第三章 无约束优化方法

§3.4 共轭梯度法

天津大学数学学院

2020年3月

◆ 共轭梯度法最早是由Hestenes和Stiefel在1952年提出的，用于解正定系数矩阵的线性方程组；

◆ 1964年，Fletcher和Reeves在上述工作基础上提出了解非线性最优化问题的共轭梯度法.

教学内容

- 二次函数极小化的共轭方向法（ § 3.4.1）
- 二次函数极小化的共轭梯度法（ §3.4.2 ）
- 一般函数极小化的共轭梯度法（ § 3.4.3）

§3.4.1 二次函数极小化的共轭方向法

（一）共轭方向及其性质

定义3.4.1 (共轭向量组)

设$\boldsymbol{G}$为n阶对称正定矩阵，$\boldsymbol{d}^0,\boldsymbol{d}^1,\cdots,\boldsymbol{d}^{k-1}(k\le n)$为$n$维向量组，若

$$(\boldsymbol{d}^i)^{\mathrm{T}}\boldsymbol{G}\,\boldsymbol{d}^j=0,\ i,j\in\{0,1,\cdots,k-1\},\ \ i\ne j,$$

则称向量组$\boldsymbol{d}^0,\boldsymbol{d}^1,\cdots,\boldsymbol{d}^{k-1}$关于$G$共轭.

注 当$\boldsymbol{G}=\boldsymbol{I}$为单位矩阵，且$\boldsymbol{d}^0,\boldsymbol{d}^1,\cdots,\boldsymbol{d}^{k-1}(k\le n)$为$n$维非零向量组，则

$$(\boldsymbol{d}^i)^{\mathrm{T}}\boldsymbol{d}^j=0,$$

则 $\boldsymbol{d}^i\perp\boldsymbol{d}^j$，$i\ne j$，从而向量组$\boldsymbol{d}^0,\boldsymbol{d}^1,\cdots,\boldsymbol{d}^{k-1}$是正交组.

定理3.4.1　设$\boldsymbol{G}$为n阶对称正定矩阵，非零n维向量组$\boldsymbol{d}^0,\boldsymbol{d}^1,\cdots,\boldsymbol{d}^{k-1}$ $(k \le n)$关于$\boldsymbol{G}$共轭，则

(i)　向量组$\boldsymbol{d}^0,\boldsymbol{d}^1,\cdots,\boldsymbol{d}^{k-1}$线性无关；

(ii) 若$k=n$,则向量组$\boldsymbol{d}^0,\boldsymbol{d}^1,\cdots,\boldsymbol{d}^{n-1}$构成$\mathbb{R}^n$的一组基；

(iii) 若$k=n$,且向量$\boldsymbol{v}$与向量组$\boldsymbol{d}^0,\boldsymbol{d}^1,\cdots,\boldsymbol{d}^{n-1}$关于$\boldsymbol{G}$共轭，则$\boldsymbol{v}=\boldsymbol{0}$.

（二） 共轭方向法的理论基础

定义3.4.2 (k维超平面)

设n维向量组$d^0, d^1, \cdots, d^{k-1} (k \le n)$线性无关，给定$x^0 \in \mathbb{R}^n$,称集合

$$H_k := \{x^0 + \sum_{i=0}^{k-1} \lambda_i d^i \mid \lambda_i \in \mathbb{R}, i \in \{0,1,\cdots k-1\}\}$$

为由点x^0与$d^0, d^1, \cdots, d^{k-1}$生成的$k$维超平面.

引理3.4.1 设f为连续可微的严格凸函数，$d^0, d^1, \cdots, d^{k-1} (k \le n)$为一组线性无关的$n$维向量，$x^0 \in \mathbb{R}^n$,则

$$x^k = x^0 + \sum_{i=0}^{k-1} \hat{\lambda}_i d^i$$

是函数f在x^0与$d^0, d^1, \cdots, d^{k-1}$生成的$k$维超平面$H_k$上的唯一极小点

$\Leftrightarrow g_k{}^{\mathrm{T}} d^i = 0, \quad i \in \{0,1,\cdots k-1\}.$ （3.4.2）

证明　定义函数

$$h(\lambda_0,\lambda_1,\cdots,\lambda_{k-1}) := f\left(x^0+\sum_{i=0}^{k-1}\lambda_i d^i\right).$$

那么,

(i) $\nabla h(\hat{\lambda}_0,\hat{\lambda}_1,\cdots,\hat{\lambda}_{k-1})=(g_k^{\mathrm{T}}d^0,g_k^{\mathrm{T}}d^1,\cdots,g_k^{\mathrm{T}}d^{k-1})^{\mathrm{T}}$;

(ii) h 是 k 维空间 $\mathbb{R}^k$ 上的严格凸函数;

(iii) x^k 是函数 f 在 H_k 上的极小值点当且仅当 $(\hat{\lambda}_0,\hat{\lambda}_1,\cdots,\hat{\lambda}_{k-1})^{\mathrm{T}}$ 是函数 $h(\lambda_0,\lambda_1,\cdots,\lambda_{k-1})$ 在 $\mathbb{R}^k$ 上的极小值点.

一方面，若 x^k 是 f 在 H_k 上的极小值点，则由定理 3.1.1 中的 (i) 可知, $\nabla h(\hat{\lambda}_0,\hat{\lambda}_1,\ldots,\hat{\lambda}_{k-1})=0$. 这与上述结论 (i) 相结合可得 (3.4.2) 式成立.

另一方面，若 (3.4.2) 式成立，则结合上述结论 (i)，可得 $\nabla h(\hat{\lambda}_0,\hat{\lambda}_1,\cdots,\hat{\lambda}_{k-1})=0$. 又由上述结论 (ii) 可知, h 是 k 维空间 $\mathbb{R}^k$ 上的严格凸函数. 根据定理 3.1.3, 可得 $(\hat{\lambda}_0,\hat{\lambda}_1,\cdots,\hat{\lambda}_{k-1})^{\mathrm{T}}$ 是 h 在 $\mathbb{R}^k$ 上的极小值点. 这与上述结论 (iii) 相结合可得: x^k 是 f 在 H_k 上的唯一极小值点. 引理得证. $\square$

共轭方向的基本定理

定理3.4.2（子空间扩展定理）设G为n阶对称正定矩阵，非零n维向量组$d^0, d^1, \cdots, d^{k-1} (k \le n)$关于$G$共轭，且二次函数$f(x) = \frac{1}{2} x^{\mathrm{T}} G x + q^{\mathrm{T}} x$. 由任意初始点$x^0$开始，依次进行$k$次精确一维搜索，得

$$x^{i+1} = x^i + \lambda_i d^i,\ i \in \{0, 1, \cdots k-1\},$$

则x^k是二次函数f在H_k上的极小点.

注 对于正定二次函数$f(x) = \frac{1}{2} x^{\mathrm{T}} G x + q^{\mathrm{T}} x$的极小化问题,若迭代方向$d^0, d^1, \cdots, d^{k-1}$关于$G$共轭,步长由精确一维线搜索得到，则

$$g_k^{\mathrm{T}} d^i = 0,\ \ i \in \{0, 1, \cdots k-1\}.$$

证明 由引理 3.4.1 可知, 只需证明 (3.4.2) 式成立. 因为

$$
\begin{aligned}
g_k^{\mathrm{T}} d^i &= g_k^{\mathrm{T}} d^i - g_{k-1}^{\mathrm{T}} d^i + g_{k-1}^{\mathrm{T}} d^i - g_{k-2}^{\mathrm{T}} d^i + \cdots + g_{i+2}^{\mathrm{T}} d^i - g_{i+1}^{\mathrm{T}} d^i + g_{i+1}^{\mathrm{T}} d^i \\
&= g_{i+1}^{\mathrm{T}} d^i + \sum_{j=i+1}^{k-1} (g_{j+1} - g_j)^{\mathrm{T}} d^i, \quad \forall i \in \{0, 1, \cdots, k-1\},
\end{aligned}
$$

且 $g_{j+1} - g_j = G(x^{j+1} - x^j) = \lambda_j G d^j$, 所以,

$$
g_k^{\mathrm{T}} d^i = g_{i+1}^{\mathrm{T}} d^i + \sum_{j=i+1}^{k-1} \lambda_j (d^j)^{\mathrm{T}} G d^i, \quad \forall i \in \{0, 1, \cdots, k-1\}. \tag{3.4.3}
$$

一方面, 利用精确一维线搜索可得 $g_{i+1}^{\mathrm{T}} d^i = 0$; 另一方面, 由向量组的共轭性可得, 对任意的 $j \in \{i+1, \cdots, k-1\}$ 都有 $(d^j)^{\mathrm{T}} G d^i = 0$. 因此, 由 (3.4.3) 式可得 (3.4.2) 式成立. 定理得证. □

共轭方向法的算法框架

算法3.4.1 (二次函数极小化的共轭方向法)

设二次函数 $f(x)=\dfrac{1}{2}x^{\mathrm{T}}Gx+q^{\mathrm{T}}x$. 给定初始点$x^0\in\mathbb{R}^n$及下降方向$d^0$, 计算$g_0$.

如果$g_0=\mathbf{0}$，算法终止；否则，置$k:=0$.

步1　由精确一维搜索 $f(x^k+\lambda_k d^k)=\min\limits_{\lambda\geq 0} f(x^k+\lambda d^k)$ 计算步长λ_k.

步2　置 $x^{k+1}:=x^k+\lambda_k d^k$.

步3　计算g_{k+1}.　如果$g_{k+1}=\mathbf{0}$，算法终止.　否则, 转步4.

步4　取共轭方向d^{k+1},使得 $(d^{k+1})^{\mathrm{T}}Gd^i=0,\ \forall\ i\in\{0,1,\cdots,k\}$.

步5　置$k:=k+1$, 转步1.

如何构造共轭方向?

Gram-Schmidt正交化过程：

从线性无关组$v_0, v_1, \ldots, v_m$ $\xrightarrow{\text{构造与之等价的正交组}}$ $u_0, u_1, \ldots, u_m$

(满足$< u_i, u_j >= u_i^T u_j = 0, i \neq j$)

投影算子：$\mathrm{Proj}_u(v) = \dfrac{< v, u >}{< u, u >} u$

$$u_0 = v_0$$

$$u_1 = v_1 + \beta_0 u_0, \quad \text{其中} \beta_0 = -\frac{< v_1, u_0 >}{< u_0, u_0 >} = -\frac{v_1^T u_0}{u_0^T u_0};$$

$$u_2 = v_2 + \beta_0 u_0 + \beta_1 u_1, \text{其中} \beta_0 = -\frac{< v_2, u_0 >}{< u_0, u_0 >}, \beta_1 = -\frac{< v_2, u_1 >}{< u_1, u_1 >};$$

……

$$u_k = v_k + \sum_{i=0}^{k-1} \beta_i u_i, \quad \text{其中} \beta_i = -\frac{< v_k, u_i >}{< u_i, u_i >} = -\frac{v_k^T u_i}{u_i^T u_i}.$$

关于I共轭 $\Rightarrow$ 关于Hesse阵G共轭

把Gram-Schmidt正交化过程 $\Rightarrow$ 构造共轭方向的方法，

构造G共轭向量组 $d^0, d^1, \ldots, d^m$（满足 $<d^i, d^i>_G = (d^i)^T G d^j = 0, i \neq j$）

投影算子：$\mathrm{Proj}_u(v) = \dfrac{<v,u>_G}{<u,u>_G} u$

考虑最速下降法，每一步的迭代$d^k = -g_k$,满足$d^k \perp d^{k+1}$，$g_{k+1}^T g_k = 0.$

$d^0 = -g_0$,（取初始下降方向），

$$d^1 = -g_1 + \beta_0 d^0,\quad \text{其中}\beta_0 = -\frac{<-g_1, d^0>_G}{<d^0, d^0>_G} = \frac{g_1^{\ T} G d^0}{(d^0)^T G d^0};$$

……

$$d^k = -g_k + \sum_{i=0}^{k-1} \beta_i d^i,\quad \text{其中}\beta_i = -\frac{<-g_k, d^i>_G}{<d^i, d^i>_G} = \frac{g_k^{\ T} G d^i}{(d^i)^T G d^i}.$$

$d^0 = -g_0,$

$d^1 = -g_1 + \beta_0 d^0,$　其中取$\beta_0 = \dfrac{g_1^T G d^0}{(d^0)^T G d^0}$,保证了$d^1$与$d^0$关于G共轭;

设$d^2 = -g_2 + \beta_0 d^0 + \beta_1 d^1$,令

$0 = (d^2)^T G d^0 = -g_2^T G d^0 + \beta_0 (d^0)^T G d^0 + \beta_1 (d^1)^T G d^0 = -g_2^T G d^0 + \beta_0 (d^0)^T G d^0$

而$g_2^T G d^0 = g_2^T \dfrac{1}{\lambda_0}(g_1 - g_0) = \dfrac{1}{\lambda_0}(g_2^T g_1 - g_2^T g_0) = 0$, 故$\beta_0=0$，$d^2 = -g_2 + \beta_1 d^1.$

(备注：事实上由引理3.4.1，有　$g_2^T g_0 = -g_2^T d^0 = 0.$

$0 = g_2^T d^1 = g_2^T(-g_1 + \beta_0 d^0) = -g_2^T g_1 + \beta_0 g_2^T d^0 = -g_2^T g_1$.)

猜测取　$d^k = -g_k + \beta_{k-1} d^{k-1}$,　其中$\beta_k = \dfrac{g_k^T G d^{k-1}}{(d^{k-1})^T G d^{k-1}}$,

这样产生的方向$d^0, d^0, \ldots, d^k (k \le n-1)$关于G共轭.

§3.4.2 二次函数极小化的共轭梯度法

定理3.4.3 对于正定二次函数 $f(x)=\frac{1}{2}x^{\mathrm{T}}Gx+q^{\mathrm{T}}x$(其中$G$为$n$阶对称正定矩阵），采用精确一维线搜索的共轭梯度法的迭代方向是由以下公式产生的共轭方向$d^0,d^1,\cdots,d^{m-1}(m\le n)$：

$$\begin{cases} d^0=-g_0, \\ d^k=-g_k+\beta_{k-1}d^{k-1},\ \text{其中}\beta_{k-1}=\dfrac{g_k{}^T G d^{k-1}}{(d^{k-1})^T G d^{k-1}}, \quad 0<k\le m \end{cases}$$

共轭梯度法$m-1(m\le n)$次迭代可求得二次函数 f的极小点, 并且对于所有的$i\in\{0,1,\cdots,m-1\}$有：

(1) $(d^i)^T G d^j=0, \forall j\in\{0,1,\cdots,i-1\}$

(2) $g_i{}^T g_j=0, \forall j\in\{0,1,\cdots,i-1\}$

(3) $g_i{}^T d^{\mathrm{j}}=0, \forall j\in\{0,1,\cdots,i-1\}$

(4) $g_i{}^T d^i=-g_i{}^T g_i$

举例，利用定理3.4.3的（2)可以证明 $d^k = -g_k + \beta_{k-1} d^{k-1}$.

构造G共轭向量组 $d^0, d^1, \ldots, d^m$(满足$(d^i)^T G d^j = 0, i \neq j$)

对于正定二次函数 $f(x) = \frac{1}{2} x^{\mathrm{T}} G x + q^{\mathrm{T}} x$， $g_k = G x^k + q$,

$g_{k+1} - g_k = G(x^{k+1} - x^k)$， 将$x^{k+1} - x^k = \lambda_k d^k$代入得,

$g_{k+1} - g_k = \lambda_k G d^k$,或 $\quad G d^k = \frac{1}{\lambda_k}(g_{k+1} - g_k)$

$d^k = -g_k + \sum_{i=0}^{k-1} \beta_i d^i$,其中$\beta_i = -\frac{< -g_k, d^i >_G}{< d^i, d^i >_G} = \frac{g_k{}^T G d^i}{(d^i)^T G d^i}$,

当$i < k-1$时， $\beta_i = -\frac{< -g_k, d^i >_G}{< d^i, d^i >_G} = \frac{g_k{}^T G d^i}{(d^i)^T G d^i} = \frac{\frac{1}{\lambda_i} g_k{}^T (g_{i+1} - g_i)}{(d^i)^T G d^i} = 0.$

(根据定理3.4.3的（2）可得$g_i{}^T g_j = 0, j < i$)

所以 $d^k = -g_k + \beta_{k-1} d^{k-1}$.

§3.4.3 一般函数极小化的共轭梯度法

（一）系数 β_{k-1} 的其它形式：

(1) FR公式（Flecher-Reeves） 1964年

$$\beta_{k-1} = \frac{g_k^{\mathrm{T}} g_k}{g_{k-1}^{\mathrm{T}} g_{k-1}}.$$

(2) PRP公式（Polak-Ribiere-Polyak)

$$\beta_{k-1} = \frac{g_k^{\mathrm{T}} (g_k - g_{k-1})}{g_{k-1}^{\mathrm{T}} g_{k-1}}.$$

(3) Dixon公式 $\beta_{k-1} = -\dfrac{g_k^{\mathrm{T}} g_k}{g_{k-1}^{\mathrm{T}} d^{k-1}}.$

(4) DY公式(Dai-Yuan) $\beta_{k-1} = \dfrac{g_k^{\mathrm{T}} g_k}{(g_k - g_{k-1})^{\mathrm{T}} d^{k-1}}.$

求解一般函数极小化问题的FR共轭梯度法算法框架

（二） **算法3.4.2** （FR共轭梯度法)

设函数 $f:\mathbb{R}^n\to\mathbb{R}$,给定初始点 $x^0\in\mathbb{R}^n$,置$k:=0$.

步1　计算g_k.　如果$g_k=0$，算法终止. 否则, 转步2.

步2　取 $$d^k=\begin{cases}-g_k, & \text{若}k=0,\\ -g_k+\dfrac{g_k{}^{\mathrm{T}}g_k}{g_{k-1}{}^{\mathrm{T}}g_{k-1}}d^{k-1}, & \text{否则}.\end{cases}$$

步3　由线搜索计算步长λ_k.

步4　置$x^{k+1}:=x^k+\lambda_k d^k$，$k=k+1$,转步1.

注1:采用精确一维搜索的FR共轭梯度法、PRP共轭梯度法、Dixon共轭梯度、DY共轭梯度法均为下降算法.

注2:采用精确一维搜索的FR共轭梯度法具有全局收敛性和线性收敛速度. （定理3. 4. 4）

注3：共轭梯度法计算量小，适合于求解大规模无约束问题.

定理 3.4.4 设由问题 (3.0.1) 给出函数 f 在有界水平集 $\mathscr{L}(f)=\{x\in\mathbb{R}^n|f(x)\leqslant f(x^0)\}$ 上连续可微且有下界. 序列 $\{x^k\}$ 是由采用精确一维线搜索的 FR 共轭梯度法产生的序列. 如果 $\{x^k\}$ 有限，那么 $\{x^k\}$ 有限地终止于 f 的一个稳定点; 否则，$\{x^k\}$ 的每个聚点都是 f 的一个稳定点.

证明 不妨假设序列 $\{x^k\}$ 是无限点列. 迭代方向 d^k 是下降方向，结合函数 f 的函数值在水平集上有下界, 因而 $\{f(x^k)\}$ 是单调下降且有下界的序列. 所以，$\lim_{k\to\infty} f(x^k)$ 存在. 又由于 $\{x^k\}\subseteq\mathscr{L}$ 且 $\mathscr{L}$ 有界，所以 $\{x^k\}$ 为有界点列，因而存在收敛子列. 不失一般性，不妨设 $\lim_{k\to\infty} x^k=x^*$. 那么由 f 的连续性可知, $\lim_{k\to\infty} f(x^k)=f(\lim_{k\to\infty} x^k)=f(x^*)$.

下面采用反证法证明 $g_* = 0$. 假设 $g_* \neq 0$. 那么必存在一下降方向 d^*, 使得对任意充分小的 $\lambda_* > 0$, 有

$$f(x^* + \lambda_* d^*) < f(x^*). \tag{3.4.5}$$

而利用精确一维线搜索可知, $f(x^{k+1}) = f(x^k + \lambda_k d^k) \leqslant f(x^k + \lambda_* d^k)$, 两边取极限可得

$$f(x^*) \leqslant f(x^* + \lambda_* d^*).$$

这与 (3.4.5) 式相矛盾. 定理得证. □

例 3.4.1 采用精确一维线搜索的 FR 共轭梯度法求解下述无约束优化问题

$$\min\ f(x) = \frac{1}{2}x_1^2 + x_2^2,$$

其中初始点取为 $x^0 = (2,1)^{\mathrm{T}}$.

解 通过简单计算可得 $g(x) = (x_1, 2x_2)^{\mathrm{T}}$. 由于 $g_0 = (2,2)^{\mathrm{T}} \neq (0,0)^{\mathrm{T}}$, 因而取 $d^0 = (-2,-2)^{\mathrm{T}}$.

从 x^0 出发沿 d^0 进行精确一维线搜索，即求

$$\min\ f(x^0 + \lambda d^0) = 6\lambda^2 - 8\lambda + 3$$

的极小值点，得步长 $\lambda_0 = \frac{2}{3}$. 进而，$x^1 = x^0 + \lambda_0 d^0 = \left(\frac{2}{3}, -\frac{1}{3}\right)^{\mathrm{T}}$,

且 $g_1=(\frac{2}{3},-\frac{2}{3})^{\mathrm{T}}\neq 0.$

由FR公式，$\beta_0=\frac{g_1{}^{\mathrm{T}}g_1}{g_0{}^{\mathrm{T}}g_0}=\frac{1}{9}$，$d^1=-g_1+\beta_0 d^0=(-\frac{8}{9},\frac{4}{9})^{\mathrm{T}}.$

从 x^1 出发沿 d^1 进行精确一维线搜索，即求

$$\min\ f(x^1+\lambda d^1)=\frac{1}{162}(96\lambda^2-144\lambda+54)$$

的极小值点，得步长 $\lambda_1=\frac{3}{4}$. 进而，$x^2=x^1+\lambda_1 d^1=(0,0)^{\mathrm{T}}$，并且 $g_2=(0,0)^{\mathrm{T}}$.

因此，所求无约束优化问题的最优解为 $x^*=(0,0)^{\mathrm{T}}$，最优值为 $f^*=0$. □

n步重新开始的共轭梯度法

n步重新开始的共轭梯度法基本思想:

每迭代n次，就重新取负梯度方向作为搜索方向. 可以保证在最优解附近，通过重新取负梯度方向作为迭代方向，那么后面的迭代，将产生近似共轭方向，从而较快地收敛到最优解.

注 n步重新开始的共轭梯度法具有全局收敛性和线性收敛速度.

n步重新开始的共轭梯度法

算法 3.4.3 (n 步重新开始的 FR 共轭梯度法)　设函数 f 由问题 (3.0.1) 给出. 给定初始点 $x^0 \in \mathbb{R}^n$. 置 $k := 0$.

步 1　计算 g_k. 如果 $g_k = 0$, 算法终止. 否则, 转步 2.

步 2　若 k 是 n 的倍数, 则取 $d^k = -g_k$, 否则, 取

$$d^k = -g_k + \frac{g_k^{\mathrm{T}} g_k}{g_{k-1}^{\mathrm{T}} g_{k-1}} d^{k-1}.$$

步 3　由精确一维线搜索 $f(x^k + \lambda_k d^k) = \min_{\lambda \geqslant 0} f(x^k + \lambda d^k)$, 计算步长 λ_k.

步 4　置 $x^{k+1} := x^k + \lambda_k d^k$ 且 $k := k + 1$, 转步 1.

谢谢！