一种新的凸组合形式的混合共轭梯度法*

陈秀芳

（滇西科技师范学院 数理学院，云南 临沧 677000）

［摘 要］本文提出一种新的混合共轭梯度法来求解无约束优化问题，该方法由文献［13］简记（MHS 方法）和 Dai-Yuan（DY 方法）作凸组合，得到新的共轭参数 β_k^{HXF}，且凸组合形式下的共轭梯度法产生的搜索方向满足共轭条件，进一步证明该方法不依赖于任何线搜索而满足充分下降性．最后，在一般假设条件下，使用强 wolfe 线搜索，证明了新算法框架具有全局收敛性.

［关键词］凸组合；混合共轭梯度法；全局收敛性；充分下降性

［中图分类号］O221. 2 ［文献标志码］A ［文章编号］1674 – 5639（2023）03 – 0050 – 08

DOI：10. 14091/j. cnki. kmxyxb. 2023. 03. 008

共轭梯度法发展历史悠久，最初由 Hestenes 和 Stiefel[1] 于 1952 年求解线性方程组提出，随后，1964 年，由 Fletcher 和 Reeves[2] 推广到非线性优化领域，且共轭梯度法从求解二次函数极小化问题入手，逐步推广到求解大规模无约束优化问题，比如在石油勘探、大气模拟、航空航天、图像识别、压缩感知及其他应用领域中的问题. 然求解无约束优化问题的方法不在少数，如牛顿法，虽收敛速度快，但每一步迭代中都需要计算二阶导数，且计算量复杂，拟牛顿法也如此. 相比之下，非线性共轭梯度法是求解无约束优化问题最有效的方法之一，其具有算法简单、存储性低和数值性能较好的特点，被广泛用于求解生活中的实际问题，如信号恢复、图像处理、深度学习和最优控制等领域[3-6]，本文进一步研究非线性共轭梯度法，考虑以下非线性无约束优化问题：

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

其中，f：$R^n \to R$ 为连续可微单调函数，给定初始点 $x_0 \in \boldsymbol{R}^n$，共轭梯度法产生迭代系列 $\{x_k\} \subset R^n$，满足

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

α_k 为步长，由某种线搜索决定，d_k 为搜索方向，满足：

$$d_k = \begin{cases} -g_k, & k=1; \\ -g_k + \beta_k d_{k-1}, & k \geqslant 2. \end{cases} \tag{3}$$

其中，$g_k = g(x_k) = \nabla f(x_k)$，精确线搜索需要计算一个单变量函数的极小值，成本高且计算复杂，故在设计算法时一般采用非精确线搜索，如强 wolfe 线搜索（SWP）和弱 wolfe（WWP）线搜索[2]，公式分别记为：

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \delta\alpha_k g_k^T d_k, \quad |g_{k+1}^T d_k| \leqslant -\sigma g_k^T d_k; \tag{4}$$

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \delta\alpha_k g_k^T d_k, \quad g_{k+1}^T d_k \geqslant \sigma g_k^T d_k. \tag{5}$$

记 $y_k = g_{k+1} - g_k$，参数 δ 和 σ 满足 $0 < \delta < \sigma < 1$，β_k 为共轭参数，不同的共轭参数和步长决定不同的共轭梯度法. 传统共轭梯度法分为两大类：一类是数值性能一般，而收敛性好的算法，如 FR 方法[7]、DY 方法[8] 和 CD 方法[7]；另一类是收敛性不是很好，数值性能较好的算法，如 PRP 方法[9]、HS 方法[10] 和 LS 方法[11]，对应的共轭参数公式有：

---

* ［收稿日期］2022 – 12 – 07

［作者简介］陈秀芳，女，云南曲靖人，滇西科技师范学院助教，硕士，研究方向为最优化理论与算法、高等数学教学.

［基金项目］云南省教育厅科学研究基金项目（2022J1021）；滇西科技师范学院校级科研项目（DXXY202226）.

$$\beta_k^{FR} = \frac{g_k^T g_k}{\|g_{k-1}\|^2},\ \beta_k^{DY} = \frac{g_k^T g_k}{d_{k-1}^T(g_k - g_{k-1})},\ \beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}; \tag{6}$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2},\ \beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})},\ \beta_k^{LS} = \frac{g_k^T(g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}}. \tag{7}$$

考虑到以上两类方法各有优缺，故许多学者在两类方法上进行改进，或将两类方法与其他方法混合，在传统二项共轭梯度法的基础上进行改进得到凸组合，或者修改谱系数、三项和更多项形式的共轭梯度法，这些新方法的共同点都是改进共轭系数 β_k，从而得到一个收敛性质较好，数值性能也较好的共轭梯度算法，相关研究参见文献：

2006 年，Wei 等[12]改进经典 PRP 方法，其参数为：

$$\beta_k^{WYL} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{\|g_{k-1}\|^2}, \tag{8}$$

Wei 等分别在两种线搜索下，证明改进后的 WYL 方法较 PRP 方法更有效，具有良好的收敛性和数值性能，能解决更多的测试问题.

2007 年，Yao 等[13]延续 Wei 等人的思想，在 HS 方法上作细微的改动，修改参数 β_k^{HS} 得到 β_k^{MHS}，其参数为：

$$\beta_k^{MHS} = \frac{g_k^T\left(g_k - \frac{g_k}{g_{k-1}}g_{k-1}\right)}{d_{k-1}^T(g_k - g_{k-1})}. \tag{9}$$

在一般 wolfe 线搜索下，修改后的 MHS 方法比传统的 DY 方法和 HS 方法更有效．经典共轭梯度法的变体很多，其中，凸组合形式的共轭梯度法也深受许多学者的青睐. 近年来，文献［14－16］给出：任何一种形式的共轭梯度法，都是其他两种共轭梯度法凸组合的构成，其搜索方向 d_k 满足牛顿方向，在某种特定条件下，d_k 也满足 DL 共轭条件，以及经典共轭条件；在经典共轭梯度法基础上，修改共轭参数，其目的都是寻找收敛性和数值性能更好的算法.

通过整理文献，给出以下几种比较常见凸组合形式的混合共轭梯度法（表 1）.

表 1　几种常见凸组合形式的混合共轭梯度法

序号	公式	作者
1	$\beta_k^c = (1-\theta_k)\beta_k^{HS} + \theta_k\beta_k^{DY}$	Andiei[17]
2	$\beta_k^{hyb} = (1-\theta_k)\beta_k^{HS} + \theta_k\beta_k^{FR}$	Djordjevic’[18]
3	$\beta_k^{hyb} = (1-\theta_k)\beta_k^{LS} + \theta_k\beta_k^{FR}$	Djordjevic`[19]
4	$\beta_k = (1-\theta_k)\beta_k^{LS} + \theta_k\beta_k^{DY}$	Liu 等[20]
5	$\beta_k^{hFRBA} = (1-\theta_k)\beta_k^{FR} + \theta_k\beta_k^{BA}$	Delladji[21]

受以上文献的思路启发，本文利用 β_k^{MHS} 方法数值性能好的优点和 β_k^{DY} 方法很强的收敛性来重新定义共轭系数 β_k，记为 β_k^{HXF}，由 β_k^{MHS} 和 β_k^{DY} 作凸组合，公式为：

$$\beta_k^{HXF} = (1-\theta_k)\beta_k^{MHS} + \theta_k\beta_k^{DY}, \tag{10}$$

记

$$d_0 = -g_0,\ d_{k+1} = -g_{k+1} + \beta_k^{HXF} d_k, \tag{11}$$

其中，θ_k 为凸组合参数，$0 \leq \theta_k \leq 1$.

$$\beta_k^{HXF} = (1-\theta_k)\frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{d_{k-1}^T(g_k - g_{k-1})} + \theta_k\frac{g_k^T g_k}{d_{k-1}^T(g_k - g_{k-1})}, \tag{12}$$

将（12）式代入（11）式，

$$d_{k+1} = -g_{k+1} + (1-\theta_k)\frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{d_{k-1}^T(g_k - g_{k-1})}d_k + \theta_k\frac{g_k^T g_k}{d_{k-1}^T(g_k - g_{k-1})}d_k,$$

若 $\theta_k=0$，$\beta_k^{HXF}=\beta_k^{MHS}$；若 $\theta_k=1$，$\beta_k^{HXF}=\beta_k^{DY}$；若 $0<\theta_k<1$，β_k^{HXF} 是 β_k^{MHS} 和 β_k^{DY} 的凸组合. 后文将介绍新的混合共轭梯度法，获得凸组合参数 θ_k，证明在强 wolfe 线搜索下算法的充分下降性，并在一般假设条件下讨论算法的全局收敛性.

1 新算法及其下降性

假定在每一步迭代中，θ_k 不依赖于任何线搜索而满足共轭条件：$d_{k+1}^T y_k=0$，

定义搜索方向：

$$d_{k+1}=-g_{k+1}+(1-\theta_k)\frac{g_k^T\left(g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{d_{k-1}^T(g_k-g_{k-1})}d_k+\theta_k\frac{g_k^Tg_k}{d_{k-1}^T(g_k-g_{k-1})}d_k, \tag{13}$$

进一步，等式两边同时乘以 y_k，得到：

$$d_{k+1}^Ty_k=-g_{k+1}^Ty_k+(1-\theta_k)\frac{g_k^T\left(g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{d_{k-1}^T(g_k-g_{k-1})}d_k^Ty_k+\theta_k\frac{g_k^Tg_k}{d_{k-1}^T(g_k-g_{k-1})}d_k^Ty_k,$$

$$0=-g_{k+1}^Ty_k+(1-\theta_k)\frac{g_k^T\left(g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{d_{k-1}^T(g_k-g_{k-1})}d_k^Ty_k+\theta_k\frac{g_k^Tg_k}{d_{k-1}^T(g_k-g_{k-1})}d_k^Ty_k,$$

解出 θ_k，并记为 θ_k^N，则：

$$\theta_k^N=\frac{g_{k+1}^Ty_k-\dfrac{g_k^T\left(g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{d_{k-1}^Ty_{k-1}}}{\dfrac{g_k^Tg_k}{d_{k-1}^Ty_{k-1}}d_k^Ty_k-\dfrac{g_k^T\left(g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{d_{k-1}^Ty_{k-1}}d_k^Ty_k},$$

进一步化简 θ_k^N 为：

$$\theta_k^N=\frac{g_{k+1}^Ty_k\cdot d_{k-1}^Ty_{k-1}-g_k^T\left(g_k-\frac{g_k}{g_{k-1}}g_{k-1}\right)\cdot d_k^Ty_k}{g_k{}^2d_k^Ty_k-g_k^T\left(g_k-\frac{g_k}{g_{k-1}}g_{k-1}\right)d_k^Ty_k}. \tag{14}$$

HMHSDY 方法如下：

步骤1：给定 $x_0\in\boldsymbol{R}^n$，$\varepsilon\geqslant0$，计算 $f(x_0)$ 和 $g(x_0)$，设 $d_0=-g_0$，假定 $\alpha_0=\frac{1}{\|g_0\|}$，若 $\|g_0\|\leqslant\varepsilon$，则停止；不然，进行步骤2；

步骤2：由强 wolfe 线搜索（4）计算 α_k；

步骤3：$x_{k+1}=x_k+\alpha_kd_k$，$g_{k+1}=g(x_{k+1})$，计算 s_k 和 y_k；

步骤4：若 $\|g_k\|^2d_k^Ty_k-g_k^T\left(g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)d_k^Ty_k=0$，则有 $\theta_k=0$，不然 $\theta_k=\theta_k^N$ 来计算；

步骤5：若 $\theta_k\leqslant0$，计算 $\beta_k=\beta_k^{MHS}$；若 $\theta_k\geqslant1$，计算 $\beta_k=\beta_k^{DY}$；若 $0<\theta_k<1$，β_k 由（12）式计算；

步骤6：计算 $d=-g_{k+1}+\beta_k^{HXF}d_k$，若 Powell 重启条件

$$|g_{k+1}^Tg_k|\geqslant0.2\|g_{k+1}\|^2, \tag{15}$$

满足，则有 $d_{k+1}=-g_{k+1}$，不然，$d_{k+1}=d$，设初始点 $\alpha_k=1$；

步骤7：令 $k:=k+1$，返回步骤1.

引理1.1 若（11）和（12）式成立，则有 $d_{k+1}^{HMHSDY}=(1-\theta_k)d_{k+1}^{MHS}+\theta_kd_{k+1}^{DY}$成立.

证： 因

$$d_{k+1}=-g_{k+1}+\beta_k^{HXF}d_k, \tag{16}$$

将（10）式代入（16）式有：

$$d_{k+1}^{HMHSDY} = -(\theta_k g_{k+1} + (1-\theta_k) g_{k+1}) + \beta_k^{HXF} d_k = -(\theta_k g_{k+1} + (1-\theta_k) g_{k+1}) + (1-\theta_k)\beta_k^{MHS} d_k + \theta_k \beta_k^{DY} d_k =$$
$$(1-\theta_k)(-g_{k+1} + \beta_k^{MHS} d_k) + \theta_k(-g_{k+1} + \beta_k^{DY} d_k) = (1-\theta_k) d_{k+1}^{MHS} + \theta_k d_{k+1}^{DY}.$$

因此，

$$d_{k+1}^{HMHSDY} = (1-\theta_k) d_{k+1}^{MHS} + \theta_k d_{k+1}^{DY}. \tag{17}$$

故引理 1.1 得证.

新的混合共轭梯度法在任何线搜索下都具有充分下降性，则有：

定理 1.1　序列 $\{g_k\}$ 和 $\{d_k\}$ 由 HMHSDY 方法产生，则有搜索 d_k 满足充分下降条件：

$$g_k^T d_k \leqslant -c\|g_k\|^2,\ \forall k \geqslant 0,\ 0 < c < 1. \tag{18}$$

证：由 HMHSDY 方法知，若（15）式成立，则有 $d_k = -g_k$ 和（18）式成立；

因此，假定（15）式不成立，则有 $|g_{k+1}^T g_k| < 0.2\|g_{k+1}\|^2$.

当 $k=0$ 时，$g_0^T d_0 = -\|g_0\|^2$，则（18）式成立；（17）式两端同乘 g_{k+1}^T，得：

$$g_{k+1}^T d_{k+1}^{HMHSDY} = (1-\theta_k) g_{k+1}^T d_{k+1}^{MHS} + \theta_k g_{k+1}^T d_{k+1}^{DY}. \tag{19}$$

1）当 $\theta_k = 0$ 时，$d_{k+1}^{HMHSDY} = d_{k+1}^{MHS}$.

记 $d_{k+1}^{MHS} = -g_{k+1} + \beta_k^{MHS} d_k$，两边同乘 g_{k+1}^T：

$$g_{k+1}^T d_{k+1}^{MHS} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T\left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right)}{d_k^T y_k} \cdot (g_{k+1}^T d_k),$$

$$g_{k+1}^T d_{k+1}^{MHS} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T d_k}{d_k^T y_k}\|g_{k+1}\|^2 - \frac{\frac{\|g_{k+1}\|}{\|g_k\|}(g_{k+1}^T g_k)(g_{k+1}^T d_k)}{d_k^T y_k}.$$

记 $A = \dfrac{\|g_{k+1}\|}{\|g_k\|} > 0$，则有：

$$g_{k+1}^T d_{k+1}^{MHS} \leqslant -\|g_{k+1}\|^2 + \left|\frac{g_{k+1}^T d_k}{d_k^T y_k}\right| \cdot \|g_{k+1}\|^2 + \left|\frac{g_{k+1}^T d_k}{d_k^T y_k}\right| \cdot |g_{k+1}^T g_k|. \tag{20}$$

若（15）式成立，则 $d_{k+1} = -g_{k+1}$，进而 $g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2$，因此 d_{k+1} 满足充分下降条件；若（15）式不成立，则有 $|g_{k+1}^T g_k| < 0.2\|g_{k+1}\|^2$.

由强 wolfe 线搜索的第 2 个条件知：

$|g_{k+1}^T d_k| \leqslant -\sigma g_k^T d_k$，进而有 $\sigma g_k^T d_k \leqslant g_{k+1}^T d_k \leqslant -\sigma g_k^T d_k$.

$$y_k^T d_k \geqslant \sigma g_k^T d_k - g_k^T d_k = -(1-\sigma) g_k^T d_k, \tag{21}$$

则有：

$$\left|\frac{g_{k+1}^T d_k}{y_k^T d_k}\right| \leqslant \frac{\sigma}{1-\sigma}. \tag{22}$$

（20）式化简得：

$$g_{k+1}^T d_{k+1}^{MHS} \leqslant -\|g_{k+1}\|^2 + \frac{\sigma}{1-\sigma}\|g_{k+1}\|^2 + \frac{0.2\sigma}{1-\sigma}\|g_{k+1}\|^2 \leqslant -\|g_{k+1}\|^2 + \frac{1.2\sigma}{1-\sigma}\|g_{k+1}\|^2 \leqslant -\frac{1-2.2\sigma}{1-\sigma}\|g_{k+1}\|^2$$

只需 $1-2.2\sigma > 0$，解得 $\sigma < \dfrac{5}{11}$.

$$g_{k+1}^T d_{k+1}^{MHS} \leqslant -c_1 g\|g_{k+1}\|^2，其中，c_1 = \frac{1-2.2\sigma}{1-\sigma},\ \sigma < \frac{5}{11}.$$

2）当 $\theta_k = 1$ 时，$d_{k+1} = d_{k+1}^{DY}$.

由强 wolfe 线搜索第 2 式变形得：

$$y_k^T d_k \geqslant \sigma g_k^T d_k - g_k^T d_k = -(1-\sigma) g_k^T d_k \geqslant 0.$$

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \beta_k d_k^T g_{k+1}. \tag{23}$$

结合（22）式：

$$d_{k+1}^T g_{k+1} \leqslant -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{d_k^T y_k} |d_k^T g_{k+1}| \leqslant -\frac{1-2\sigma}{1-\sigma}\|g_{k+1}\|^2.$$

只需 $1-2\sigma>0$，解得 $\sigma<\dfrac{1}{2}$；

$g_{k+1}^T d_{k+1}^{DY} \leqslant -c_2 g\|g_{k+1}\|^2$，其中，$c_2=\dfrac{1-2\sigma}{1-\sigma}$，$\sigma<\dfrac{1}{2}$.

3）当 $0<\theta_k<1$，β_k 由（10）式计算，结合（23）式：

当 $\|g_{k+1}\|^2 > \dfrac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k$ 时，则有 $\left|\dfrac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{d_k^T y_k}\right| \leqslant \left|\dfrac{\|g_{k+1}\|^2}{d_k^T y_k}\right|$.

$$\begin{aligned} d_{k+1}^T g_{k+1} &\leqslant -\|g_{k+1}\|^2 + |\beta_k^{MHS}| \cdot |d_k^T g_{k+1}| + |\beta_k^{DY}| \cdot |d_k^T g_{k+1}| \leqslant \\ &-\|g_{k+1}\|^2 + \left|\frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{d_k^T y_k}\right| \cdot |d_k^T g_{k+1}| + \frac{\|g_{k+1}\|^2}{|d_k^T y_k|} \cdot |d_k^T g_{k+1}| \leqslant \\ &-\|g_{k+1}\|^2 + \left|\frac{\|g_{k+1}\|^2}{d_k^T y_k}\right| \cdot |d_k^T g_{k+1}| + \frac{\|g_{k+1}\|^2}{|d_k^T y_k|} \cdot |d_k^T g_{k+1}| \leqslant \\ &-\|g_{k+1}\|^2 + \left|\frac{g_{k+1}^T d_k}{d_k^T y_k}\right| \cdot \|g_{k+1}\|^2 + \left|\frac{g_{k+1}^T d_k}{d_k^T y_k}\right| \cdot g_{k+1}{}^2 \leqslant \\ &-\|g_{k+1}\|^2 + \frac{\sigma}{1-\sigma}\|g_{k+1}\|^2 + \frac{\sigma}{1-\sigma}\|g_{k+1}\|^2 = -\frac{1-3\sigma}{1-\sigma}\|g_{k+1}\|^2. \end{aligned}$$

只需 $1-3\sigma>0$，得 $\sigma<1/3$.

因此，$g_{k+1}^T d_{k+1}^{HMHSDY} \leqslant -c_3\|g_{k+1}\|^2$，其中，$c_3=\dfrac{1-3\sigma}{1-\sigma}$，$\sigma<1/3$.

综上所述，

$$g_{k+1}^T d_{k+1}^{HMHSDY} \leqslant -c\|g_{k+1}\|^2. \tag{24}$$

进而搜索方向 d_k 满足充分下降条件（18）式，证毕.

2 算法的全局收敛性

为证明 HMHSDY 方法的全局收敛性，对给定目标函数做出如下假设：

假设 H

1）目标函数 $f(x)$ 在水平集 $\psi=\{x\in\boldsymbol{R}^n \mid f(x)\leqslant f(x_1)\}$ 上有下界，x_1 是 HMHSDY 方法的初始点；

2）目标函数 $f(x)$ 在水平集 ψ 的一个邻域 ϑ 内连续可微，其梯度 $g(x)$ 满足 Lipschitz 连续，即存在常数 $L>0$，使得

$$\|g(x)-g(y)\| \leqslant L\|x-y\|,\ \ \forall x,\ y\in\boldsymbol{N}, \tag{25}$$

在假设条件下，存在常数 $\eta\geqslant 0$，使得 $\|g(x)\|\leqslant\eta$，$\forall x\in\boldsymbol{N}$.

下面给出 Zoutendijk 条件.

引理 2.1 若假设条件 H(1) 和 H(2) 成立，考虑迭代序列 $x_{k+1}=x_k+\alpha_k d_k$，若搜索方向 d_k 满足下降条件，则有 $\sum\limits_{k=1}^{\infty}\dfrac{(g_k^T d_k)^2}{\|d_k\|^2}<\infty$ 成立，特别地，当搜索方向 d_k 满足充分下降条件 $g_k^T d_k\leqslant -c\|g_k\|^2$，$(0<c<1)$ 时，进一步有 $\sum\limits_{k=1}^{\infty}\dfrac{\|g_k\|^4}{\|d_k\|^2}<\infty$.

证：

$$|g_{k+1}^T d_k| \leqslant -\sigma g_k^T d_k. \tag{26}$$

有：

$$(\sigma-1)g_k^T d_k \leqslant d_k^T(g_{k+1}-g_k) \leqslant \|d_k\|\|g_{k+1}-g_k\| \leqslant L\alpha_k\|d_k\|^2,$$

$$\alpha_k \geqslant \frac{(\sigma-1)g_k^T d_k}{L\|d_k\|^2} > 0\ . \tag{27}$$

把（27）式代入（4）中：

$$f(x_k+\alpha_k d_k)\leqslant f(x_k)-\frac{\delta(1-\sigma)}{L}\frac{(g_k^T d_k)^2}{d_k^{\ 2}}，令\mu=\delta\frac{1-\sigma}{L}>0.$$

得：

$$f(x_k)-f(x_{k+1})\geqslant\mu\frac{(g_k^T d_k)^2}{\|d_k\|},$$

因此，

$$\sum_{k=0}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2}\leqslant\sum_{k=0}^{\infty}f(x_k)-f(x_{k+1}).\tag{28}$$

由假设条件 H(1) 知，$\{f_k\}$ 是一个有下界的递减数列. 因此，数列 $\{f_k\}$ 是收敛的，下面证明级数 $\sum_{k=0}^{\infty}f(x_k)-f(x_{k+1})$ 的收敛性.

令 $s_n=\sum_{k=0}^{\infty}f(x_k)-f(x_{k+1})=f_0-f_1+f_1-f_2+\cdots+f_n-f_{n+1}=f_0-f_{n+1}$.

从而，

$$\lim_{n\to\infty}s_k=\lim_{n\to\infty}f_0-f_{n+1}=f_0-\lim_{n\to\infty}f_{n+1}.\tag{29}$$

表明 $\sum_{k=0}^{\infty}f(x_k)-f(x_{k+1})<\infty$，由级数的比较审敛法知 $\sum_{k=0}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2}<\infty$.

当 d_k 满足充分下降条件 $g_k^T d_k\leqslant -c\|g_k\|^2$，$(0<c<1)$，

有：

$$\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2}<\infty.\tag{30}$$

故引理 2.1 得证.

引理 2.2　若假设 H(1) 和 H(2) 成立，d_k 是搜索方向，步长 α_k 满足：

$$g_{k+1}^T d_k\geqslant\sigma g_k^T d_k,\ \sigma<1,\tag{31}$$

则有：

$$\alpha_k\geqslant\frac{1-\sigma}{L}\frac{|d_k^T g_k|}{\|d_k\|^2},\tag{32}$$

证：由（25）式和（31）式，结合柯西不等式有下式成立：

$$-(1-\sigma)g_k^T d_k\leqslant d_k^T(g_{k+1}-g_k)\leqslant L\alpha_k\|d_k\|^2,$$

因 d_k 是下降方向和 $\sigma<1$，显然（4.8）式成立；从强 wolfe 线搜索和（3.12）式知，步长 α_k 满足（32）式. 根据假设条件 H(1) 和 H(2)、（24）式得 $g_k^T d_k\neq 0$，$\forall k\geqslant 0$，因此，$\alpha_k=0$ 不满足（14）式，则存在常数 $\varpi>0$，

进一步有：

$$\alpha_k\geqslant\varpi,\ \forall k\geqslant 0.\tag{33}$$

以下给出 HMHSDY 算法在强 wolfe 线搜索下的全局收敛性.

定理 2.2　若假设条件 H(1) 和 H(2) 成立，则考虑算法 HMHSDY 产生的 $\{g_k\}$ 和 $\{d_k\}$，α_k 满足强 wolfe 线搜索，d_{k+1} 表示搜索方向，则有

$$\liminf_{k\to\infty}\|g_k\|=0.\tag{34}$$

证：反证法. 若结论不成立，则必存在常数 $\gamma>0$，使得

$$\|g_k\|\geqslant\gamma,\ \forall k\geqslant 1.\tag{35}$$

联立强 wolfe 线搜索第 2 式和（18）式，

则有：

$$d_k^T y_k=d_k^T g_{k+1}-d_k^T g_k\geqslant-(1-\sigma)g_k^T d_k\geqslant c(1-\sigma)\|g_k\|^2,\tag{36}$$

由假设 H(2) 知：

$$\|y_k\|=\|g_{k+1}-g_k\|\leqslant L\|x_{k+1}-x_k\|\leqslant L\|s_k\|\leqslant LD,\tag{37}$$

$$D = \max\{\|x-y\|,\ x,\ y \in \boldsymbol{N}\},$$

由前面的式子有：

$$|\beta_k^{HMHSDY}| = |(1-\theta_k)\beta_k^{MHS} + \theta_k\beta_k^{DY}|,$$

则有：

$$|\beta_k| \leqslant |\beta_k^{MHS}| + |\beta_k^{DY}| \leqslant \left|\frac{g_{k+1}^T\left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|}g_k\right)}{d_k^T y_k}\right| + \frac{\|g_{k+1}\|^2}{|d_k^T y_k|} \leqslant$$

$$\frac{|g_{k+1}^T(g_{k+1}-g_k)|}{|d_k^T y_k|} + \frac{\|g_{k+1}\|^2}{|d_k^T y_k|} \leqslant \frac{|g_{k+1}^T y_k|}{|d_k^T y_k|} + \frac{\|g_{k+1}\|^2}{|d_k^T y_k|} \leqslant$$

$$\frac{\|g_{k+1}\| \cdot \|y_k\|}{|d_k^T y_k|} + \frac{\|g_{k+1}\|^2}{|d_k^T y_k|} \leqslant \frac{\eta LD}{c(1-\sigma)\|g_k\|^2} + \frac{\eta^2}{c(1-\sigma)\|g_k\|^2} = \frac{\eta LD + \eta^2}{c(1-\sigma)\|g_k\|^2} = M. \tag{38}$$

当 θ_k 由步骤 5 定义，$\theta_k \notin (0,1)$ 时，（38）不等式成立，

则有：

$$\|d_{k+1}\| \leqslant \|g_{k+1}\| + |\beta_k| \cdot \|d_k\|, \tag{39}$$

因，

$$s_k = \alpha_k d_k,\ d_k = \frac{s_k}{\alpha_k}, \tag{40}$$

将（39）式写成：

$$\|d_{k+1}\| = \|g_{k+1}\| + \frac{|\beta_k| \cdot \|s_k\|}{\alpha_k} \leqslant \eta + \frac{MD}{\varpi} = F,$$

则有：

$$\|d_{k+1}\| \leqslant F, \sum_{k \geqslant 0} \frac{1}{\|d_k\|^2} = +\infty. \tag{41}$$

另一方面，从（18）式、引理 3.1 和（35）式得：

$$c^2\gamma^4 \sum_{k \geqslant 0} \frac{1}{\|d_k\|^2} \leqslant \sum_{k \geqslant 0} \frac{c^2\|g_k\|^4}{\|d_k\|^2} \leqslant \sum_{k \geqslant 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{42}$$

与（41）式矛盾，故（34）成立，证毕.

3 结论

本文结合 MHS 方法和 DY 方法做凸组合，得到新的共轭系数 β_k 和凸组合系数 θ_k，公式为 $\beta_k^{HXF} = (1-\theta_k)\beta_k^{MHS} + \theta_k\beta_k^{DY}$，使得搜索方向 d_k 不依赖于任何线搜索都充分下降，从而得到一种新的求解大规模无约束优化问题的混合共轭梯度法，并在强 wolfe 线搜索下，证明新算法是全局收敛的.

[参考文献]

[1] HESTENES M R，STIEFEL E. Method of conjugate gradient for solving linear equations [J]. Journal of Research National Bureau of Standards，1952，49：409-436

[2] 戴彧虹，袁亚湘. 非线性共轭梯度法 [M]. 上海：上海科技出版社，1999：1-3.

[3] 王静，乌彩英. 三项共轭梯度法在信号恢复问题中的应用 [J]. 内蒙古大学学报（自然科学版），2021，52（1）：27-32.

[4] 闵涛，寇婷，邹学文. 图像处理中几种数学方法的探讨 [J]. 计算机工程与应用，2006（12）：30-33.

[5] 沈海荣，刘超，宫宁生. 基于神经网络控制的共轭梯度法 [J]. 南京工业大学学报（自然科学版），2006（6）：91-94.

[6] 董永强. 基于近似共轭梯度法链路价格调整的速率控制算法 [J]. 解放军理工大学学报（自然科学版），2008，9（5）：475-478.

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

[8] 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.
[9] POLYAK B T. The conjugate gradient method in extremal problems [J]. USSR Computational Mathematics and Mathematical Physics，1969，9（4）：94-112.
[10] HESTENES M R，STIEFEL E. Method of conjugate gradient for solving linear equations [J]. Journal of Research National Bureau of Standards，1952，49：409-436
[11] LIU Y，STOREY C. Efficient generalized conjugate gradient algorithms，part 1：Theory [J]. Journal of Optimization Theory and Applications，1991，69（1）：129-137.
[12] WEI Z，YAO S，LIU L. The convergence properties of some new conjugate gradient methods [J]. Applied Mathematics and Computation，2006，183（2）：1341-1350.
[13] YAO S W，WEI Z X，HUANG H. A note about WYL's conjugate gradient method and its applications [J]. Applied Mathematics and Computation，2007，191（2）：381-388.
[14] 李丹丹，王松华. 基于凸组合技术的加速 FR 型共轭梯度算法 [J]. 广西科学，2021，28（2）：160-166.
[15] 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.
[16] 曾乐雅，许华，王天睿. 改进的凸组合最小均方算法 [J]. 北京邮电大学学报，2016，39（4）：114-117.
[17] ANDREI N. Another nonlinear conjugate gradient algorithm for unconstrained optimization [J]. Optimization Methods and Software，2008，24：89-104.
[18] DJORDJEVIC S S. New hybrid conjugate gradient method as a convex combination of LS and CD methods [J]. Filomat，2019，31：1813-1825.
[19] DJORDJEVIC S S. New hybrid conjugate gradient method as a convex combination of HS and FR conjugate gradient methods [J]. Journal of Applied Mathematics and Computation，2018，2：366-378.
[20] LIU J K，LI S J. New hybrid conjugate gradient method for unconstrained optimization [J]. Applied Mathematics and Computation，2014，245：36-43
[21] DELLADJI S，BELLOUFI M，BADREDDINE B，New hybrid conjugate gradient method as a convex combination of FR and BA methods [J]. Journal of Information and Optimization Sciences，2020，42（16）：1-12.

A Hybrid Conjugate Gradient Method as a New Convex Combination

CHEN Xiufang

(*School of Mathematics and Physics*, *West Yunnan University*, *Lincang*, *Yunnan*, *China* 677000)

Abstract: In this paper, a new hybrid conjugate gradient method for constrained optimization problems is presented. This method includes the convex combination of MHS method and Dai Yuan method in reference [13] to obtain new conjugate parameters β_k^{HXF}, and the search direction generated by the conjugate gradient method in the form of convex combination meets the conjugate conditions. It is further proved that the method does not rely on any line search and meets the sufficient descent property. Finally, under the general assumptions, strong wolfe line search is used to prove that the new algorithm framework has global convergence.

Key words: convex combination; hybrid conjugate gradient method; global convergence; the sufficient descent

（责任编辑：陈伟超）