4-3 共轭梯度法

共轭梯度法

共轭梯度法最早是由Hestenes和Stiefel(1952)提出来的，用于解正定系数矩阵的线性方程组. 在此基础上, Fletcher和Reeves(1964)首先提出了解非线性最优化问题的共轭梯度法。

共轭梯度法是介于最速下降法与牛顿法之间的一种优化方法，其主要利用函数的一阶导数信息建立起来的方法, 对于正定二次函数具有二次终止性。

共轭梯度法在每步迭代中所需的计算量和存储量比Newton法少，不仅是解大型线性方程组最有用的方法之一, 也是解大型非线性最优化问题最有效的算法之一。

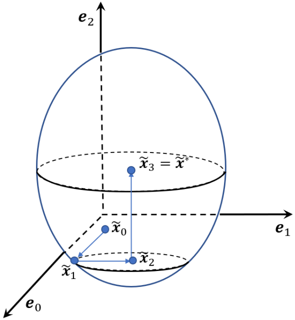
特殊正定二次函数

针对正定二次函数：

$$\tilde{f}(\tilde{\boldsymbol{x}}) = \frac{1}{2}\tilde{\boldsymbol{x}}^T G\tilde{\boldsymbol{x}} + \boldsymbol{b}^T\tilde{\boldsymbol{x}} \tag{4.3.1}$$

其中G为$n \times n$的对角阵, 对角元素为正。

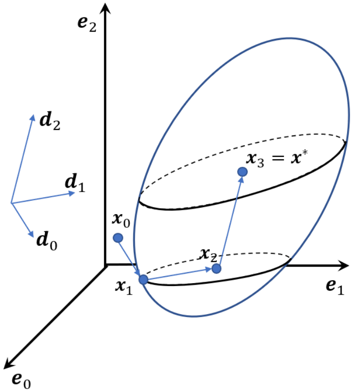
考虑$n = 3$情形，从某初始点$\tilde{\boldsymbol{x}}_0$出发，分别沿着平行与坐标轴的方向$\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2$做精确线性搜索，得到迭代点列$\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \tilde{\boldsymbol{x}}_3$。经过3步迭代可精确找到正定二次函数的极小点。



一般正定二次函数

对于一般的正定二次函数$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x}$，即$G$为对称正定矩阵，从任意点出发，沿$\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2$做精确线性搜索，显然不能期望得到极小点。

欲保持二次终止性，沿着$\boldsymbol{d}_0, \boldsymbol{d}_1, \boldsymbol{d}_2$方向分别进行精确线性搜索。这些方向如何构造？



一般正定二次函数

考虑变换:

$$\boldsymbol{x} = D\tilde{\boldsymbol{x}}$$

这里

$$D = [\boldsymbol{d}_0, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{n-1}]_{n\times n}$$

关于$\tilde{x}$的函数形式为:

$$\tilde{f}(\tilde{\boldsymbol{x}}) = f(D\tilde{\boldsymbol{x}}) = \frac{1}{2}(D\tilde{\boldsymbol{x}})^T G(D\tilde{\boldsymbol{x}}) + b^T(D\tilde{\boldsymbol{x}}) = \frac{1}{2}\tilde{\boldsymbol{x}}^T(D^T GD)\tilde{\boldsymbol{x}} + \boldsymbol{b}^T(D\tilde{\boldsymbol{x}})$$

若对$i \neq j$, 有:

$$\boldsymbol{d}_i^T G\boldsymbol{d}_j = 0$$

成立, 则以上函数的海森矩阵$D^T GD$为对角阵。

共轭方向

定义4.3.1: 设G是$n\times n$对称正定矩阵, $\boldsymbol{d}', \boldsymbol{d}''$ 是n维非零向量。如果

$$\boldsymbol{d}'^T G \boldsymbol{d}'' = 0, \tag{4.3.2}$$

则称向量$\boldsymbol{d}'$和$\boldsymbol{d}''$是关于矩阵G**共轭的**(G−共轭或G−正交)。

设$\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{n-1}$是$\mathbb{R}^n$中一组非零向量,如果

$$\boldsymbol{d}_i^T G \boldsymbol{d}_j = 0, \quad (i, j = 0, \ldots, n-1, i \neq j) \tag{4.3.3}$$

则称向量组$\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{n-1}$是关于矩阵$G$**共轭的**。

可以证明: 若向量组$\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{n-1}$是$G$−共轭的,则它们是线性无关的。特别地取$G$为单位阵, 则共轭性即为通常的正交性。

变换矩阵

变换矩阵D可以取为正定矩阵平方根逆矩阵。G的平方根矩阵定义与性质如下。

定义4.3.2 设G为$n\times n$对称正定矩阵. 若存在对称正定矩阵T，使得$T^2=G$，则称T为G的**平方根矩阵**，记为：$T=\sqrt{G}$或$T=G^{\frac{1}{2}}$。

定理4.3.1 平方根矩阵的存在唯一性.

若G是$n\times n$对称正定矩阵，则G的平方根矩阵存在且唯一。

取$D^{-1}=\sqrt{G}$, 则有：$\boldsymbol{e}_i^T\boldsymbol{e}_j=(\sqrt{G}\boldsymbol{d}_i)^T(\sqrt{G}\boldsymbol{d}_j)=\boldsymbol{d}_i^TG\boldsymbol{d}_j=0$。对于$\tilde{f}(\tilde{\boldsymbol{x}})$沿着$\boldsymbol{e}_0,\ldots,\boldsymbol{e}_{n-1}$方向依次进行迭代，通过变换$\boldsymbol{x}=D\tilde{\boldsymbol{x}}$, 就成为$f(\boldsymbol{x})$沿着$\boldsymbol{d}_0,\ldots,\boldsymbol{d}_{n-1}$方向依次进行迭代。

算法4.3.1 一般共轭方向法

步1 给出初始点$\boldsymbol{x}_0, \epsilon > 0, k := 0$. 计算$\boldsymbol{g}_0 = \boldsymbol{g}(x_0)$和初始下降方向$\boldsymbol{d}_0$, 使$\boldsymbol{d}_0^T \boldsymbol{g}_0 < 0$.

步2 如果$\|\boldsymbol{g}_k\| \le \epsilon$, 停止迭代。

步3 计算α_k 和$\boldsymbol{x}_{k+1}$, 使得

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) = \min_{\alpha \ge 0} f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k),$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k.$$

步4 采用某种共轭方向法计算$\boldsymbol{d}_{k+1}$ 使得

$$\boldsymbol{d}_{k+1}^T G \boldsymbol{d}_j = 0, j = 0, 1, \cdots, k.$$

步5 令$k := k + 1$, 转步2.

共轭方向法

极小化正定二次函数：

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}, \tag{4.3.4}$$

由一阶最优性必要条件，以上问题等价于解线性方程组：

$$G\boldsymbol{x} = \boldsymbol{b}, \tag{4.3.5}$$

记：$r(\boldsymbol{x}) \triangleq G\boldsymbol{x} - \boldsymbol{b} = g(\boldsymbol{x})$(4.3.5)的残量等于目标函数梯度。

共轭方向法的迭代格式：$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$. 由**精确线性搜索**确定步长因子为：

$$\alpha_k = -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T G \boldsymbol{d}_k}, \tag{4.3.6}$$

其中：$\boldsymbol{g}_k = G\boldsymbol{x}_k - \boldsymbol{b} \triangleq \boldsymbol{r}_k$.

共轭方向

定理4.3.1 共轭方向基本定理.

设$\boldsymbol{x}_0 \in \mathbb{R}^n$ 是任意初始点。对于极小化二次函数(4.3.4), 共轭方向法至多经n步精确线性搜索终止;且每一迭代点$\boldsymbol{x}_{k+1}$都是$f(\boldsymbol{x})$在$\boldsymbol{x}_0$和方向$\boldsymbol{d}_0, \cdots, \boldsymbol{d}_k$ 所张成的线性流形

$$\left\{ \boldsymbol{x} \;\middle|\; \boldsymbol{x} = \boldsymbol{x}_0 + \sum_{j=0}^{k} \beta_j \boldsymbol{d}_j,\ \forall \beta_j \right\}$$

中的极小点。

该定理表明：在精确线性搜索条件下,共轭方向法具有二次终止性(即对于正定二次函数, 方法是有限步终止的)。

证明： 因G正定,且向量组

$$\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots$$

线性无关, 故只需要证明对所有的$k \leq n-1$,有

$$\boldsymbol{g}_{k+1}^T \boldsymbol{d}_j = 0, \quad j = 0, \cdots, k, \tag{4.3.7}$$

便可得出定理结论。

由于

$$\boldsymbol{g}_{i+1} - \boldsymbol{g}_i = G(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) = \alpha_i G \boldsymbol{d}_i \tag{4.3.8}$$

结合在精确线性搜索得：

$$\boldsymbol{g}_{i+1}^T \boldsymbol{d}_i = 0, \tag{4.3.9}$$

定理4.3.1证明续

证明续: 故有当$j < k$时,

$$\boldsymbol{g}_{k+1}^T \boldsymbol{d}_j = \boldsymbol{g}_{j+1}^T \boldsymbol{d}_j + \sum_{i=j+1}^{k} (\boldsymbol{g}_{i+1} - \boldsymbol{g}_i)^T \boldsymbol{d}_j$$
$$= \boldsymbol{g}_{j+1}^T \boldsymbol{d}_j + \sum_{i=j+1}^{k} \alpha_i \boldsymbol{d}_i^T G \boldsymbol{d}_j = 0. \qquad (4.3.10)$$

在上式中两项为零分别由精确线性搜索和共轭性得到。

当$j = k$时,直接由精确线性搜索可知

$$\boldsymbol{g}_{k+1}^T \boldsymbol{d}_k = 0. \qquad (4.3.11)$$

从而(4.3.7)成立。

定理4.3.1证明续

证明续: 对$\boldsymbol{x}_k$与任意给定线性流形中的$\boldsymbol{x}$，有：

$$\boldsymbol{x}_k = \boldsymbol{x}_0 + \sum_{j=1}^{k-1} \alpha_j \boldsymbol{d}_j, \quad \boldsymbol{x} = \boldsymbol{x}_0 + \sum_{j=1}^{k-1} \beta_j \boldsymbol{d}_j,$$

由G的正定性与(4.3.7)式，有：

$$\begin{aligned}
f(\boldsymbol{x}) &= f(\boldsymbol{x}_k) + (\boldsymbol{x} - \boldsymbol{x}_k)^T \boldsymbol{g}_k + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^T G(\boldsymbol{x} - \boldsymbol{x}_k) \\
&\geq f(\boldsymbol{x}_k) + (\boldsymbol{x} - \boldsymbol{x}_k)^T \boldsymbol{g}_k \\
&= f(\boldsymbol{x}_k) + \sum_{j=0}^{k-1} (\beta_j - \alpha_j) \boldsymbol{d}_j^T \boldsymbol{g}_k \\
&= f(\boldsymbol{x}_k)
\end{aligned}$$

■

共轭梯度法

共轭梯度法(conjugate gradient method)是一种典型的共轭方向法，搜索方向构造要求如下：

- 所有搜索方向是互相共轭的。
- 搜索方向$\boldsymbol{d}_k$仅仅是负梯度方向$-\boldsymbol{g}_k$与上一次迭代的搜索方向$\boldsymbol{d}_{k-1}$的组合。

记

$$\boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_{k-1}\boldsymbol{d}_{k-1}, \tag{4.3.12}$$

上式两边左乘$\boldsymbol{d}_{k-1}^T G$, 利用方向共轭$\boldsymbol{d}_{k-1}^T G \boldsymbol{d}_k = 0$得：

$$\beta_{k-1} = \frac{\boldsymbol{g}_k^T G \boldsymbol{d}_{k-1}}{\boldsymbol{d}_{k-1}^T G \boldsymbol{d}_{k-1}}. \quad \text{(Hestenes-Stiefel公式)} \tag{4.3.13}$$

利用(4.3.8)和(4.3.9),上式也可以写成

$$\beta_{k-1}=\frac{\boldsymbol{g}_k^T(\boldsymbol{g}_k-\boldsymbol{g}_{k-1})}{\boldsymbol{d}_{k-1}^T(\boldsymbol{g}_k-\boldsymbol{g}_{k-1})},\text{(Crowder-Wolfe公式)} \tag{4.3.14}$$

$$=\frac{\boldsymbol{g}_k^T\boldsymbol{g}_k}{\boldsymbol{g}_{k-1}^T\boldsymbol{g}_{k-1}}.\text{(Fletcher-Reeves公式)} \tag{4.3.15}$$

另外三个常用的等价共轭梯度公式

$$\beta_{k-1}=\frac{\boldsymbol{g}_k^T(\boldsymbol{g}_k-\boldsymbol{g}_{k-1})}{\boldsymbol{g}_{k-1}^T\boldsymbol{g}_{k-1}},\qquad\text{(Polak-Ribiere-Polyak公式)} \tag{4.3.16}$$

$$\beta_{k-1}=-\frac{\boldsymbol{g}_k^T\boldsymbol{g}_k}{\boldsymbol{d}_{k-1}^T\boldsymbol{g}_{k-1}},\qquad\text{(Dixon公式)}$$

$$\beta_{k-1}=\frac{\boldsymbol{g}_k^T\boldsymbol{g}_k}{\boldsymbol{d}_{k-1}^T(\boldsymbol{g}_k-\boldsymbol{g}_{k-1})}.\qquad\text{(Dai-Yuan公式)}$$

记：

$$\boldsymbol{g}_k = G\boldsymbol{x}_k - \boldsymbol{b} \triangleq \boldsymbol{r}_k,$$

由上式以及：

$$\boldsymbol{g}_k^T \boldsymbol{d}_k = \boldsymbol{g}_k^T(-\boldsymbol{g}_k + \beta_{k-1}\boldsymbol{d}_{k-1}) = -\boldsymbol{g}_k^T \boldsymbol{g}_k = -\boldsymbol{r}_k^T \boldsymbol{r}_k$$

得：

$$\boldsymbol{r}_{k+1} - \boldsymbol{r}_k = G\boldsymbol{x}_{k+1} - G\boldsymbol{x}_k = \alpha_k G\boldsymbol{d}_k,$$

和

$$\alpha_k = -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T G \boldsymbol{d}_k} = \frac{\boldsymbol{r}_k^T \boldsymbol{r}_k}{\boldsymbol{d}_k^T G \boldsymbol{d}_k},$$

共轭梯度法基本步骤

算法4.3.2 共轭梯度法

步1 初始步:给出$\boldsymbol{x}_0, \epsilon > 0$,计算$\boldsymbol{r}_0 = G\boldsymbol{x}_0 - \boldsymbol{b}$, 令$\boldsymbol{d}_0 = -\boldsymbol{r}_0, k := 0$.

步2 如果$\|\boldsymbol{r}_k\| \le \epsilon$, 停止。

步3 计算:

$$
\begin{array}{rcll}
\alpha_k &=& \dfrac{\boldsymbol{r}_k^T\boldsymbol{r}_k}{\boldsymbol{d}_k^T G\boldsymbol{d}_k}, & \text{步长因子} \\
\boldsymbol{x}_{k+1} &=& \boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k, & \text{迭代点} \\
\boldsymbol{r}_{k+1} &=& \boldsymbol{r}_k + \alpha_k G\boldsymbol{d}_k & \text{残差} \\
\beta_k &=& \dfrac{\boldsymbol{r}_{k+1}^T\boldsymbol{r}_{k+1}}{\boldsymbol{r}_k^T\boldsymbol{r}_k}, & \text{组合系数} \\
\boldsymbol{d}_{k+1} &=& -\boldsymbol{r}_{k+1} + \beta_k\boldsymbol{d}_k. & \text{共轭方向}
\end{array}
$$

步4 令$k := k + 1$, 转步2.

注意: 充分利用$G\boldsymbol{d}_k$, 可计算因子α_k中分母内积向量, 以及迭代方式产生$\boldsymbol{r}_{k+1}$。另外, $\boldsymbol{r}_k^T\boldsymbol{r}_k$也可重复利用。

例4.3.1 用FR共轭梯度法解极小化问题

$$\min f(\boldsymbol{x}) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1.$$

设初始点为：$(0,0)^T$.

解： 目标函数写成矩阵向量乘积形式：

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$$

其中：

$$G = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad \boldsymbol{r}(\boldsymbol{x}) = G\boldsymbol{x} - \boldsymbol{b}.$$

设$\boldsymbol{x}_0 = (0,0)^T$,得

$$\boldsymbol{r}_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \quad \boldsymbol{d}_0 = -\boldsymbol{r}_0 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad \alpha_0 = \frac{\boldsymbol{r}_0^T\boldsymbol{r}_0}{\boldsymbol{d}_0^T G\boldsymbol{d}_0} = \frac{1}{3},$$

共轭梯度法算例

解续：

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0 \boldsymbol{d}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix}$$

$$\boldsymbol{r}_1 = G\boldsymbol{x}_1 - \boldsymbol{b} = \begin{pmatrix} 0 \\ -\frac{2}{3} \end{pmatrix}, \quad \beta_0 = \frac{\boldsymbol{r}_1^T \boldsymbol{r}_1}{\boldsymbol{r}_0^T \boldsymbol{r}_0} = \frac{1}{9}$$

$$\boldsymbol{d}_1 = -\boldsymbol{r}_1 + \beta_0 \boldsymbol{d}_0 = \begin{pmatrix} 0 \\ \frac{2}{3} \end{pmatrix} + \frac{1}{9}\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{9} \\ \frac{2}{3} \end{pmatrix}, \quad \alpha_1 = \frac{\boldsymbol{r}_1^T \boldsymbol{r}_1^T}{\boldsymbol{d}_1^T G \boldsymbol{d}_1} = \frac{3}{2}$$

$$\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1 \boldsymbol{d}_1 = \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} + \frac{3}{2}\begin{pmatrix} \frac{2}{9} \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\boldsymbol{r}_2 = G\boldsymbol{x}_2 - b = \begin{pmatrix} 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

所以$\boldsymbol{x}_2 = (1,1)^T$是所求的极小点。 ■

定理4.3.2 共轭梯度法基本性质.

对于正定二次函数, 对于任意初始点$\boldsymbol{x}_0 \in \mathbb{R}$, 取$\boldsymbol{d}_0 = -\boldsymbol{g}_0$, 采用精确线性搜索的共轭梯度法在$m \le n$步后终止(二次终止性),且对$1 \le i \le n$成立下列关系式:

$$\text{共轭性}: \ \boldsymbol{d}_i^T G \boldsymbol{d}_j = 0, \ j = 0, 1, \cdots, i-1, \tag{4.3.17a}$$

$$\text{正交性}: \ \boldsymbol{g}_i^T \boldsymbol{g}_j = 0, \ j = 0, 1, \cdots, i-1, \tag{4.3.17b}$$

$$\text{下降性}: \ \boldsymbol{d}_i^T \boldsymbol{g}_i = -\boldsymbol{g}_i^T \boldsymbol{g}_i. \tag{4.3.17c}$$

以及:

$$\operatorname{span}\{\boldsymbol{g}_0, \ldots, \boldsymbol{g}_k\} = \operatorname{span}\{\boldsymbol{g}_0, G\boldsymbol{g}_0, \ldots, G^k \boldsymbol{g}_0\} \tag{4.3.17d}$$

$$\operatorname{span}\{\boldsymbol{d}_0, \ldots, \boldsymbol{d}_k\} = \operatorname{span}\{\boldsymbol{d}_0, G\boldsymbol{d}_0, \ldots, G^k \boldsymbol{d}_0\} \tag{4.3.17e}$$

定理4.3.2证明

证明：由共轭方向的推导可得(4.3.17a)式，由(4.3.7)与(4.3.12)可以推导出(4.3.17b)式。由共轭梯度法的定义有：

$$
\begin{aligned}
\boldsymbol{d}_k^T \boldsymbol{g}_k &= (-\boldsymbol{g}_k + \beta_{k-1}\boldsymbol{d}_{k-1})^T \boldsymbol{g}_k \\
&= -\boldsymbol{g}_k^T \boldsymbol{g}_k < 0,
\end{aligned}
$$

即(4.3.17c)成立。

由数学归纳法可以证明(4.3.17d)和(4.3.17e)两式。(略) ■

共轭梯度法收敛速度

可以证明: 共轭梯度产生的点列误差估计:

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_G \le \left(\frac{\sqrt{\kappa(G)}-1}{\sqrt{\kappa(G)}+1}\right)^k \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_G. \tag{4.3.18}$$

其中

$$\kappa(G) = \|G\|_2\|G^{-1}\|_2 = \frac{\lambda_1}{\lambda_n}.$$

共轭梯度法的收敛速度:

- 依赖于$\sqrt{\kappa(G)}$, 与最速下降法的收敛速度$\kappa(G)$相比，具有更快的收敛速度。
- 通过改善G的条件数，可加快收敛速度。

Fletcher和Reeves(1964)提出了针对非二次函数的共轭梯度法。

算法4.3.3 FR共轭梯度法

步1 给出$\boldsymbol{x}_0, \varepsilon > 0, \boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0)$, 令$\boldsymbol{d}_0 = -\boldsymbol{g}_0, k := 0$.

步2 若$\|\boldsymbol{g}_k\| \leq \varepsilon$, 停止。

步3 进行一维线性搜索求步长因子α_k,并令

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k.$$

步4 $\beta_k = \dfrac{\boldsymbol{g}_{k+1}^T \boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^T \boldsymbol{g}_k}$ $, \boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k.$

步5 令$k := k + 1$, 转步2.

Fletcher和Revees提出在共轭梯度法中使用非精确线性搜索。

特点： 程序简单，计算量小，无矩阵存储与计算，适合解大型非线性规划。

FR共轭梯度法的下降性

在非线性的共轭梯度法中，使用精确线性搜索与非精确线性搜索是有区别的。使用精确线性搜索可以保证迭代方向的下降性。若采用非精确线性搜索，FR方法可能产生上升的方向。只有当使用强Wolfe线搜索并保证$0 < \sigma < 1/2$时，得到的方向是下降方向。

定理4.3.3 FR方法的下降性.

对于FR共轭梯度法，若α_k满足强Wolfe准则，且$\sigma \in \left(0, \frac{1}{2}\right)$,则$\boldsymbol{d}_k$满足：

$$-\frac{1}{1-\sigma} \leq \frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2} \leq \frac{2\sigma - 1}{1-\sigma} \tag{4.3.19}$$

$\boldsymbol{d}_k$是下降方向。

定理4.3.3 FR方法的收敛性.

设$f(\boldsymbol{x})$有下界，$\boldsymbol{g}(\boldsymbol{x})$满足Lipschitz条件，对使用精确线性搜索准则的FR方法，则或存在N, 使得$\boldsymbol{g}_N = 0$, 或者

$$\liminf_{k\to\infty} \|\boldsymbol{g}_k\| = 0.$$