第 37 卷第 1 期 重庆工商大学学报(自然科学版) 2020 年 2 月
Vol. 37 No. 1 J Chongqing Technol & Business Univ(Nat Sci Ed) Feb. 2020

doi:10. 16055/j. issn. 1672-058X. 2020. 0001. 008

一类具有充分下降性的 DL 共轭梯度法

谢 丽

(重庆师范大学 数学科学学院,重庆 401331)

摘 要:针对无约束优化问题,利用两项共轭梯度法(DL 方法)去逼近改进的 HS 三项共轭梯度法,提出了改进的 DL 共轭梯度法即 MDL 共轭梯度法. 该方法相对于 DL 方法具有一个更好的性质,即该共轭梯度法的搜索方向不依赖任何线搜索就可满足充分下降条件,理论上证明了该方法在 Wolfe 线搜索条件下对一般函数具有全局收敛性.

关键词:共轭梯度法;Wolfe 线搜索;充分下降性;全局收敛性

中图分类号:O224 **文献标志码**:A **文章编号**:1672-058X(2020)01-0049-05

0 引 言

考虑以下的无约束优化问题:

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

其中,$f:\boldsymbol{R}^n\to\mathbf{R}$ 连续可微.

共轭梯度法是求解式(1)这类无约束优化问题的重要方法之一,其迭代格式如下:

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

其中 α_k 为通过某种线搜索获得的步长,$\boldsymbol{d}_k$ 为搜索方向,一般迭代格式如下:

$$\boldsymbol{d}_k=\begin{cases}-\boldsymbol{g}_k,k=0\\ -\boldsymbol{g}_k+\beta_k\boldsymbol{d}_{k-1},k\geqslant 1\end{cases} \tag{3}$$

其中 $\boldsymbol{g}_k=\nabla f(x_k)$,$\beta_k\in\mathbf{R}$ 为共轭参数,不同的参数 β_k 对应不同的共轭梯度法. 其中经典的共轭梯度方法有 FR[1] 方法,HS[2] 方法,PRP[3] 方法和 DY[4] 方法,它们的表达式分别为

$$\beta_k^{\mathrm{FR}}=\frac{\|\boldsymbol{g}_k\|^2}{\|\boldsymbol{g}_{k-1}\|^2},\beta_k^{\mathrm{HS}}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{T}\boldsymbol{y}_{k-1}}$$

$$\beta_k^{\mathrm{PRP}}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2},\beta_k^{\mathrm{DY}}=\frac{\|\boldsymbol{g}_k\|^2}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}$$

$\|.\|$表示欧式范数.

2006 年,Zhang 等[5] 提出了一类三项共轭梯度法——ZZL 方法,该方法有一个优点,即满足充分下降条件:$\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k=-\|\boldsymbol{g}_k\|^2$, $\forall k\geqslant 0$,其搜索方向的迭代格式为

$$\boldsymbol{d}_k^{\mathrm{ZZL}}=-\boldsymbol{g}_k+\frac{\boldsymbol{g}_k^{\mathrm{T}}y_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}\boldsymbol{d}_{k-1}-\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}\boldsymbol{y}_{k-1}$$

2015 年,基于文献[5-6],Saman 等[7] 利用两项共轭梯度法 DL 方法去逼近三项共轭梯度法即 ZZL 方法,通过求解极小化问题 $\min\limits_{k\geqslant 0}\|\boldsymbol{d}_k^{\mathrm{DL}}-\boldsymbol{d}_k^{\mathrm{ZZL}}\|^2$,得到了改进的共轭梯度法,该方法的共轭参数形式如下:

$$\beta_k^{\mathrm{NDL}}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}-\frac{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{y}_k}{\|\boldsymbol{s}_k\|^2}\frac{\boldsymbol{g}_k^{\mathrm{T}}s_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}$$

2018 年,Li[8] 基于无记忆 BFGS 方法提出了一类修正的 HS 三项共轭梯度法,其共轭参数如下:

$$\beta_k^{\mathrm{JHS}}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}-\frac{\|\boldsymbol{y}_{k-1}\|^2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{(\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1})^2}+m_k\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}$$

其中 $0\leqslant m_k\leqslant\overline{m}<1$ 且 $\overline{m}=0.3$,同时,

$$m_k=\min\left\{0.3,\max\left\{0,1-\frac{\boldsymbol{y}_{k-1}^{\mathrm{T}}\boldsymbol{s}_{k-1}}{\|\boldsymbol{y}_{k-1}\|^2}\right\}\right\}$$

收稿日期:2019-05-30;**修回日期**:2019-07-11.

作者简介:谢丽(1994—),女,四川广安人,硕士研究生,从事最优化理论与算法研究.

1 一类新的共轭梯度法

本文受文献[7-8]的启发,利用两项共轭梯度法 DL 方法去逼近文献[8]提出的三项共轭梯度法,提出了一类满足充分下降条件的 DL 方法,其共轭参数形式如下:

$$\beta_k^{\mathrm{MDL}}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}-(1-m_k)\frac{\|\boldsymbol{y}_{k-1}\|^2}{\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}} \tag{4}$$

其中:

$m_k=\min\left\{0.3,\max\left\{0,1-\frac{\boldsymbol{y}_{k-1}^{\mathrm{T}}\boldsymbol{s}_{k-1}}{\|\boldsymbol{y}_{k-1}\|^2}\right\}\right\}$,且$\bar{m}=0.3$. 以上方法可以看作为改进的 DL 方法,其中参数 $t=(1-m_k)\frac{\|\boldsymbol{y}_{k-1}\|^2}{\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}$,将由式(2)式(3)式(4)组成的方法称作 MDL 方法. 然而在精确线搜索下,MDL 方法等价于 PRP 方法. 根据 Powell 的反例知,MDL 方法对一般函数不一定收敛,类似 DK+方法的截断思想,对 MDL 方法采取截断方式得到了 MDL+方法,其共轭参数形式如下:

$$\beta_k=\beta_k^{\mathrm{MDL+}}=\max\left\{\beta_k^{\mathrm{MDL}},\eta\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{d}_{k-1}\|^2}\right\} \tag{5}$$

其中 $0\leqslant\eta<1$,将由式(2)式(3)式(5)组成的方法称为 MDL+方法.

在收敛分析中,充分下降条件起着十分重要的作用,即 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k\leqslant-c\|\boldsymbol{g}_k\|^2$ 成立. 其中 c 为常数且 $c>0$,步长 α_k 一般要求满足 Wolfe 线搜索条件:

$$f(\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k)-f(\boldsymbol{x}_k)\leqslant\delta\alpha_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \tag{6}$$

$$\boldsymbol{g}(\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k\geqslant\sigma\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \tag{7}$$

其中 $0<\delta<\sigma<1$.

2 收敛性分析

以下假设 $\boldsymbol{g}_k\neq0$,对所有 k 成立. 为了证明 MDL 方法的全局收敛性,本文做如下两个假设:

(A)目标函数 f 在以下水平集:$\Omega=\{\boldsymbol{x}\in\boldsymbol{R}^n:f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_0)\}$ 上有下界,其中 $\boldsymbol{x}_0\in\boldsymbol{R}^n$ 为算法初始点,即存在一个常数 $B>0$,使得

$$\|\boldsymbol{x}\|\leqslant B,\ \forall\boldsymbol{x}\in\Omega \tag{8}$$

(B)目标函数 f 在 Ω 的某个领域 N 是连续可微的,且梯度是 Lipschitz 连续的,即存在常数 $L>0$,使得

$$\|\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{g}(\bar{\boldsymbol{x}})\|\leqslant L\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|,\ \forall\boldsymbol{x},\bar{\boldsymbol{x}}\in N \tag{9}$$

由对目标函数 f 的假设知,存在一个常数$\bar{\gamma}\geqslant0$,使得

$$\|\boldsymbol{g}(\boldsymbol{x})\|\leqslant\bar{\gamma},\ \forall\boldsymbol{x}\in\Omega \tag{10}$$

在上述假设下,给出如下的引理 1,该引理由 Dai-Kou 证得,通常被称为 Zoutendijk 条件.

引理 1 假设(A)(B)成立,考虑具有式(2)、式(3)形式的迭代方法,其中 $\boldsymbol{d}_k$ 是一个下降方向,α_k 满足 Wolfe 条件,则有

$$\sum_{k\geqslant1}\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2}<\infty \tag{11}$$

引理 2 假设(A)(B)成立,考虑具有式(2)式(3)形式的迭代方法,其中 $\boldsymbol{d}_k$ 是一个下降方向,α_k 满足 Wolfe 条件,若

$$\sum_{k\geqslant1}\frac{1}{\|\boldsymbol{d}_k\|^2}=\infty$$

则有

$$\lim_{k\to\infty}\inf\|\boldsymbol{g}_k\|=0 \tag{12}$$

引理 3 假设(A)(B)成立,考虑形如式(2)式(3)式(4)的 MDL 方法,以及式(2)式(3)式(5)的 MDL+方法,若 $\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}\neq0$,则方法满足充分下降性,即 $\forall k$,有

$$\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k\leqslant\bar{c}\|\boldsymbol{g}_k\|^2 \tag{13}$$

证明 用归纳法证明:

当 $k=0$,$0\leqslant m_k\leqslant\bar{m}<1$ 时,

$$\boldsymbol{g}_o^{\mathrm{T}}\boldsymbol{d}_0=-\|\boldsymbol{g}_0\|^2\leqslant-\left(1-\frac{(1+\bar{m})^2}{4}\right)\|\boldsymbol{g}_0\|^2$$

成立。

假设当 $n=k-1$,$0\leqslant m_k\leqslant\bar{m}<1$ 时,有 $\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}\leqslant-\left(1-\frac{(1+\bar{m})^2}{4}\right)\|\boldsymbol{g}_{k-1}\|^2$ 成立,则

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

$$\frac{\|\boldsymbol{y}_{k-1}\|^2(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1})^2}{(\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1})^2}=-\|\boldsymbol{g}_k\|^2+\frac{(1+\overline{m}_k)^2}{4}\|\boldsymbol{g}_k\|^2\leqslant -\left(1-\frac{(1+\overline{m})^2}{4}\right)\|\boldsymbol{g}_k\|^2$$

其中 $0\leqslant m_k\leqslant\overline{m}<1$,则充分下降条件式(13)成立.

对于 MDL+方法的充分下降性证明,只需考虑如下情况,当 $k\geqslant2$ 时,有

$$\beta_k^{\mathrm{MDL+}}=\eta\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{d}_{k-1}\|^2}$$

则有

$$\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k=-\|\boldsymbol{g}_k\|^2+\eta\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1})^2}{\|\boldsymbol{d}_{k-1}\|^2}\leqslant-(1-\eta)\|\boldsymbol{g}_k\|^2 \tag{14}$$

将上述两种情况合并,有

$$\bar{c}=\min\left\{1-\frac{(1+\overline{m})^2}{4},1-\eta\right\}$$

证毕。

下证 MDL 方法采用 Wolfe 线搜索时,对一致凸函数是强收敛的.

定理 1 假设 f 是一致凸函数且假设(A)(B)成立,考虑 MDL 方法. 如果 α_k 是由 Wolfe 线搜索计算得到,且 $\sigma\in[0,1]$,则

$$\lim_{k\to+\infty}\|\boldsymbol{g}_k\|=0 \tag{15}$$

证明 由 Lipschitze 条件以及 f 为一致凸函数,有

$$\|\boldsymbol{y}_{k-1}\|\leqslant L\|\boldsymbol{s}_{k-1}\|,\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}\geqslant\mu\|\boldsymbol{d}_{k-1}\|\|\boldsymbol{s}_{k-1}\|$$

由式(8)和式(9),有

$$\left|\frac{\|\boldsymbol{y}_{k-1}\|^2}{\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right|\leqslant\frac{L^2}{\mu}$$

使用 Wolfe 条件时,有

$$\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}\geqslant(\sigma-1)\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}>0$$

式(4)中的 β_k^{MDL} 可写成

$$\beta_k^{\mathrm{MDL}}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}-t_k\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}$$

其中参数

$$t_k=(1-m_k)\frac{\|\boldsymbol{g}_{k-1}\|^2}{\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}},0\leqslant m_k\leqslant\overline{m}<1$$

由式(10)和 $0\leqslant m_k\leqslant\overline{m}<1$ 知:

$$|t_k|=\left|(1-m_k)\frac{\|\boldsymbol{g}_{k-1}\|^2}{\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right|<\frac{L^2}{\mu}$$

则有

$$\|\boldsymbol{d}_k\|\leqslant\|\boldsymbol{g}_k\|+\left|\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}-t_k\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right|\|\boldsymbol{d}_{k-1}\|\leqslant \left(1+\frac{L\|\boldsymbol{s}_{k-1}\|\|\boldsymbol{d}_{k-1}\|}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}+|t_k|\frac{\|\boldsymbol{d}_{k-1}\|\|\boldsymbol{s}_{k-1}\|}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right)\|\boldsymbol{g}_k\|\leqslant \left(1+\frac{L}{\mu}+|t_k|\frac{1}{\mu}\right)|\boldsymbol{g}_k| \tag{16}$$

由充分下降条件式(13)和引理 1 的 Zoutentijk 条件知

$$\sum_{k\geqslant1}\frac{\|\boldsymbol{g}_k\|^4}{\|\boldsymbol{d}_k\|^2}<\infty$$

由式(16)有

$$\sum_{k\geqslant1}\|\boldsymbol{g}_k\|^2<\infty \tag{17}$$

故式(15)成立.

下证 MDL+方法对一般函数的全局收敛性,首先需证明两个重要引理.

引理 4 假设(A)(B)成立,考虑式(2)(3)(5)的 MDL+共轭梯度法,其中 $\boldsymbol{d}_k$ 是一个下降方向,α_k 由 Wolfe 线搜索计算得到,如果存在一个常数 $\gamma>0$,使得

$$\|\boldsymbol{g}_k\|\geqslant\gamma,\ \forall k\geqslant1 \tag{18}$$

则 $\boldsymbol{d}_k\neq0$ 且

$$\sum_{k\geqslant1}\|\boldsymbol{\mu}_k-\boldsymbol{\mu}_{k-1}\|^2<\infty$$

其中 $\boldsymbol{\mu}_k=\dfrac{\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}$.

证明 首先,注意到 $\boldsymbol{d}_k\neq0$,否则式(13)不成立,因此 μ_k 有意义. 此外,由式(18)和引理 2,有

$$\sum_{k\geqslant1}\frac{1}{\|\boldsymbol{d}_k\|^2}<\infty$$

否则式(12)成立,与式(18)矛盾. 现在将 $\beta_k^{\mathrm{MDL+}}$ 分为如下两部分:

$$\beta_k^{(1)}=\max\left\{\beta_k^{\mathrm{MDL}}-\eta\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{d}_{k-1}\|^2},0\right\}$$

$$\beta_k^{(2)}=\eta\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{d}_{k-1}\|^2} \tag{19}$$

假设

$$\boldsymbol{w}_k=\frac{\boldsymbol{v}_k}{\|\boldsymbol{d}_k\|},\boldsymbol{\delta}_k=\beta_k^{(1)}\frac{\|\boldsymbol{d}_{k-1}\|}{\|\boldsymbol{d}_k\|} \tag{20}$$

其中 $\boldsymbol{v}_k=-\boldsymbol{g}_k+\beta_k^{(2)}\boldsymbol{d}_{k-1}$.

当 $k\geqslant2$ 时,由式(3),有

$$\boldsymbol{\mu}_k=\boldsymbol{\gamma}_k+\boldsymbol{\delta}_k\boldsymbol{\mu}_{k-1} \tag{21}$$

使用等式 $\|\boldsymbol{\mu}_k\|=\|\boldsymbol{\mu}_{k-1}\|=1$ 和方程(21),有

$$\|\boldsymbol{\gamma}_k\|=\|\boldsymbol{\mu}_k-\boldsymbol{\delta}_k\boldsymbol{\mu}_{k-1}\|=\|\boldsymbol{\delta}_k\boldsymbol{\mu}_k-\boldsymbol{\mu}_{k-1}\| \tag{22}$$

由 $\boldsymbol{\delta}_k\geqslant 0$,再根据式(22),得到:

$$\|\boldsymbol{\mu}_k-\boldsymbol{\mu}_{k-1}\|\leqslant\|(1+\boldsymbol{\delta}_k)\boldsymbol{\mu}_k-(1+\boldsymbol{\delta}_k)\boldsymbol{\mu}_{k-1}\|=$$
$$\|\boldsymbol{\delta}_k\boldsymbol{\mu}_k-\boldsymbol{\mu}_{k-1}\|\leqslant\|\boldsymbol{\mu}_k-\boldsymbol{\delta}_k\mu_{k-1}\|+\|\boldsymbol{\delta}_k\boldsymbol{\mu}_k-\boldsymbol{\mu}_{k-1}\|=2\|\boldsymbol{w}_k\| \tag{23}$$

由式(19)得:

$$\|\boldsymbol{v}_k\|\leqslant\|\boldsymbol{g}_k\|+\|\beta_k^{(2)}\|\|\boldsymbol{d}_{k-1}\|\leqslant(1+\eta)\|\boldsymbol{g}_k\| \tag{24}$$

由式(20)(23)(24)得:

$$\|\boldsymbol{\mu}_k-\boldsymbol{\mu}_{k-1}\|\leqslant 2\|\boldsymbol{\mu}_k\|\leqslant 2(1+\eta)\frac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{d}_k\|} \tag{25}$$

通过式(11)(13)(18),有

$$\sum_{k\geqslant 1}\frac{\|\boldsymbol{g}_k\|^2}{\|\boldsymbol{d}_k\|^2}\leqslant\frac{1}{\gamma^2}\sum_{k\geqslant 1}\frac{\|\boldsymbol{g}_k\|^4}{\|\boldsymbol{d}_k\|^2}\leqslant\frac{1}{\gamma^2\bar{c}^2}\sum_{k\geqslant 1}\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2}<+\infty$$

最后由式(25)可知:

$$\sum_{k\geqslant 1}\|\boldsymbol{\mu}_k-\boldsymbol{\mu}_{k-1}\|^2\leqslant 4(1+\eta)\sum_{k\geqslant 1}\frac{\|\boldsymbol{g}_k\|^2}{\|\boldsymbol{d}_k\|^2}<\infty$$

证毕.

定义正整数集为 N^*,常数 $\lambda>0$ 和一个正整数 Δ,记:$K_{k,\Delta}^{\lambda}:=\{i\in\boldsymbol{N}^*:k\leqslant i\leqslant k+\Delta-1,\|\boldsymbol{s}_{i-1}\|>\lambda\}$,令 $|K_{k,\Delta}^{\lambda}|$ 为 $K_{k,\Delta}^{\lambda}$ 中的元素个数.

引理 5 假设(A)(B)成立,考虑形如式(2)(3)(5)的 MDL+方法,步长 α_k 通过 Wolfe 线搜索式(6)和(7)计算得到,如果式(17)成立,则存在常数 $\lambda>0$,使得 $\forall\Delta\in\boldsymbol{N}^*$ 和任意指标集 k_0,存在 $k\geqslant k_0$,$|K_{k,\Delta}^{\lambda}|>\dfrac{\Delta}{2}$成立。

证明过程类似于文献[10]中引理 3.5 的证明.

下面是 MDL+方法的全局收敛性定理.

定理 2 假设(A)(B)成立,考虑形如式(2)(3)(5)的 MDL+共轭梯度法,线搜索 α_k 满足 Wolfe 线搜索,则有

$$\lim_{k\to\infty}\inf\|\boldsymbol{g}_k\|=0$$

证明 证明这个定理利用反证法. 如果$\lim\limits_{k\to\infty}\inf\|\boldsymbol{g}_k\|>0$,则式(18)一定成立,且引理 4 和引理 5 的条件成立. 定义$\boldsymbol{\mu}_i=\dfrac{\boldsymbol{d}_i}{\|\boldsymbol{d}_i\|}$,对任意指标 l,k 且 $l\geqslant k$,有

$$\boldsymbol{x}_l-\boldsymbol{x}_{k-1}=\sum_{i=k}^{l}\boldsymbol{x}_i-x_{i-1}=\sum_{i=k}^{l}\boldsymbol{\alpha}_{i-1}\boldsymbol{d}_{i-1}=\sum_{i=k}^{l}\boldsymbol{\mu}_{i-1}\|\boldsymbol{s}_{i-1}\|=$$
$$\sum_{i=k}^{l}\|\boldsymbol{s}_{i-1}\|\boldsymbol{\mu}_{k-1}+\sum_{i=k}^{l}\|\boldsymbol{s}_{i-1}\|(\boldsymbol{\mu}_{i-1}-\boldsymbol{\mu}_{k-1}) \tag{26}$$

由式(26),$\|\boldsymbol{\mu}_i\|=1$ 和式(8),有

$$\sum_{i=k}^{l}\|\boldsymbol{s}_{i-1}\|\leqslant\|\boldsymbol{x}_l-\boldsymbol{x}_{k-1}\|+\sum_{i=k}^{l}\|\boldsymbol{s}_{i-1}\|\|\boldsymbol{\mu}_{i-1}-\boldsymbol{\mu}_{k-1}\|\leqslant$$
$$2B+\sum_{i=k}^{l}\|\boldsymbol{s}_{i-1}\|\|\boldsymbol{\mu}_{i-1}-\boldsymbol{\mu}_{k-1}\| \tag{27}$$

让引理 5 中的 $\lambda>0$,且定义 $\Delta:=\left|\dfrac{8B}{\lambda}\right|$是不小于$\dfrac{8B}{\lambda}$的最小整数. 根据引理 4,可以找到这样的一个指标 $k_0\geqslant 1$,式(28)成立.

$$\sum_{l\geqslant k_0}\|\boldsymbol{\mu}_{i-1}-\boldsymbol{\mu}_{k-1}\|^2\leqslant\frac{1}{4\Delta} \tag{28}$$

对于这样的 Δ 和 k_0,引理 5 给出了一个指标 $k\geqslant k_0$,有

$$|K_{k,\Delta}^{\lambda}|>\frac{\Delta}{2} \tag{29}$$

对于任何指数 $i\in[k,k+\Delta-1]$,由 Cauchy-Schwartz 不等式和式(28),有

$$\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_{k-1}\|\leqslant\sum_{j=k}^{i}\|\boldsymbol{\mu}_j-\boldsymbol{\mu}_{j-1}\|\leqslant$$
$$(i-k+1)^{\frac{1}{2}}\Big(\sum_{j=k}^{i}\|\boldsymbol{\mu}_j-\boldsymbol{\mu}_{j-1}\|^2\Big)^{\frac{1}{2}}\leqslant$$
$$\Delta^{\frac{1}{2}}\left(\frac{1}{4\Delta}\right)^{\frac{1}{2}}=\frac{1}{2} \tag{30}$$

从这些关系式(29)和式(30)中,取式(27)中的$l=k+\Delta-1$,得到:

$$2B\geqslant\frac{1}{2}\sum_{i=k}^{k+\Delta-1}\|\boldsymbol{s}_{i-1}\|>\frac{\lambda}{2}|K_{k,\Delta}^{\lambda}|>\frac{\lambda\Delta}{4}$$

因此 $\Delta<\dfrac{8B}{\lambda}$与 Δ 的定义矛盾,证明完成.

3 结束语

本文针对大规模无约束最优化问题,在 DL 共轭梯度方法的基础上提出了改进的 DL 方法即 MDL 共轭梯度法. MDL 共轭梯度法具有一个很好的性质,即不依赖于任何线搜索就可满足充分下降性,同时在 Wolfe 线搜索条件下对一般函数具有全局收敛性,理论上相对于 DL 方法有了很大的进步. 本文主要思想是利用两项共轭梯度法去逼近三项共轭梯度法来获取 DL 方法里面参数 t 的最优选择. 文献[11]是通过改进 DL 方法里面的$\dfrac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}$,得到 NVLS*DL 方法,与本文有着本质的区别.

参考文献(References)：

[1] FLETCHER R, REEVES C. Function Minimization by Conjugate Gradients[J]. Computation, 1964, 7(2): 149—154

[2] HESTENES M, STIEFEL E. Methods of Conjugate Gradients for Solving Linear Systems[M]. Washington DC: NBS, 1952

[3] POLA E, RIBIERE G. Note Sur La Convergence De Methodes De Directions Conjugées[J]. Rev Française Informat Recherche Operationelle, 1969(3): 35—43

[4] LIU Y, STOREY C. Efficient Generalized Conjugate Gradient Algorithms(part 1)[J]. Journal of Optimization Theory and Applications, 1991, 69(1):129—137

[5] ZHANG L, ZHOU W, LI D H. A Descent Modified Polak - Ribière - Polyak Conjugate Gradient Method and Its Global Convergence[J]. IMA Journal of Numerical Analysis, 2006, 26(4): 629—640

[6] DAI Y H, LIAO L Z. New Conjugacy Conditions and Related Nonlinear Conjugate Gradient Methods[J]. Applied Mathematics and Optimization, 2001, 43(1): 87—101

[7] BABAIE-KAFAKI S, GHANBARI R. Two Optimal Dai Liao Conjugate Gradient Methods[J]. Optimization, 2015, 64(11): 2277—2287

[8] LI M. A Family of Three - term Nonlinear Conjugate Gradient Methods Close to the Memoryless BFGS Method[J]. Optimization Letters, 2018, 12(8): 1911—1927

[9] ZOUTENDIJK G. Nonlinear Programming, Computational Methods[J]. Integer and Nonlinear Programming, 1970, 143(1):37—86

[10] GILBERT J C, NOCEDAL J. Global Convergence Properties of Conjugate Gradient Methods for Optimization[J]. SIAM Journal on Optimization, 1992,2(1):21—42

[11] 张莉林. 一类具有充分下降性修正的 DL 共轭梯度法[J]. 重庆工商大学学报(自然科学版),2017,34(6):38—41

ZHANG L L. A Class of DL Conjugate Gradient Method with Sufficient Descent Correction[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2017,34(6):38—41(in Chinese)

A Class of DL Conjugate Gradient Method with Sufficient Descent

XIE Li

(College of Mathematical Science, Chongqing Normal University, Chongqing 401331,China)

Abstract: For unconstrained optimization problems, the improved HS three-term conjugate gradient method is approximated by the two-term conjugate gradient method DL method, and the improved DL conjugate gradient method MDL conjugate gradient method is proposed. Compared with DL conjugate gradient method, this method has a better property, that is, the search direction of the conjugate gradient method can satisfy sufficient descent condition without relying on any line search. It is theoretically proved that the method has global convergence for general functions under Wolfe line search condition.

Key words: conjugate gradient method; Wolfe line search; sufficient descent property; global convergence

责任编辑：李翠薇

引用本文/Cite this paper:

谢丽. 一类具有充分下降性的 DL 共轭梯度法[J]. 重庆工商大学学报(自然科学版),2020,37(1):49—53

XIE L. A Class of DL Conjugate Gradient Method with Sufficient Descent[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2020,37(1):49—53