Doi:10.12051/j.issn.1674-4942.2021.03.003

一类改进的谱共轭梯度法

李 景,景书杰,牛海峰
(河南理工大学 数学与信息科学学院,河南 焦作 454000)

摘 要:谱共轭梯度法是在共轭梯度法基础上发展起来的新型算法,其特点是有两个方向控制参数,是解决大规模无约束优化问题的有效方法,也是优化工作者研究的热点。本文基于已有的非线性谱共轭梯度法提出了一类新的谱共轭梯度法,利用新构造的共轭方向调控参数β_k构建了新的算法,并保证了该算法在任何线搜索下都满足共轭条件,进而在迭代时产生的搜索方向都是充分下降的。在Wolfe线搜索下,该方法的全局收敛性得以验证。

关键词:无约束优化;谱共轭梯度法;Wolfe线搜索;全局收敛性

中图分类号:O224 **文献标志码**:A **文章编号**:1674-4942(2021)03-0269-05

A Kind of Improved Spectral Conjugate Gradient Method

LI Jing, JING Shujie, NIU Haifeng
(*School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo* 454000, *China*)

Abstract: The spectral conjugate gradient method is a new algorithm developed on the basis of the conjugate gradient method. Its characteristic is to control parameters in two directions. It is an effective method to solve large-scale unconstrained optimization problems and a hot topic for optimization workers. In this paper, based on the existing nonlinear spectral conjugate gradient method, a new class of spectral conjugate gradient method was proposed. Using the newly constructed conjugate direction control parameter β_k, a new algorithm was constructed, and it is guaranteed that the algorithm meets the conjugate condition under any line search, and the search direction generated during iteration is fully reduced. Under the Wolfe line search, the global convergence of the method was verified.

Keywords: unconstrained optimization; spectral conjugate gradient method; Wolfe line search; global convergence

考虑求解如下类型的无约束非线性优化问题

$$\min_{x \in \mathbf{R}^n} f(x), \tag{1}$$

其中,x是决策变量,目标函数$f:\mathbf{R}^n \to R$连续可微,梯度向量$\nabla f(x_k) = g_k$。共轭梯度法是求解式(1)问题最为常用的方法,其优点是只需要目标函数f的值和梯度向量$\nabla f(x_k)$的值,不需大量的矩阵存储,数值效果却比较显著[1]。迭代格式的一般形式为

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

$$d_k = \begin{cases} -g_k & k = 1 \\ -g_k + \beta_k d_{k-1} & k > 1 \end{cases}, \tag{3}$$

其中,k是迭代次数,d_k是搜索方向,α_k是步长因子,β_k是共轭方向调控参数。β_k的不同选取分别对应不同的共轭梯度法,著名的β_k的计算公式有[2-8]

收稿日期:2021-05-27
基金项目:国家自然科学基金项目(U1504104)
第一作者:李景(1996—),河南长垣人,硕士研究生,研究方向为运筹学与控制论。E-mail:1358213150@qq.com

$$\beta_k^{FR} = \frac{\left\|g_k\right\|^2}{\left\|g_{k-1}\right\|^2}, \beta_k^{PRP} = -\frac{g_k^{\mathrm{T}}\left(g_k - g_{k-1}\right)}{\left\|g_{k-1}\right\|^2}, \beta_k^{HS} = -\frac{g_k^{\mathrm{T}}\left(g_k - g_{k-1}\right)}{d_{k-1}^{\mathrm{T}}\left(g_k - g_{k-1}\right)},$$

$$\beta_k^{DY} = \frac{\left\|g_k\right\|^2}{d_{k-1}^{\mathrm{T}}\left(g_k - g_{k-1}\right)}, \beta_k^{CD} = -\frac{\left\|g_k\right\|^2}{g_{k-1}^{\mathrm{T}} d_{k-1}}, \beta_k^{LS} = -\frac{g_k^{\mathrm{T}}\left(g_k - g_{k-1}\right)}{d_{k-1}^{\mathrm{T}} g_{k-1}},$$

其中$\|\cdot\|$表示欧式范数。以上公式对应的算法中,FR、DY和CD方法具有较强的收敛性,但易受干扰,数值性能较差;PRP、HS和LS可以避免干扰,数值性能较好,但可能收敛性不足。优化工作者在这些公式的基础上进行各种探讨性研究,一方面工作是对经典β_k进行修改,另一方面是将不同的β_k进行组合,并保证β_k的一些基本特征,得到了一些收敛性与数值效果都相对较好的方法。

2001年,Birgin和Martinez将谱思想和共轭梯度法相结合,给出了一类谱共轭梯度法[9],并将搜索方向d_k的迭代格式推广为如下形式:

$$d_k = \begin{cases} -g_k, & k = 1 \\ -\theta_k g_k + \beta_k d_{k-1}, & k > 1 \end{cases}, \tag{4}$$

其中,$\beta_k = \dfrac{g_k^{\mathrm{T}}\left(\theta_k y_{k-1} - s_{k-1}\right)}{d_{k-1}^{\mathrm{T}} y_{k-1}}$,$\theta_k = \dfrac{s_{k-1}^{\mathrm{T}} s_{k-1}}{s_{k-1}^{\mathrm{T}} y_{k-1}}$,这里$\theta_k$是谱参数,$y_{k-1} = g_k - g_{k-1}$;$s_{k-1} = x_k - x_{k-1}$。

1 算法及其下降性

Wang等在HS方法的基础上提出了两种修正的谱共轭方法,分别称之为DHS、WHS方法[10],两种方法均选取以下的β_k:

$$\beta_k^{WHS} = \frac{g_k^{\mathrm{T}}\left(g_k - g_{k-1}\right)}{\mu\left(g_k^{\mathrm{T}} d_{k-1}\right)^2 - d_{k-1}^{\mathrm{T}} g_{k-1}}, \tag{5}$$

这里参数μ是一个常数,且$0 \leqslant \mu \leqslant 1$,不同的参数$\mu$具有不同的迭代效果。两种方法均在Wolfe线搜索下进行

$$\begin{cases} f\left(x_k + \alpha_k d_k\right) \leqslant f\left(x_k\right) + \rho \alpha_k g_k^{\mathrm{T}} d_k \\ g\left(x_k + \alpha_k d_k\right)^{\mathrm{T}} d_k \geqslant \delta g_k^{\mathrm{T}} d_k \end{cases}。 \tag{6}$$

其中,在DHS方法中,式(6)中的系数满足$0 < \delta < \sigma < \dfrac{1}{4}$,而在WHS方法中,式(6)中的系数满足$0 < \delta < \sigma < \dfrac{1}{3}$。在对目标函数的一般性假设条件下,这两种算法具有足够的下降性质,并且总是满足全局收敛的。数值结果表明,该算法优于HS共轭梯度方法。

Shan等在时间序列分析和无约束优化理论的基础上利用一种新的谱共轭梯度方法与自回归积分运动平均组合模型(FHS谱共轭—ARIMA组合模型)对实际时间序列拟合和模拟,首先提出了FHS算法。FHS满足自动下降特性,具有HS法的较好的收敛性,在一般性假设和Wolfe线搜索下,全局收敛性得以证明[11]。其中,共轭方向调控参数

$$\beta_k^{FHS} = \frac{g_k^{\mathrm{T}}\left(g_k - \dfrac{\left\|g_k\right\|}{\left\|g_{k-1}\right\|} g_{k-1}\right)}{\left(g_k^{\mathrm{T}} d_{k-1}\right)^2 - d_{k-1}^{\mathrm{T}} g_{k-1}} \tag{7}$$

也是在Wolfe线搜索下进行的,式(6)里的系数满足条件$0 < \delta < \sigma < \dfrac{1}{4}$。该算法与ARIMA模型的结合为谱共轭梯度寻求了一个新的发展方向,时间序列分析在系统描述、系统分析、预测未来、决策和控制的同时,也

充分利用了谱共轭梯度法的优势,形成了一种新的更有效的竞争混合算法。李梓等对改进的谱共轭梯度法和共轭梯度法进行了凸组合,即与ARIMA模型相结合,证明了新算法的充分下降性及其完全收敛性[12]。黄元元通过非精确求解不可微分凸优化问题的充分下降共轭梯度法,将无约束优化问题推广到非光滑优化问题上,分析了算法的可行性[13],有一定理论价值。

结合上述文献的基本思想,本文提出了一种新的混合谱共轭梯度法,简记为LH法。其中β_k^{LH}定义为

$$\beta_k^{LH}=\frac{\left\|g_k\right\|^2-\max\left\{0,\frac{\left\|g_k\right\|}{\left\|g_{k-1}\right\|}g_k^{\mathrm{T}}g_{k-1},g_k^{\mathrm{T}}g_{k-1}\right\}}{\left(g_k^{\mathrm{T}}d_{k-1}\right)^2-d_{k-1}^{\mathrm{T}}g_{k-1}},\tag{8}$$

谱系数选取为$\theta_k=1+\beta_k^{LH}\frac{g_k^{\mathrm{T}}d_{k-1}}{\left\|g_k\right\|^2}$,基于$\beta_k^{LH}$给出一种新的谱共轭梯度算法。

算法:

Step 0:给定初始点$x_1\in\mathbf{R}^n,\varepsilon>0,0<\rho<\delta<1,d_1=-g_1,k:=1$。

Step 1:计算g_k,若$\left\|g_1\right\|\leqslant\varepsilon$,停止。否则,转Step 2。

Step 2:计算搜索方法d_k。

$$d_k=\begin{cases}-g_k, & k=1\\ -\theta_k g_k+\beta_k^{LH}d_{k-1}, & k>1\end{cases},\tag{9}$$

$$\theta_k=1+\beta_k^{LH}\frac{g_k^{\mathrm{T}}d_{k-1}}{\left\|g_k\right\|^2}。\tag{10}$$

Step 3:计算步长,使其满足式(6)。

Step 4:计算$x_{k+1}=x_k+\alpha_k d_k,g_{k+1}=g\left(x_{k+1}\right)$。若$\left\|g_{k+1}\right\|\leqslant\varepsilon$,则停止。

Step 5:$k:=k+1$,转Step 2。

引理1.1 设序列$\left\{g_k\right\},\left\{d_k\right\}$都由上述算法产生,并且满足充分下降性,则对任意$k\geqslant1$,都有

$$g_k^{\mathrm{T}}d_k=-\left\|g_k\right\|^2。\tag{11}$$

证明 (1)$k=1$时,由式(9),有$d_1=-g_1$,则$g_k^{\mathrm{T}}d_k=g_1^{\mathrm{T}}d_1=-\left\|g_1\right\|^2<0$成立。

(2)$k>1$时,由式(9),有$d_k=-\theta_k g_k+\beta_k^{LH}d_{k-1}$,结合式(10),

$$\begin{aligned}g_k^{\mathrm{T}}d_k&=-\theta_k\left\|g_k\right\|^2+\beta_k^{LH}g_k^{\mathrm{T}}d_{k-1}\\&=-\left(1+\beta_k^{LH}\frac{g_k^{\mathrm{T}}d_{k-1}}{\left\|g_k\right\|^2}\right)\left\|g_k\right\|^2+\beta_k^{LH}g_k^{\mathrm{T}}d_{k-1}\\&=-\left\|g_k\right\|^2。\end{aligned}$$

因此,该方法对于任意的k都满足充分下降条件。

引理1.2 设序列$\left\{g_k\right\}$和$\left\{d_k\right\}$由算法A生成,对于任意的$k\geqslant1$,有$0<\beta_k^{LH}<\frac{\left\|g_k\right\|^2}{\left\|g_{k-1}\right\|^2}$成立。

证明 由β_k^{LH}的定义式(10)显然可得$\beta_k^{LH}\leqslant\frac{\left\|g_k\right\|^2}{\left(g_k^{\mathrm{T}}d_{k-1}\right)^2-d_{k-1}^{\mathrm{T}}g_{k-1}}$,由引理1.1中式(11)可知$g_{k-1}^{\mathrm{T}}d_{k-1}=-\left\|g_{k-1}\right\|^2$,那么$\beta_k^{LH}\leqslant\frac{\left\|g_k\right\|^2}{\left(g_k^{\mathrm{T}}d_{k-1}\right)^2+\left\|g_{k-1}\right\|^2}\leqslant\frac{\left\|g_k\right\|^2}{\left\|g_{k-1}\right\|^2}$,结合$\beta_k^{LH}$的定义,可以得到$0<\beta_k^{LH}<\frac{\left\|g_k\right\|^2}{\left\|g_{k-1}\right\|^2}$。

2 算法收敛性

基本假设H[14]:

(H1)目标函数$f(x)$在水平集$\Omega=\left\{x\in\mathbf{R}^n \middle| f(x)\leqslant f\left(x_1\right)\right\}$上有下界,其中$x_1$为初始迭代点。

(H2)目标函数$f(x)$在水平集Ω的某个邻域Λ内是连续可微,并且梯度函数$g(x)$满足Lipschitz连续,即存在常数$L>0$,使得

$$\left\|g(x)-g\left(y\right)\right\|=L\left\|x-y\right\|,\forall x,y\in\Lambda。\tag{12}$$

引理2.1 假设H成立,如果搜索方向d_k满足$g_k^{\mathrm{T}}d_k<0$,由算法1产生的序列$\left\{g_k\right\}$和$\left\{d_k\right\}$满足Zoutendijk条件,那么$\sum\frac{\left(g_k^{\mathrm{T}}d_k\right)^2}{\left\|d_k\right\|^2}<+\infty$。

定理1 假设H成立,算法1产生的序列$\left\{g_k\right\}$有

$$\lim_{k\to\infty}\inf\left\|g_k\right\|=0。\tag{13}$$

证明 用反证法证明,假设结论不成立,则存在常数$\gamma>0$,使得

$$\left\|g_k\right\|\geqslant\gamma,k\geqslant 1。\tag{14}$$

由式(9)可知

$$d_k+\theta_k g_k=\beta_k^{LH}d_{k-1},\tag{15}$$

把上式两边取平方模

$$\left(d_k+\theta_k g_k\right)^{\mathrm{T}}\left(d_k+\theta_k g_k\right)=\left(\beta_k^{LH}\right)^2\left\|d_{k-1}\right\|^2,\tag{16}$$

展开有

$$\left\|d_k\right\|^2+2\theta_k g_k^{\mathrm{T}}d_k+{\theta_k}^2\left\|g_k\right\|^2=\left(\beta_k^{LH}\right)^2\left\|d_{k-1}\right\|^2,\tag{17}$$

移项有

$$\left\|d_k\right\|^2=\left(\beta_k^{LH}\right)^2\left\|d_{k-1}\right\|^2-2\theta_k g_k^{\mathrm{T}}d_k-{\theta_k}^2\left\|g_k\right\|^2,\tag{18}$$

两边同时除以$\left(g_k^{\mathrm{T}}d_k\right)^2$有

$$\frac{\left\|d_k\right\|^2}{\left(g_k^{\mathrm{T}}d_k\right)^2}=\frac{\left(\beta_k^{LH}\right)^2\left\|d_{k-1}\right\|^2}{\left(g_k^{\mathrm{T}}d_k\right)^2}-\frac{2\theta_k}{g_k^{\mathrm{T}}d_k}-\frac{{\theta_k}^2\left\|g_k\right\|^2}{\left(g_k^{\mathrm{T}}d_k\right)^2}。\tag{19}$$

由引理1.2可以得知

$$0<\beta_k^{LH}<\frac{\left\|g_k\right\|^2}{\left\|g_{k-1}\right\|^2},$$

进而有

$$\frac{\left\|d_k\right\|^2}{\left(g_k^{\mathrm{T}}d_k\right)^2}\leqslant\frac{\left\|g_k\right\|^4}{\left\|g_{k-1}\right\|^4}\cdot\frac{\left\|d_{k-1}\right\|^2}{\left(g_k^{\mathrm{T}}d_k\right)^2}+2\frac{\theta_k}{g_k^{\mathrm{T}}d_k}-\frac{\theta_k^2\left\|g_k\right\|^2}{\left(g_k^{\mathrm{T}}d_k\right)^2},$$

结合式(11)可得

$$\begin{aligned}\frac{\left\|d_k\right\|^2}{\left(g_k^{\mathrm{T}}d_k\right)^2}&\leqslant\frac{\left\|d_{k-1}\right\|^2}{\left\|g_{k-1}\right\|^4}+2\frac{\theta_k}{\left\|g_k\right\|^2}-\frac{\theta_k^2}{\left\|g_k\right\|^2}=\frac{\left\|d_{k-1}\right\|^2}{\left\|g_{k-1}\right\|^4}-\frac{1}{\left\|g_k\right\|^2}\left(\theta_k^2-2\theta_k+1-1\right)\\&=\frac{\left\|d_{k-1}\right\|^2}{\left\|g_{k-1}\right\|^4}-\frac{\left(\theta_k-1\right)^2}{\left\|g_k\right\|^2}+\frac{1}{\left\|g_k\right\|^2}\leqslant\frac{\left\|d_{k-1}\right\|^2}{\left\|g_{k-1}\right\|^4}+\frac{1}{\left\|g_k\right\|^2}。\end{aligned}\tag{20}$$

当$k = 1$时,由式(10)显然可以得到$\frac{\left\|d_1\right\|^2}{\left(g_1^{\mathrm{T}} d_1\right)^2} = \frac{1}{\left\|g_1\right\|^2}$。

当$k > 1$时,由式(20)可得

$$\sum \frac{\left\|d_k\right\|^2}{\left(g_k^{\mathrm{T}} d_k\right)^2} \leqslant \sum \frac{1}{\left\|g_i\right\|^2} + l = \frac{k}{\gamma^2} + l, \tag{21}$$

那么

$$\sum \frac{\left(g_k^{\mathrm{T}} d_k\right)^2}{\left\|d_k\right\|^2} \geqslant r^2 \sum \frac{1}{k + r^2 l} \rightarrow +\infty, \tag{22}$$

与上述引理2.1矛盾,故

$$\lim_{k \to \infty} \inf \left\|g_k\right\| = 0。$$

即定理1得证。

参考文献:

[1] 王宜举,修乃华. 非线性规划理论与算法[M]. 西安:陕西科学技术出版社,2008.

[2] FLETCHER R, REEVES C M. Function minimization by conjugate gradients[J]. The Computer Journal, 1964, 7(2): 149-154.

[3] POLAK E, RIBIERE G. Note sur la convergence de méthodes de directions conjuguées[J]. Revue Française d'Informatique et De Recherche Opérationnelle Série Rouge, 1969, 3(16): 35-43.

[4] POLYAK B T. The conjugate gradient method in extremal problems[J]. USSR Computational Mathematics and Mathematical Physics, 1969, 9(4): 94-112.

[5] HESTENES M R, STIEFEL E. Methods of conjugate gradients for solving linear systems[J]. Journal of Research of the National Bureau of Standards, 1952, 49(6): 409-436.

[6] DAI Y H, YUAN Y. A nonlinear conjugate gradient method with a strong global convergence property[J]. SIAM Journal on Optimization, 1999, 10(1): 177-182.

[7] HAGER W W, ZHANG H C. A new conjugate gradient method with guaranteed descent and an efficient line search[J]. SIAM Journal on Optimization, 2005, 16(1): 170-192.

[8] LIU Y, STOREY C. Efficient generalized conjugate gradient algorithms, part 1: Theory[J]. Journal of Optimization Theory and Applications, 1991, 69(1): 129-137.

[9] BIRGIN E G, MARTÍNEZ J M. A spectral conjugate gradient method for unconstrained optimization[J]. Applied Mathematics and Optimization, 2001, 43(2): 117-128.

[10] WANG G F, SHAN R, HUANG W, et al. Two new spectral conjugate gradient algorithms based on Hestenes-Stiefel[J]. Journal of Algorithms & Computational Technology, 2017, 11(4): 345-352.

[11] SHAN R, WANG G F, HUANG W, et al. A new spectral conjugate gradient method and ARIMA combined forecasting model and application[J]. Journal of Algorithms & Computational Technology, 2018, 12(3): 245-252.

[12] 李梓,单锐. 基于凸组合共轭梯度法的ARIMA模型参数估计[J]. 数学的实践与认识,2021,51(1):223-229.

[13] 黄元元. 求解不可微凸优化问题的充分下降共轭梯度法[J]. 河南科技大学学报(自然科学版),2021,42(2):94-99,10.

[14] 简金宝,江羡珍,尹江华. 非线性共轭梯度法研究进展[J]. 玉林师范学院学报,2016,37(2):3-10.

责任编辑:刘 炜