_1-x_0)$)

令$p_1=-\nabla f(x_1)+\alpha_0 p_0$

目的：求解$\alpha_0$使得$p_1^T Ap_0=0$

# 一、FR共轭梯度法

注意$\nabla f(x)=Ax+b$

$\nabla f(x_0)=Ax_0+b$

$\nabla f(x_1)=Ax_1+b$

$\nabla f(x_1)-\nabla f(x_0)=A(x_1-x_0)$

要求$p_1^T A p_0=0$

$\rightarrow p_1^T A(\frac{1}{t_0}(x_1-x_0))=0$

$\rightarrow (-\nabla f(x_1)-\alpha_0\nabla f(x_0))^T(\nabla f(x_1)-\nabla f(x_0))=0$

$\rightarrow \alpha_0=\frac{\|\nabla f(x_1)\|^2}{\|\nabla f(x_0)\|^2}$

有$p_0=\frac{1}{t_0}(x_1-x_0)$

$$p_1=-\nabla f(x_1)+\alpha_0 p_0$$
$$=-\nabla f(x_1)-\alpha_0\nabla f(x_0)$$

## 二、FR共轭梯度法计算步骤

Step 1. 选定初始点$x^1$。

Step 2. 如果$\|g_1\| \le \varepsilon$，算法停止，$x^* = x^1$；否则转Step 3;

Step 3. 取 $p^1 = -g_1, k=1$;

Step 4. 精确一维搜索找最佳步长$\lambda_k$，令 $x^{k+1} = x^k + \lambda_k p^k$;

Step 5. 如果$\|g_{k+1}\| \le \varepsilon$，算法停止，$x^* = x^{k+1}$，否转Step 6;

Step 6. 如果$k=n$，令 $x^1 = x^{k+1}, p^1 = -g_{k+1}$，算法停止，$k=1$，转Step 4;

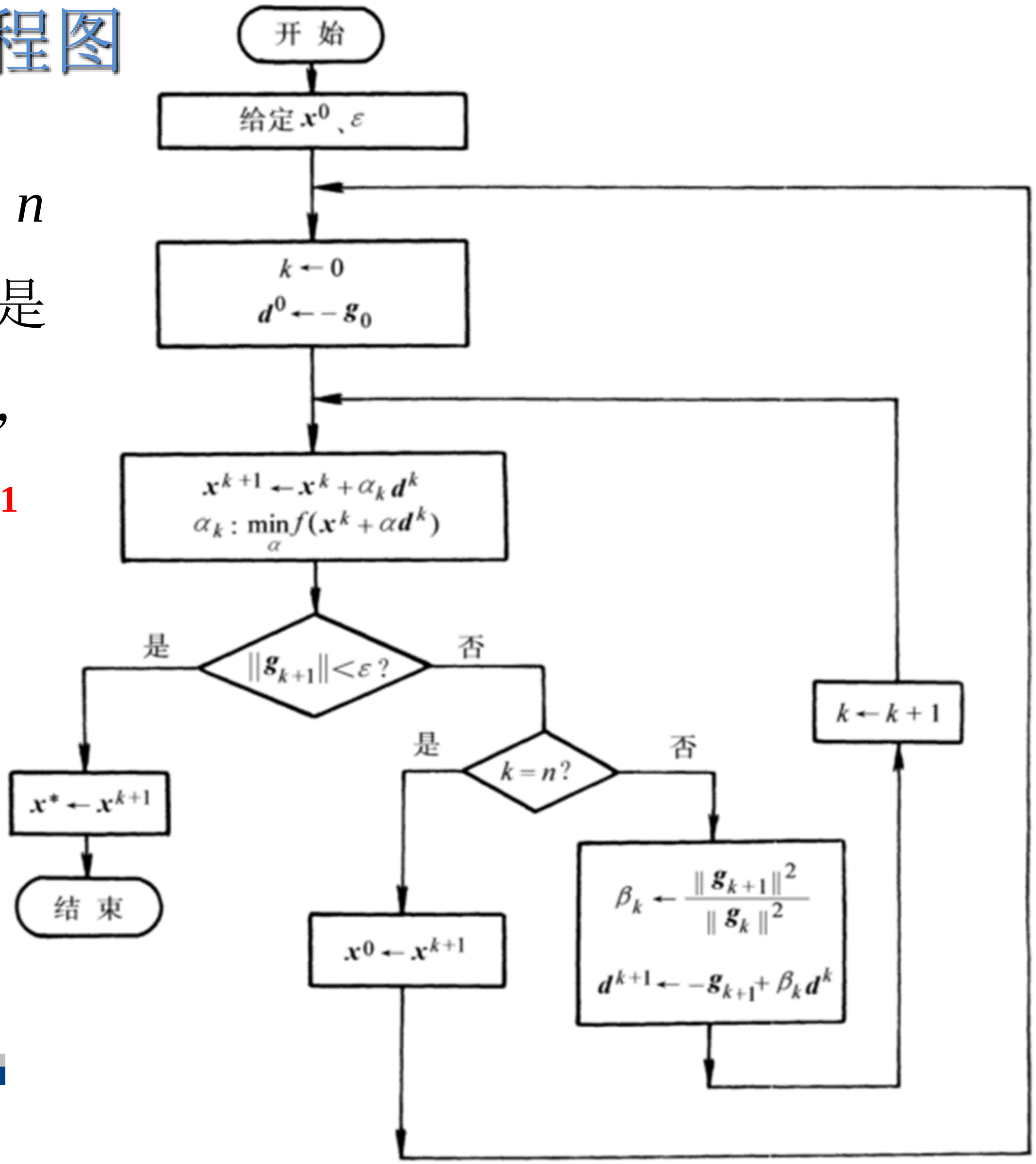
否则转步骤7;

Step 7. 计算 $\alpha_k = \dfrac{\|g_{k+1}\|^2}{\|g_k\|^2}$，$p^{k+1} = -g_{k+1} + \alpha_k p^k$，$k = k+1$，转Step 4。

# FR共轭梯度法

## 三、FR共轭梯度法流程图

$R^n$中共轭方向最多有$n$个，$n$步后构造的搜索方向不再是共轭的，会降低收敛速度，因此**Step 6 重新开始：$x^{n+1}$作为新的$x^1$**



# 四、FR共轭梯度法举例

例 用$FR$共轭梯度法$f(\boldsymbol{x})=x_1^2+2x_2^2-4x_1-2x_1x_2$的极小值，已知初始点为$(1,1)^{\mathrm{T}}$，迭代精度$\varepsilon=0.001$。

解: (1)第一次迭代：沿负梯度方向搜寻。

计算初始点处的梯度

$$g_0=\nabla f(x^0)=\begin{pmatrix}2x_1-2x_2-4\\4x_2-2x_1\end{pmatrix}\Bigg|_{x=x^0}=\begin{pmatrix}-4\\2\end{pmatrix}$$

$$p^0=-g_0=\begin{pmatrix}4\\-2\end{pmatrix}$$

$$x^0+\lambda p^0=\begin{pmatrix}1\\1\end{pmatrix}+\lambda\begin{pmatrix}4\\-2\end{pmatrix}=\begin{pmatrix}1+4\lambda\\1-2\lambda\end{pmatrix}$$

# 四、FR共轭梯度法举例

例 用FR共轭梯度法 $f(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 4x_1 - 2x_1x_2$ 的极小值，已知初始点为 $(1,1)^{\mathrm{T}}$，迭代精度 $\varepsilon = 0.001$。

精确一维搜索求最佳步长，$\phi(\lambda)=f(x^0+\lambda p^0)=f(1+4\lambda,1-2\lambda)=40\lambda^2-20\lambda-3$,

令 $0=\phi'(\lambda)=80\lambda-20$,

得 $\lambda_0 = 0.25$

$$x^1 = x^0 + \lambda_0 p^0 = \begin{pmatrix} 1+4\lambda \\ 1-2\lambda \end{pmatrix} = \begin{pmatrix} 2 \\ 0.5 \end{pmatrix}$$

$$g_1 = \nabla f(x^1) = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

不满足精度，继续迭代：

# 四、FR共轭梯度法举例

例 用*FR*共轭梯度法$f(\mathbf{x})=x_1^2+2x_2^2-4x_1-2x_1x_2$的极小值，已知初始点为$(1,1)^{\mathrm{T}}$，迭代精度$\varepsilon=0.001$。

**（2）第二次迭代：**

$$g_0=\begin{pmatrix}-4\\2\end{pmatrix},\quad g_1=\begin{pmatrix}-1\\-2\end{pmatrix},\quad p^0=\begin{pmatrix}4\\-2\end{pmatrix},\quad x^1=\begin{pmatrix}2\\0.5\end{pmatrix},$$

$$\alpha_0=\frac{\|g_1\|^2}{\|g_0\|^2}=\frac{5}{20}=0.25\qquad p^1=-g_1+\alpha_0p^0=\begin{pmatrix}2\\1.5\end{pmatrix},$$

$$x^1+\lambda p^1=\begin{pmatrix}2\\0.5\end{pmatrix}+\lambda\begin{pmatrix}2\\1.5\end{pmatrix}=\begin{pmatrix}2+2\lambda\\0.5+1.5\lambda\end{pmatrix},$$

精确一维搜索求最佳步长，$\phi(\lambda)=f(x^1+\lambda p^1)=f(2+2\lambda,0.5+1.5\lambda)$

$$=(2+2\lambda)^2+2(0.5+1.5\lambda)^2-2(2+2\lambda)(0.5+1.5\lambda)-4(2+2\lambda)$$

## 四、FR共轭梯度法举例

例 用$FR$共轭梯度法$f(\boldsymbol{x})=x_1^2+2x_2^2-4x_1-2x_1x_2$的极小值，已知初始点为$(1,1)^{\mathrm{T}}$，迭代精度$\varepsilon=0.001$。

$$\phi(\lambda)=(2+2\lambda)^2+2(0.5+1.5\lambda)^2-2(2+2\lambda)(0.5+1.5\lambda)-4(2+2\lambda)$$

令 $0=\phi'(\lambda)$,

得 $\lambda_1=1$,

$$x^2=x^1+\lambda_1 p^1=\begin{pmatrix}2+2\lambda\\0.5+1.5\lambda\end{pmatrix}=\begin{pmatrix}4\\2\end{pmatrix},\qquad g_2=\nabla f(x^2)=\begin{pmatrix}0\\0\end{pmatrix},$$

因 $\|g_2\|=0<\varepsilon$

得到最优解: $x^*=x^2=\begin{pmatrix}4\\2\end{pmatrix}$.

# FR共轭梯度法

Rosenbrock function
举例

OUTPUT OF STEEPEST DESCENT FOR ROSENBROCK'S FUNCTION

| ITERATION | FUNCTIONAL VALUE | INDEPENDENT VARIABLE VALUES | |
|---|---|---|---|
| 1 | 0.800000D+01 | -0.100000D+01 | -0.100000D+01 |
| 2 | 0.812503D+00 | 0.499005D+00 | -0.500332D+00 |
| 3 | 0.451564D+00 | 0.350837D+00 | -0.505562D-01 |
| 4 | 0.262715D+00 | 0.644972D+00 | 0.463002D-01 |
| 5 | 0.184997D+00 | 0.581630D+00 | 0.238476D+00 |
| 6 | 0.133115D+00 | 0.727438D+00 | 0.286628D+00 |
| 7 | 0.101903D+00 | 0.688701D+00 | 0.403629D+00 |
| 8 | 0.788786D-01 | 0.781869D+00 | 0.434408D+00 |
| 9 | 0.629818D-01 | 0.754935D+00 | 0.515845D+00 |
| 10 | 0.506272D-01 | 0.821078D+00 | 0.537735D+00 |
| 11 | 0.414636D-01 | 0.800961D+00 | 0.598561D+00 |
| 12 | 0.341084D-01 | 0.850827D+00 | 0.615022D+00 |
| 13 | 0.284133D-01 | 0.835152D+00 | 0.662288D+00 |
| 14 | 0.237435D-01 | 0.874150D+00 | 0.675226D+00 |
| 15 | 0.200203D-01 | 0.861559D+00 | 0.713055D+00 |
| 16 | 0.169190D-01 | 0.892833D+00 | 0.723433D+00 |
| 17 | 0.143938D-01 | 0.882566D+00 | 0.754365D+00 |
| 18 | 0.122669D-01 | 0.908130D+00 | 0.762839D+00 |
| 19 | 0.105079D-01 | 0.899626D+00 | 0.788518D+00 |
| 20 | 0.901350D-02 | 0.920832D+00 | 0.795530D+00 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 180 | 0.268216D-10 | 0.999996D+00 | 0.999988D+00 |
| 181 | 0.237863D-10 | 0.999995D+00 | 0.999989D+00 |
| 182 | 0.210944D-10 | 0.999996D+00 | 0.999990D+00 |

THE MINIMAL SOLUTION FOUND

F(X) = 0.210944D-10

X( 1) = 0.999996D+00
X( 2) = 0.999990D+00

# FR共轭梯度法

OUTPUT OF FLETCHER-REEVES FOR ROSENBROCK'S FUNCTION
(RESETTING $d_k = -\nabla f(x_k)$ AFTER EACH N ITERATIONS)

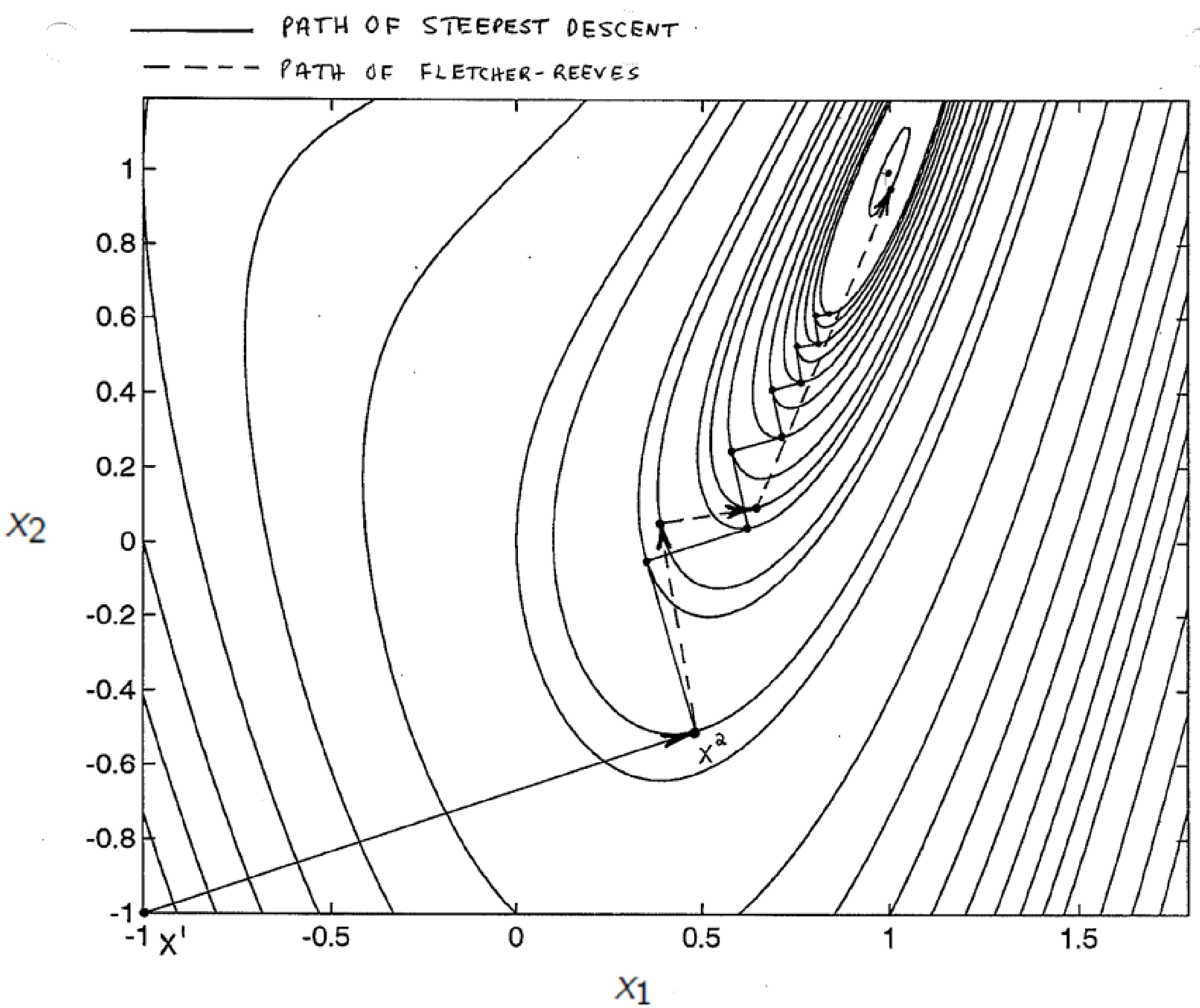
| ITERATION | FUNCTIONAL VALUE | INDEPENDENT VARIABLE VALUES | |
|---|---|---|---|
| 1 | .800000D+01 | -.100000D+01 | -.100000D+01 |
| 2 | .812503D+00 | .500966D+00 | -.499678D+00 |
| 3 | .383289D+00 | .390233D+00 | .451686D-01 |
| 4 | .227756D+00 | .647871D+00 | .976166D-01 |
| 5 | .262354D-02 | .967714D+00 | .976233D+00 |
| 6 | .193254D-03 | .987238D+00 | .969126D+00 |
| 7 | .121901D-05 | .999491D+00 | .999963D+00 |
| 8 | .847627D-08 | .999915D+00 | .999794D+00 |
| 9 | .111635D-08 | .999977D+00 | .999977D+00 |
| 10 | .148806D-09 | .999989D+00 | .999973D+00 |
| 11 | .711720D-12 | .100000D+01 | .100000D+01 |
| 12 | .468360D-14 | .100000D+01 | .100000D+01 |
| 13 | .430046D-15 | .100000D+01 | .100000D+01 |

THE MINIMAL SOLUTION FOUND

F(X) = .430046D-15

X( 1) = .100000D+01
X( 2) = .100000D+01

# FR共轭梯度法


PATH OF STEEPEST DESCENT
PATH OF FLETCHER-REEVES
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
$x_2$
-1
-0.5
0
0.5
1
1.5
$x_1$
$x^1$
$x^2$

# 无约束最优化方法

1 共轭方向法

2 共轭梯度法

3 FR共轭梯度法

4 算法总结

# 一、共轭梯度法

共轭梯度法对正定二次函数，具有二次收敛性；

对非二次函数，由于目标函数的Hesse矩阵不再是常数矩阵，因而产生的方向不再是共轭方向；

共轭梯度法在一定条件下也是收敛的，且收敛速度通常优于最速下降法，具有较高的求解效率。

# 一、共轭梯度法

设 $f(x)$ 存在连续一阶偏导数，且函数为凸函数，且水平集 $L=\{x \mid f(x)\le f(x^0)\}$ 有界，则由共轭梯度法得到的点列 $\{x^k\}$ 有如下性质：

(1) $\{f(x^k)\}$ 为严格单调下降序列，且 $\lim\limits_{k\to\infty} f(x^k)$ 存在；

(2) $\{x^k\}$ 的任意聚点 $x^*$ 都是 $f(x)$ 的最优解。

共轭梯度法在无约束优化方法中占有重要的地位，是目前最常用的方法之一。

# FR共轭梯度法

## 二、共轭梯度法的特点

(1) 对凸函数全局收敛（下降算法）；

(2) 计算公式简单，不用求Hesse矩阵或者逆矩阵，计算量小，存储量小，每步迭代只需存储若干向量，适用于大规模问题；

(3) 具有二次收敛性；

(对于正定二次函数，至多$n$次迭代可达最优解）

(4) Crowder和Wolfe证明，共轭梯度法的收敛速率不坏于最速下降法。如果初始方向不用负梯度方向，则其收敛速率是线性收敛的；

(5) 共轭梯度法是目前求解无约束优化问题最常用的方法之一。

注：不同的$\alpha_k$公式，对于正定二次函数是等价的，

对非正定二次函数，有不同的效果，经验上$PPR$ 效果较好。

# FR共轭梯度法

**第四次作业:**

习题五（ P119 ）

第1,2,3,4,6题

西安電子科技大學
XIDIAN UNIVERSITY

智能感知与图像理解

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education

# *THE END*

Thanks for your participation!