DOI: 10.13718/j.cnki.xdzk.2016.07.019

一种新的求解无约束优化问题的谱共轭梯度法[①]

李智群[1]，　林浦任[1]，　韦增欣[2]

1. 钦州学院理学院，广西 钦州 535011；2. 广西大学 数学与信息科学学院，南宁 530004

摘要：在修正的 Wei-Yao-Liu 共轭梯度法基础上，给出一种新的求解无约束优化问题的谱共轭梯度算法，该算法在强 wolfe 型线搜索下具有充分下降性和全局收敛性，数值实验结果表明该算法是有效的.

关　键　词：无约束优化；谱共轭梯度法；充分下降性；全局收敛性；数值实验

中图分类号：O224　　**文献标志码**：A　　**文章编号**：1673-9868(2016)07-0115-06

考虑无约束优化问题

$$\min\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^n\} \tag{1}$$

其中 $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ 为一阶连续可微的非线性函数，其梯度向量 $\boldsymbol{g}(\boldsymbol{x}) \triangleq \nabla f(\boldsymbol{x})$，令 $\boldsymbol{g}_k = \boldsymbol{g}(\boldsymbol{x}_k)$. 非线性共轭梯度法是解决无约束优化问题的有效方法，尤其是大规模的无约束优化问题.

非线性共轭梯度法一般的迭代公式为：

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \tag{2}$$

$$\boldsymbol{d}_k = \begin{cases} -\boldsymbol{g}_k & k = 1 \\ -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1} & k \geqslant 2 \end{cases} \tag{3}$$

其中：$\boldsymbol{d}_k$ 为搜索方向，β_k 为参数，α_k 为步长因子由线搜索决定.

文章考虑的线搜索是强 wolfe 型线搜索：

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \leqslant f(\boldsymbol{x}_k) + \delta \alpha_k \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k \tag{4}$$

$$|\boldsymbol{g}(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)^{\mathrm{T}} \boldsymbol{d}_k| \leqslant \sigma |\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k| \tag{5}$$

其中 $0 < \delta < \sigma < 1$.

不同的参数表达式 β_k 对应着不同的共轭梯度法，较著名的非线性共轭梯度法参数有 β_k^{FR}，β_k^{PRP}，β_k^{HS} 等[1].

PRP 方法是目前认为数值表现最好的共轭梯度法之一，但其收敛性质不是很好[1]. 为此，很多研究者对 PRP 方法进行改进，得到一系列修正的 PRP 方法. 比如：

韦增欣等人[2] 提出改进的 PRP 方法，我们简称为 WYL 方法，即

$$\beta_k^{\mathrm{WYL}} = \frac{\boldsymbol{g}_k^{\mathrm{T}}\left(\boldsymbol{g}_k - \dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}\boldsymbol{g}_{k-1}\right)}{\|\boldsymbol{g}_{k-1}\|^2} \tag{6}$$

根据研究表明，WYL 方法不仅有良好的数值试验结果，而且有良好的收敛性[2-3].

后来，文献[4] 对 WYL 方法作一些细微的改动，构造修正的 WYL 方法参数为

① 收稿日期：2015-03-11
基金项目：国家自然科学基金项目(11161003).
作者简介：李智群(1979-)，女，广西贵港市人，讲师，硕士，主要从事最优化的研究.

$$\beta_k^{\mathrm{NPRP}}=\frac{\|\boldsymbol{g}_k\|^2-\dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}\,|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2} \tag{7}$$

并证明了当参数 $\sigma\in\left(0,\dfrac{1}{2}\right)$ 时，NPRP方法在强wolfe型线搜索下具有充分下降性和全局收敛性. 文献[5]也对WYL方法作了一些修改，得到的修正共轭梯度法的参数为：

$$\beta_k^{*}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\left(\dfrac{\|\boldsymbol{g}_{k-1}\|}{\|\boldsymbol{g}_k\|}\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\right)}{\|\boldsymbol{g}_{k-1}\|^2}$$

该方法在wolfe型线搜索下收敛，数值表现也较好.

文献[6]提出了一类谱共轭梯度法，其搜索方向的表达式为：

$$\boldsymbol{d}_k=\begin{cases}-\boldsymbol{g}_k & k=1\\ -\dfrac{1}{\delta_k}\boldsymbol{g}_k+\beta_k\boldsymbol{d}_{k-1} & k\geqslant 2\end{cases} \tag{8}$$

这里

$$\delta_k=\frac{s_{k-1}{}^{\mathrm{T}}\boldsymbol{y}_{k-1}}{|s_{k-1}|^2}\qquad \boldsymbol{y}_{k-1}=\boldsymbol{g}_k-\boldsymbol{g}_{k-1} \tag{9}$$

是谱系数.

文献[7]为了保证谱共轭梯度算法的稳定性，对谱系数 δ_k 作了如下规定：

为了保证存在 $0<\lambda_{\min}<\lambda_{\max}$ 使得 $\lambda_{\min}<\delta_k<\lambda_{\max}$ 对所有的 k 成立，若某一步产生的 δ_{k+1} 不属于区间 $[\lambda_{\min},\lambda_{\max}]$，则令 $\delta_{k+1}=\delta_k$.

谱共轭梯度法存储需求少，能有效提高共轭梯度法的计算效率. 文献[7]在此基础上讨论了各种谱共轭梯度法的收敛性和它们的数值表现. 下面将讨论NPRP型的谱共轭梯度算法和收敛性.

1 NPRP型的谱共轭梯度算法和收敛性分析

我们在NPRP的基础上，给出NPRP型的谱共轭梯度法(简称为SNPRP方法)的参数 β_k 为：

$$\beta_k^{\mathrm{SNPRP}}=\frac{\delta_{k-1}\left(\|\boldsymbol{g}_k\|^2-\dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}\,|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|\right)}{\delta_k\|\boldsymbol{g}_{k-1}\|^2} \tag{10}$$

SNPRP算法的步骤如下：

算法1

1) 给出 $\boldsymbol{x}_1\in\mathbb{R}^n$，$0\leqslant\varepsilon$，$\boldsymbol{d}_1=-\boldsymbol{g}_1$，$k:=1$；

2) 若 $\|\boldsymbol{g}_k\|\leqslant\varepsilon$，则停止；

3) 计算步长因子 α_k，使得 α_k 满足强wolfe型线搜索条件(4)及(5)；

4) 令 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k$，$\boldsymbol{g}_{k+1}=\boldsymbol{g}(\boldsymbol{x}_{k+1})$；

5) 按(9)式计算参数 δ_{k+1}，按(10)式计算 $\beta_{k+1}{}^{\mathrm{SNPRP}}$，按(8)式计算 $\boldsymbol{d}_{k+1}$，令 $k:=k+1$，转到步骤2).

显然，SNPRP方法中的参数 β_k^{SNPRP} 满足 $0\leqslant\beta_k^{\mathrm{SNPRP}}\leqslant\dfrac{\delta_{k-1}\|\boldsymbol{g}_k\|^2}{\delta_k\|\boldsymbol{g}_{k-1}\|^2}$.

SNPRP方法具有充分下降性，即

引理1 考虑算法1，如果参数 $\sigma\in\left(0,\dfrac{1}{2}\right)$，那么对所有的 k，下式成立：

$$\frac{1-2\sigma+\sigma^k}{1-\sigma}\leqslant\frac{-\delta_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}\leqslant\frac{1-\sigma^k}{1-\sigma} \tag{11}$$

证 采用数学归纳法. 假设 $k=1$，因为 $\boldsymbol{d}_1=-\boldsymbol{g}_1$，$\delta_1=1$，(11)式显然成立.

假设对任何的 $k-1$ 结论(11)成立，下证(11)式对 k 成立.

由 $\boldsymbol{d}_k=-\dfrac{1}{\delta_k}\boldsymbol{g}_k+\beta_k^{SNPRP}\boldsymbol{d}_{k-1}$ 得

$$\frac{-\delta_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}=1-\frac{\|\boldsymbol{g}_k\|^2-\dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_k\|^2}\frac{\delta_{k-1}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}\tag{12}$$

利用(5)式得

$$1+\frac{\|\boldsymbol{g}_k\|^2-\dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|}{|\boldsymbol{g}_k|^2}\frac{\sigma\delta_{k-1}\boldsymbol{g}_{k-1}{}^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}\leqslant\frac{-\delta_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}\leqslant$$
$$1-\frac{\|\boldsymbol{g}_k\|^2-\dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_k\|^2}\sigma\frac{\delta_{k-1}\boldsymbol{g}_{k-1}{}^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}\tag{13}$$

因为

$$0\leqslant\frac{\|\boldsymbol{g}_k\|^2-\dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_k\|^2}\leqslant 1\tag{14}$$

所以(14)式变为

$$1+\frac{\sigma\delta_{k-1}\boldsymbol{g}_{k-1}{}^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}\leqslant\frac{-\delta_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}\leqslant 1-\sigma\frac{\delta_{k-1}\boldsymbol{g}_{k-1}{}^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}\tag{15}$$

由归纳假设得

$$\frac{1-2\sigma+\sigma^k}{1-\sigma}=1-\sigma\frac{1-\sigma^{k-1}}{1-\sigma}\leqslant\frac{-\delta_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}\leqslant$$
$$1+\sigma\frac{1-\sigma^{k-1}}{1-\sigma}=\frac{1-\sigma^k}{1-\sigma}\tag{16}$$

由(16)式知道，对所有的 k(11)式成立.

引理 1 充分表明，当参数 $0<\sigma<\dfrac{1}{2}$ 时，对于谱共轭梯度法 SNPRP 方法，存在常数 $C=\dfrac{1-2\sigma}{1-\sigma}>0$，使得对所有的 k 有

$$\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k\leqslant -C\frac{1}{\delta_k}\|\boldsymbol{g}_k\|^2\tag{17}$$

为了分析 SNPRP 方法的收敛性，需要对目标函数 $f(\boldsymbol{x})$ 作如下假设条件：

(i) 水平集 $\Omega=\{\boldsymbol{x}\in\mathbb{R}^n, f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_1)\}$ 有界；

(ii) $f(\boldsymbol{x})$ 的梯度函数 $\boldsymbol{g}(\boldsymbol{x})$ 在 Ω 内 Lipschitz 连续，即存在常数 $L>0$，使得对任意 $\boldsymbol{x}$，$\boldsymbol{y}\in\Omega$ 有

$$\|\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{g}(\boldsymbol{y})\|\leqslant L\|\boldsymbol{x}-\boldsymbol{y}\|\tag{18}$$

根据以上假设可知存在常数 $r>0$ 使得

$$\|\boldsymbol{g}_k\|\leqslant r\qquad\forall\boldsymbol{x}_k\in\Omega\tag{19}$$

引理 2[8-10]　考虑形如(2)和(3)式的方法，若假设(i)和(ii)成立，步长因子 α_k 由 wolfe 型线搜索方法得到，且对所有的 k 有 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k<0$，那么

$$\sum_{k\geqslant 1}\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2}<+\infty\tag{20}$$

这就是著名的 Zoutendijk 条件. 易知在上面假设条件下和强 wolfe 型线搜索下也有 Zoutendijk 条件成立.

引理 3[1]　设 $l>0$，c 为常数，如果正项级数$\{a_i\}$对所有 k 成立

$$\sum_{i=1}^{k}a_i\geqslant lk+c\tag{21}$$

则有

$$\sum_{i\geqslant 1}\frac{a_i^{\,2}}{i}=+\infty \tag{22}$$

以及

$$\sum_{k\geqslant 1}\frac{a_k^2}{\sum_{i=1}^{k}a_i}=+\infty \tag{23}$$

定理 1 若假设(i)和(ii)成立，则算法 1 产生的序列$\{\boldsymbol{x}_k,\boldsymbol{d}_k,\boldsymbol{g}_k\}$有

$$\liminf_{k\to\infty}\|\boldsymbol{g}_k\|=0 \tag{24}$$

证 反证法. 设结论不成立，则存在常数 $\gamma>0$，使得

$$\|\boldsymbol{g}_k\|\geqslant\gamma \qquad \forall k\geqslant 1 \tag{25}$$

由 $\boldsymbol{d}_k+\frac{1}{\delta_k}\boldsymbol{g}_k=\beta_k^{\mathrm{SNPRP}}\boldsymbol{d}_{k-1}$，两边取平方模移项得

$$\begin{aligned}\|\boldsymbol{d}_k\|^2=&-\frac{1}{\delta_k^2}\|\boldsymbol{g}_k\|^2-2\frac{1}{\delta_k}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k+(\beta_k^{\mathrm{SNPRP}})^2\|\boldsymbol{d}_{k-1}\|^2\leqslant\\&-\frac{1}{\delta_k^2}\|\boldsymbol{g}_k\|^2-2\frac{1}{\delta_k}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k+\left(\frac{\delta_{k-1}\|\boldsymbol{g}_k\|^2}{\delta_k\|\boldsymbol{g}_{k-1}\|^2}\right)^2\|\boldsymbol{d}_{k-1}\|^2=\\&-\frac{1}{\delta_k^2}\|\boldsymbol{g}_k\|^2-2\frac{1}{\delta_k}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k+\frac{\delta_{k-1}^2\|\boldsymbol{g}_k\|^4}{\delta_k^2\|\boldsymbol{g}_{k-1}\|^4}\|\boldsymbol{d}_{k-1}\|^2\end{aligned} \tag{26}$$

令

$$t_k=\frac{\delta_k^2\|\boldsymbol{d}_k\|^2}{\|\boldsymbol{g}_k\|^4} \qquad r_k=-\frac{\delta_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}$$

(26)式两边除以$\frac{\|\boldsymbol{g}_k\|^4}{\delta_k^2}$得到

$$t_k\leqslant t_{k-1}-\frac{1}{\|\boldsymbol{g}_k\|^2}+\frac{2r_k}{\|\boldsymbol{g}_k\|^2}$$

依次类推，且 $t_1=\frac{1}{\|\boldsymbol{g}_1\|^2}$，$r_1=1$，则有

$$t_k\leqslant-\sum_{i=1}^{k}\frac{1}{\|\boldsymbol{g}_i\|^2}+2\sum_{i=1}^{k}\frac{|r_i|}{\|\boldsymbol{g}_i\|^2} \tag{27}$$

利用(19)和(25)式，可得

$$t_k\leqslant-\frac{k}{r^2}+\frac{2}{\gamma^2}\sum_{i=1}^{k}|r_i| \tag{28}$$

故

$$t_k\leqslant\frac{2}{\gamma^2}\sum_{i=1}^{k}|r_i| \tag{29}$$

因 $t_k\geqslant 0$，由(28)式有

$$\sum_{i=1}^{k}|r_i|\geqslant\frac{\gamma^2 k}{2r^2} \tag{30}$$

根据(29),(30)式以及引理 3 知

$$\sum_{k\geqslant 1}\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2}=\sum_{k\geqslant 1}\frac{r_k^2}{t_k}=\infty$$

这与引理 2 矛盾，故(24)式成立，定理得证.

2 数值实验

本文算法的数值实验结果如表 1，算法的参数 $\delta=0.01$，$\sigma=0.1$，$\varepsilon=10^{-5}$，终止条件为 $\|\boldsymbol{g}_k\|\leqslant\varepsilon$，或

迭代次数不超过 10^4. 表 1 中 Dim 表示目标函数变量的维数；NI，NF，NG 分别表示迭代次数、目标函数值计算次数、目标函数梯度值计算次数.

表 1　数值实验

测试问题	Dim	$NI/NF/NG$	$NI/NF/NG$
FROTH2	2	15/133/27	4/61/10
BADSCB	2	15/80/24	2/53/4
JENSAM	2	8/27/17	25/152/33
HELIX	3	364/1564/593	3/59/9
BARD	3	2320/7058/3838	5/156/37
GAUSS	3	4/9/5	3/7/4
MEYER	3	7546/18439/11271	6/109/55
GULF	3	1/2/2	1/2/2
BOX	3	1/51/2	1/51/2
SING	4	16579/33435/27613	4/62/11
WOOD	4	639/1871/1066	8/217/114
KOWOSB	4	1474/3139/2445	5834/14299/6046
BD	4	56/269/173	7/209/95
WATSON	20	19936/40065/33221	3/104/5
ROSEX	8	34/423/71	20/489/48
ROSEX	50	36/431/73	3/102/5
ROSEX	100	34/331/73	6/110/10
SINGX	4	16579/33435/27613	4/62/11
PEN1	2	5/18/12	5/18/12
PEN2	4	14/181/30	79/355/157
PEN2	50	164/920/309	6/117/61
VARDIM	2	3/9/7	3/9/7
VARDIM	50	10/52/36	3/62/11
TRIG	3	17/140/32	33/259/50
TRIG	50	38/415/66	99/2431/127
IE	3	5/12/7	5/12/7
IE	50	6/13/7	7/15/8
IE	100	6/13/8	8/18/12
TRID	3	20/47/25	568/14432/796
TRID	50	31/113/38	4/59/8
TRID	100	33/73/39	4/61/9
TRID	200	32/70/38	4/62/11
BAND	3	7/64/12	15/177/26
BAND	50	18/620/25	3/115/10
BAND	100	18/760/27	6/263/14
BAND	200	18/629/26	12/563/22
LIN	2	1/3/3	1/3/3
LIN	50	1/3/3	1/3/3
LIN	500	1/3/3	1/3/3
LIN	1 000	1/3/3	1/3/3
LIN1	2	1/51/2	1/51/2
LIN1	10	1/3/3	1/3/3
LIN0	4	1/3/3	1/3/3

由表 1 可知，SNPRP 算法比 NPRP 算法在收敛性方面更好，计算效率更高.

参考文献：

[1] 戴彧虹，袁亚湘. 非线性共轭梯度法［M］. 上海：上海科学技术出版社，2001.

[2] WEIZ X，YAO S W，LIU L Y. The Convergence Properties of some Conjugate Gradient Methods［J］. Applied Mathematics and Computation，2006，183(2)：1341－1350.

[3] HUANG H，WEI Z X，YAO S W. The Proof of the Sufficient Descent Condition of the Wei-Yao-Liu Conjugate Gradient Method under the Strong Wolfe-Powell Line Search［J］. Applied Mathematics and Computation，2007，189(2)：1241－1245.

[4] ZHANG L. An Improved Wei-Yao-Liu Nonlinear Conjugate Gradient Method for Optimization Computation［J］. Applied Mathematics and Computation，2009，215(6)：2269－2274.

[5] 黎 勇. 一类修正 PRP 共轭梯度法的全局收敛性及其数值实验结果［J］. 西南大学学报(自然科学版)，2011，33(11)：23－28.

[6] BIRGIN E G，MARTINEZ J M. A Spectral Conjugate Gradient Method for Unconstrained Optimization［J］. Applied Mathematics and Optimization，2001，43(2)：117－128.

[7] 喻高航. 大规模优化问题的非线性自调比共轭梯度算法研究［D］. 广州：中山大学，2007.

[8] WOLFE P. Convergence Conditions for Ascent Methods［J］. SIAM Review，1969(11)：226－235.

[9] WOLFE P. Convergence Conditions for Ascent Methods. II：Some Correction［J］. SIAM Review，1971(31)：185－188.

[10] ZOUTENDIJK G. Nonlinear Programning，Computational Methods［M］//Integer and Nonlinear Programming. Amsterdam：North-Holland，1970.

A New Spectral Conjugate Gradient Method for Solving Unconstraints Minimization Problem

LI Zhi-qun， LIN Pu-ren， WEI Zeng-xin

1. College of Science，Qinzhou University，Qinzhou Guangxi 535011；

2. College of Mathematics and Information Science，Guangxi University，Nannin Guangxi 530004

Abstract：Base on an improved conjugate gradient method，a new spectral conjugate gradient algorithm is presented for solving unconstrained minimization problem in this paper，and the algorithm possesses sufficient descent property and global convergence with the strong Wolfe line search. Numerical results show that the new method is effective.

Key words：unconstrained optimization；spectral conjugate gradient method；sufficient descent property；global convergence；numerical experiment

责任编辑 张 枸