doi:10.16055/j.issn.1672-058X.2021.0002.005

一类混合的 DL-WYL 共轭梯度法

李　月

(重庆师范大学 数学科学学院,重庆 401331)

摘　要:共轭梯度法因为其迭代简单和低存储等特点,在工程问题、金融模型等许多实际领域中得到广泛的应用;针对大规模无约束优化问题,提出了一类混合的 DL-WYL 共轭梯度法——LHSDL 方法,它可以看作是一类修正的 DL 共轭梯度法,即利用一个数值效果和理论结果均良好的 Wei-Yao-Liu 型共轭梯度法的共轭参数去修正 DL 共轭梯度法的第一项;它也可以看作是一类修正的 WYL 共轭梯度法,通过添加 DL 共轭梯度法的第二项,使该方法可能含有一些 Hessian 信息。LHSDL 方法相对于 DL 方法具有一个较好的性质,即在强 Wolfe 线搜索条件下具有充分下降性,并且理论证明了 LHSDL 方法对于一般函数具有全局收敛性;数值实验是在 CUTEr 集的一组无约束优化测试问题上进行的,由 Dolan 和 Moré 的性能曲线图表明:LHSDL 方法略优于 DK+方法和 MNVHS 方法。

关键词:共轭梯度法;强 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(2021)02-0028-07

0　引　言

考虑无约束优化问题:

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

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

共轭梯度法在求解问题式(1)中起着重要的作用,它的一般迭代格式如下:

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

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

其中 $\boldsymbol{g}_k=\nabla f(\boldsymbol{x}_k)$,$\beta_k$ 是共轭参数,α_k 是步长,本文采用强 Wolfe 条件式(4)(5)计算 α_k:

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

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

其中 $0<\delta<\sigma<1$,$\boldsymbol{d}_k$ 表示搜索方向。

β_k 的不同代表不同的方法,经典的共轭梯度法为 HS[1],PRP[2],DY[3]方法,其 β_k 分别为

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

其中‖·‖表示 Euclidean 范数,$\boldsymbol{y}_{k-1}=\boldsymbol{g}_k-\boldsymbol{g}_{k-1}$。

2001 年,Dai 和 Liao[4] 提出了一个新的共轭梯度法,其共轭参数 β_k^{DL} 为

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

则有:

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

其中常数 $t>0$,$\boldsymbol{s}_{k-1}=\boldsymbol{x}_k-\boldsymbol{x}_{k-1}$。

2006 年,Wei 等[5] 提出了一类 WYL 型共轭梯

收稿日期:2020-04-14;**修回日期**:2020-05-20.

作者简介:李月(1994—),女,重庆綦江人,硕士研究生,从事最优化算法与理论研究.

度法,共轭参数 β_k^{WYL} 如下:

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

2012 年,Dai[6] 提出了一类修正的 WYL 共轭梯度法,其共轭参数 β_k 为

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

其中 $\mu>2$。

2016 年,Du 等[7] 提出了一类 HS 方法的变型,简称为 NVHS 方法,β_k^{NVHS} 如下所示:

$$\beta_k^{NVHS}=\frac{\boldsymbol{g}_k^T\left(\boldsymbol{g}_k-\frac{|\boldsymbol{g}_k^T\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2}\boldsymbol{g}_{k-1}\right)}{\boldsymbol{d}_{k-1}^T\boldsymbol{y}_{k-1}}$$

2018 年,陈[8] 提出了一类修正的 NVHS 方法,简称为 MNVHS 方法,β_k^{MNVHS} 为

$$\beta_k^{MNVHS}=\frac{\boldsymbol{g}_k^T\left(\boldsymbol{g}_k-\frac{|\boldsymbol{g}_k^T\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2}\boldsymbol{g}_{k-1}\right)}{\max\{\mu|\boldsymbol{g}_k^T\boldsymbol{d}_{k-1}|,\boldsymbol{d}_{k-1}^T\boldsymbol{y}_{k-1}\}}$$

其中 $\mu>2$。

1　一类修正的 LHSDL 共轭梯度法

2015 年,马[9] 提出一类 DLVHS 共轭梯度方法,其共轭参数 β_k^{DLVHS} 为

$$\beta_k^{DLVHS}=\frac{\boldsymbol{g}_k^T\left(\boldsymbol{g}_k-\frac{|\boldsymbol{g}_k^T\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2}\boldsymbol{g}_{k-1}\right)}{\boldsymbol{d}_{k-1}^T\boldsymbol{y}_{k-1}}-t\frac{\boldsymbol{g}_k^T\boldsymbol{s}_{k-1}}{\boldsymbol{d}_{k-1}^T\boldsymbol{y}_{k-1}}$$

其中 $t>0$,更多类似方法请参见文献[10-12]。

2019 年,谢[13] 提出了一类新的 DL 共轭梯度法,简称为 JHSDL 方法,β_k^{JHSDL} 如下所示:

$$\beta_k^{JHSDL}=\frac{\|\boldsymbol{g}_k\|^2-\frac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}\boldsymbol{g}_k^T\boldsymbol{g}_{k-1}}{\max\{\mu|\boldsymbol{g}_k^T\boldsymbol{d}_{k-1}|,\boldsymbol{d}_{k-1}^T\boldsymbol{y}_{k-1}\}}-t\frac{\boldsymbol{g}_k^T\boldsymbol{s}_{k-1}}{\boldsymbol{d}_{k-1}^T\boldsymbol{y}_{k-1}}$$

其中 $\mu>2$。

受文献[6-13]的启发,本文提出一类混合的 DL-WYL 共轭梯度法,共轭参数 β_k^{LHSDL} 如下:

$$\beta_k^{LHSDL}=\frac{\|\boldsymbol{g}_k\|^2-\frac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}\boldsymbol{g}_k^T\boldsymbol{g}_{k-1}}{\boldsymbol{d}_{k-1}^T\boldsymbol{y}_{k-1}+\mu|\boldsymbol{g}_k^T\boldsymbol{d}_{k-1}|}-t\frac{\boldsymbol{g}_k^T\boldsymbol{s}_{k-1}}{\boldsymbol{d}_{k-1}^T\boldsymbol{y}_{k-1}}\tag{6}$$

其中参数 $\mu>2$,当 $\mu=0$ 时,式(6)变为文献[9]中的 β_k^{DLVHS} 公式。将式(2)(3)和式(6)组成的共轭梯度法称为 LHSDL 方法。

2　LHSDL 算法

初始步　给定初始值 $\boldsymbol{x}_0\in\boldsymbol{R}^n$,置 $k:=0$,$\boldsymbol{d}_0=-\boldsymbol{g}_0$,$\varepsilon>0$。如果 $\|\boldsymbol{g}_0\|\leqslant\varepsilon$,则停止;

第一步　计算步长 $\alpha_k>0$,使其满足式(4)和式(5)。

第二步　计算 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k d_k$,如果 $\|\boldsymbol{g}_{k+1}\|\leqslant\varepsilon$,则停止。

第三步　计算式(6)的 β_k^{LHSDL} 和式(3)的 $\boldsymbol{d}_k$。

第四步　置 $k=k+1$,转第一步。

3　收敛性

首先做如下两个基本假设:

假设 A　水平集 $L=\{\boldsymbol{x}\in\boldsymbol{R}^n:f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_0)\}$ 有界,其中 $\boldsymbol{x}_0$ 为算法的初始点。即存在一个常数 $B>0$,使得

$$\|\boldsymbol{x}\|\leqslant B,\ \forall\boldsymbol{x}\in L\tag{7}$$

假设 B　目标函数 $f(\boldsymbol{x})$ 在 L 的 N 邻域内连续可微,且梯度是 Lipschitz 连续的,即 $\forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{N}$,存在

$$\|\boldsymbol{g}(x)-\boldsymbol{g}(y)\|\leqslant L\|\boldsymbol{x}-\boldsymbol{y}\|\tag{8}$$

如果目标函数 $f(\boldsymbol{x})$ 满足假设 A,则存在常数 $\bar{\gamma}>0$,使得

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

下面证明由 LHSDL 方法产生的搜索方向 $\boldsymbol{d}_k$ 是充分下降的,即

$$\boldsymbol{g}_k^T\boldsymbol{d}_k\leqslant-c\|\boldsymbol{g}_k\|^2,\ \forall k\geqslant0,c>0$$

引理 1　如下不等式(10)成立

$$0\leqslant\boldsymbol{g}_k^T\left(\boldsymbol{g}_k-\frac{|\boldsymbol{g}_k^T\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2}\boldsymbol{g}_{k-1}\right)\leqslant2\|\boldsymbol{g}_k\|^2\tag{10}$$

$$0\leqslant\boldsymbol{g}_k^T\left(\boldsymbol{g}_k-\frac{|\boldsymbol{g}_k^T\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2}\boldsymbol{g}_{k-1}\right)\leqslant2\|\boldsymbol{g}_k\|\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|\tag{11}$$

证明　首先

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

由 Cauchy-Schwarz 不等式可得:

$$\|\boldsymbol{g}_k\|^2-\frac{|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|\,|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2}\geqslant\|\boldsymbol{g}_k\|^2-\frac{\|\boldsymbol{g}_k\|^2\|\boldsymbol{g}_{k-1}\|^2}{\|\boldsymbol{g}_{k-1}\|^2}=0$$

$$\|\boldsymbol{g}_k\|^2-\frac{|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|\,|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2}\leqslant\|\boldsymbol{g}_k\|^2+\frac{\|\boldsymbol{g}_k\|^2\|\boldsymbol{g}_{k-1}\|^2}{\|\boldsymbol{g}_{k-1}\|^2}=2\|\boldsymbol{g}_k\|^2$$

综上可知,式(10)成立,式(11)的详细证明请参见文献[7]。

引理 2 考虑形如式(2)(3)和式(6)的 LHSDL 方法,其中 α_k 由强 Wolfe 条件计算所得,且 $0<\sigma<\frac{1}{3}$,$\mu>2$,则对任意的 $k\geqslant1$,有

$$\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k\leqslant\frac{-1+3\sigma}{1-\sigma}\|\boldsymbol{g}_k\|^2 \tag{12}$$

证明 当 $k=0$ 时,$\boldsymbol{d}_0=-\boldsymbol{g}_0$,则

$$\boldsymbol{g}_0^{\mathrm{T}}\boldsymbol{d}_0=-\|\boldsymbol{g}_0\|^2\leqslant\frac{-1+3\sigma}{1-\sigma}\|\boldsymbol{g}_0\|^2$$

假设对 $k-1$,$\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}\leqslant\frac{-1+3\sigma}{1-\sigma}\|\boldsymbol{g}_{k-1}\|^2$ 成立,则

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

$$-\|\boldsymbol{g}_k\|^2+\frac{\boldsymbol{g}_k^{\mathrm{T}}\left(\boldsymbol{g}_k-\frac{|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}|}{\|\boldsymbol{g}_{k-1}\|^2}\boldsymbol{g}_{k-1}\right)}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}+\mu|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}|}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}-t\,\frac{\alpha_k(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1})^2}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}$$

由强 Wolfe 条件式(5)可得

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

意味着

$$t\,\frac{\alpha_k(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1})^2}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}>0$$

因此

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

$$-\|\boldsymbol{g}_k\|^2+\frac{2\|\boldsymbol{g}_k\|^2}{d_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}\leqslant\left(-1+2\,\frac{-\sigma g_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right)\|\boldsymbol{g}_k\|^2\leqslant$$

$$\left(-1+2\,\frac{-\sigma\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}}{(\sigma-1)\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}}\right)\|\boldsymbol{g}_k\|^2\leqslant\frac{-1+3\sigma}{1-\sigma}\|\boldsymbol{g}_k\|^2$$

令 $c=\frac{-1+3\sigma}{1-\sigma}$,因为 $0<\sigma<\frac{1}{3}$,故 $c<0$ 满足充分下降性。综上,结论得证。

性质 1[14] 考虑形如式(2)和式(3)的共轭梯度法,假设存在常数 γ,$\bar{\gamma}$有

$$0<\gamma\leqslant\|\boldsymbol{g}_k\|\leqslant\bar{\gamma},\ \forall k>0 \tag{13}$$

如果存在 $b>1$ 和 $\lambda>0$,对任意的 k 都有 $|\beta_k|\leqslant b$,以及 $\|\boldsymbol{s}_{k-1}\|\leqslant\lambda\Rightarrow|\beta_k|\leqslant\frac{1}{2b}$成立,则称该共轭梯度法具有性质 1。

引理 3 假设 A,B 成立,考虑形如式(2)(3)和式(6)的 LHSDL 方法,α_k 由强 Wolfe 条件计算所得,则 LHSDL 方法具有性质 1。

证明 由式(5)可得,

$$\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}\geqslant(\sigma-1)\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}\geqslant(1-\sigma)c\gamma^2$$

其中 $c=\frac{-1+3\sigma}{1-\sigma}$,由引理 1,式(7)和式(8)可得

$$|\beta_k^{\mathrm{LHSDL}}|\leqslant\frac{2\|\boldsymbol{g}_k\|\|\boldsymbol{g}_k-\boldsymbol{g}_{k-1}\|}{|\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}|+\mu|\boldsymbol{g}_k^{\mathrm{T}}\mathrm{d}_{k-1}|}+t\,\frac{\|\boldsymbol{g}_k\|\|\boldsymbol{s}_{k-1}\|}{|\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}|}\leqslant$$

$$\frac{\|\boldsymbol{g}_k\|(2L\|\boldsymbol{s}_{k-1}\|)}{c(1-\sigma)\|\boldsymbol{g}_{k-1}\|^2}+t\,\frac{\|\boldsymbol{g}_k\|\|\boldsymbol{s}_{k-1}\|}{c(1-\sigma)\|\boldsymbol{g}_{k-1}\|^2}\leqslant\frac{2(2L+t)\bar{\gamma}B}{\gamma^2c(1-\sigma)}$$

设 $b:=\frac{2(2L+t)\bar{\gamma}B}{(1-\sigma)c\gamma^2}$,因此 $|\beta_k^{\mathrm{LHSDL}}|\leqslant b$。

又设 $\lambda:=\frac{(1-\sigma)c\gamma^2}{2b(2L+t)\bar{\gamma}}$,则 $|\beta_k^{\mathrm{LHSDL}}|\leqslant\frac{\|\boldsymbol{s}_{k-1}\|(2L+t)\bar{\gamma}}{(1-\sigma)c\gamma^2}<\frac{1}{2b}$,因此 LHSDL 方法具有性质 1。

对于非线性共轭梯度法,Dai 等[15]提出了以下一般性结论。

引理 4 假设 A,B 成立,考虑形如式(2)(3)的共轭梯度法,其中 $\boldsymbol{d}_k$ 是一个下降方向,α_k 由强 Wolfe 条件计算所得。若

$$\sum_{k\geqslant1}\frac{1}{\|\boldsymbol{d}_k\|^2}=\infty \tag{14}$$

则有

$$\liminf_{x\to\infty}\|\boldsymbol{g}_k\|=0$$

引理 5 假设 A,B 成立,考虑形如式(2)(3)和式(6)的 LHSDL 方法,α_k 由强 Wolfe 条件计算所得,其中 $0<\sigma<\frac{1}{3}$。如果存在 $\gamma>0$,使得

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

则 $\boldsymbol{d}_k\neq0$ 且

$$\sum_{k\geqslant2}\|\boldsymbol{u}_k-\boldsymbol{u}_{k-1}\|^2<+\infty$$

其中 $\boldsymbol{u}_k=\frac{\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}$。

证明　假设 $\boldsymbol{d}_k \neq 0$,否则充分下降性条件式(12)不成立,故 $\boldsymbol{u}_k$ 的定义是有意义的。此外,由引理 4 和式(13),有

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

将 β_k^{LHSDL} 分为如下两部分:

$$\beta_k^{\mathrm{LHSDL}} = \beta_k^1 + \beta_k^2$$

其中

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

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

特别地,定义

$$\omega_k := \frac{\boldsymbol{v}_k}{\|\boldsymbol{d}_k\|}, \delta_k := \beta_k^1\frac{\|\boldsymbol{d}_{k-1}\|}{\|\boldsymbol{d}_k\|} \tag{15}$$

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

由 $\boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_k^{\mathrm{LHSDL}}\boldsymbol{d}_{k-1}$,对任意的 $k \geqslant 2$,有

$$\boldsymbol{u}_k = \boldsymbol{\omega}_k + \delta_k\boldsymbol{u}_{k-1} \tag{16}$$

根据 $\|\boldsymbol{u}_k\| = \|\boldsymbol{u}_{k-1}\| = 1$ 和式(15)可得

$$\|\boldsymbol{\omega}_k\| = \|\boldsymbol{u}_k - \delta_k\boldsymbol{u}_{k-1}\| = \|\delta_k\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\|$$

由 $\delta_k \geqslant 0$ 和式(16),则

$$\|\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\| \leqslant \|(1+\delta_k)\boldsymbol{u}_k - (1+\delta_k)\boldsymbol{u}_{k-1}\| \leqslant$$
$$\|\boldsymbol{u}_k - \delta_k\boldsymbol{u}_{k-1}\| + \|\delta_k\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\| = 2\|\boldsymbol{\omega}_k\|$$

根据强 Wolfe 条件式(5)可得

$$\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1} \geqslant (\sigma - 1)\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1} \tag{17}$$

因此

$$\left|\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right| \leqslant \frac{\sigma}{1-\sigma}$$

再由 $\boldsymbol{v}_k$ 的定义,式(7),式(9)和式(17)可得

$$\|\boldsymbol{v}_k\| \leqslant \|\boldsymbol{g}_k\| + t\left|\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right|\|\boldsymbol{d}_{k-1}\| \leqslant$$
$$\|\boldsymbol{g}_k\| + t\left|\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right|\|\boldsymbol{s}_{k-1}\| \leqslant \bar{\gamma} + 2t\frac{\sigma}{1-\sigma}B$$

因此

$$\sum_{k \geqslant 1}\|\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\|^2 \leqslant \sum_{k \geqslant 1}4\|\boldsymbol{\omega}_k\|^2 \leqslant 4\sum_{k \geqslant 1}\frac{\boldsymbol{v}_k^2}{\|\boldsymbol{d}_k\|^2} \leqslant$$
$$4\left(\bar{\gamma} + 2t\frac{\sigma}{1-\sigma}B\right)^2\sum_{k \geqslant 1}\frac{1}{\|\boldsymbol{d}_k\|^2} < \infty$$

证毕。

设 N^* 为正整数集合,由 $\lambda > 0$ 以及正整数 Δ,记

$$K_{k,\Delta}^{\lambda} := \{i \in \boldsymbol{N}^* : k \leqslant i \leqslant k + \Delta - 1, \|\boldsymbol{s}_{k-1}\| > \lambda\}$$

令 $|K_{k,\Delta}^{\lambda}|$ 表示 $K_{k,\Delta}^{\lambda}$ 中元素的个数,由性质 1 可以证明以下引理。

引理 6　假设 A,B 成立,考虑形如式(2)(3)和式(6)的 LHSDL 方法,α_k 由强 Wolfe 条件计算所得,其中 $0 < \sigma < \frac{1}{3}$。如果式(17)成立,存在 $\lambda > 0$ 使得对任意 $\Delta \in N^*$ 和任意 k_0,存在 $k \geqslant k_0$,有

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

具体的证明过程可参考文献[4]和[14],因此这里省略证明过程。

引理 7　假设 A,B 成立,考虑形如式(2)(3)和式(6)的 LHSDL 方法,α_k 由强 Wolfe 条件式计算所得,其中 $0 < \sigma < \frac{1}{3}$,则

$$\liminf_{x \to \infty}\|\boldsymbol{g}_k\| = 0$$

证明　反证法,假设 $\liminf\limits_{x \to \infty}\|\boldsymbol{g}_k\| > 0$,则对所有的 k,存在一个常数 $\gamma > 0$,使得 $\|\boldsymbol{g}_k\| \geqslant \gamma$。因此引理 5 和引理 6 成立。

定义 $\boldsymbol{u}_k = \frac{\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}$,对任意指标 l, k,其中 $l \geqslant k$,有

$$\boldsymbol{x}_l - \boldsymbol{x}_{k-1} = \sum_{i=k}^{l}\boldsymbol{x}_i - \boldsymbol{x}_{i-1} =$$
$$\sum_{i=k}^{l}\alpha_{i-1}\boldsymbol{d}_{i-1} = \sum_{i=k}^{l}\boldsymbol{u}_{i-1}\|\boldsymbol{s}_{i-1}\| =$$
$$\sum_{i=k}^{l}\|\boldsymbol{s}_{k-1}\|\boldsymbol{u}_{k-1} + \sum_{i=k}^{l}\|\boldsymbol{s}_{k-1}\|(\boldsymbol{u}_{i-1} - \boldsymbol{u}_{k-1})$$

由式(7)式和 $\|\boldsymbol{u}_k\| = 1$,有

$$\sum_{i=k}^{l}\|\boldsymbol{s}_{k-1}\| \leqslant \|\boldsymbol{x}_l - \boldsymbol{x}_{k-1}\| + \sum_{i=k}^{l}\|\boldsymbol{s}_{k-1}\|\|\boldsymbol{u}_{i-1} - \boldsymbol{u}_{k-1}\| \leqslant$$
$$2B + \sum_{i=k}^{l}\|\boldsymbol{s}_{k-1}\|\|\boldsymbol{u}_{i-1} - \boldsymbol{u}_{k-1}\| \tag{18}$$

设 $\lambda > 0$ 由引理 6 给出,定义 $\Delta := \left|\frac{8B}{\lambda}\right|$,其中 $|\cdot|$ 表示向上取整函数,即最小的整数不少于 $\frac{8B}{\lambda}$。由引理 5,存在指标 $k_0 \geqslant 1$,使得

$$\sum_{k \geqslant 1}\|\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\|^2 \leqslant \frac{1}{4\Delta} \tag{19}$$

对于这样的 Δ 和 k_0,引理 5 给出的指标 $k \geqslant k_0$,

使得

$$|K_{k,\Delta}^{\lambda}|>\frac{\Delta}{2} \tag{20}$$

对任意指标 $i\in[k,k+\Delta-1]$,由 Cauchy-Schwarz 不等式和式(19)可得:

$$\|\boldsymbol{u}_{i-1}-\boldsymbol{u}_{k-1}\|\leqslant\sum_{j=k}^{i-1}\|\boldsymbol{u}_j-\boldsymbol{u}_{j-1}\|\leqslant$$
$$(i-k)^{\frac{1}{2}}\left(\sum_{j=k}^{i-1}\|\boldsymbol{u}_j-\boldsymbol{u}_{j-1}\|^2\right)^{\frac{1}{2}}\leqslant$$
$$\Delta^{\frac{1}{2}}\left(\frac{1}{4\Delta}\right)^{\frac{1}{2}}=\frac{1}{2}$$

由式(19)和式(20),其中 $l=k+\Delta-1$,可得

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

这与 $\Delta\geqslant\frac{8B}{\lambda}$矛盾,故假设不成立。

4 数值实验

Dai 和 Kou[16]在 2013 年提出一个共轭参数:

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

以及截断形式:

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

将 LHSDL 方法分别与 DK+方法和 MNVHS 方法进行数值结果比较。在强 Wolfe 线搜索条件下选取文献[17]中的 87 个测试问题进行验证。参数 $\delta=0.01$,$\sigma=0.1$,算法的测试环境为 Matlab2012a,联想 Windows10 操作系统,Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz 1.80 GHz,RAM 4.00 GB 内存。

DK+方法中参数 $\eta=0.5$; LHSDL 方法和 MNVHS 方法中选取 $\mu=2$; LHSDL 方法中选取 $\mu=2$,$t=0.01$。当下面 3 种情形之一发生时,算法将会终止:(i) $\|\boldsymbol{g}_k\|_\infty\leqslant10^{-6}$; (ii) 迭代次数 $NI>50\ 000$; (iii) $T_{\mathrm{cpu}}>300$。如果终止准则(ii)(iii)满足时,则称该方法失败,并标记为“—”。

数值结果见表 1,其中 Time 表示所耗费的 CPU 时间(单位:s),NI 表示方法的迭代次数,NF 表示方法的函数计算次数,NG 表示方法的梯度计算次数。

表 1 数值结果

Table 1 Numerical results

	DK+	MNVHS	LHSDL
Time/s	1.000 0	0.921 0	0.912 0
NI	1.000 0	1.012 3	0.997 8
NF	1.000 0	0.968 2	0.958 4
NG	1.000 0	0.966 0	0.955 3

将 DK+方法作为比较标准,数值越小计算效果越好。因此由表 1 可以看出,LHSDL 方法略优于其他方法。

通过绘制性能曲线图[18]比较每一种方法的数值效果。图 1—图 4 分别对应的是在强 Wolfe 线搜索下 DK+,MNVHS 和 LHSDL 方法的计算时间、函数计算次数、梯度计算次数以及迭代次数的性能曲线。曲线越在上方越靠近 1,则计算效率越好。因此,从图 1—图 4 中可以看出 LHSDL 方法优于 MNVHS 方法和 DK+方法。

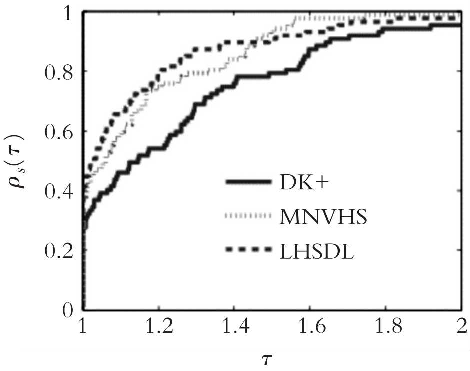

图 1 计算时间性能曲线

Fig. 1 CPU time performance curve

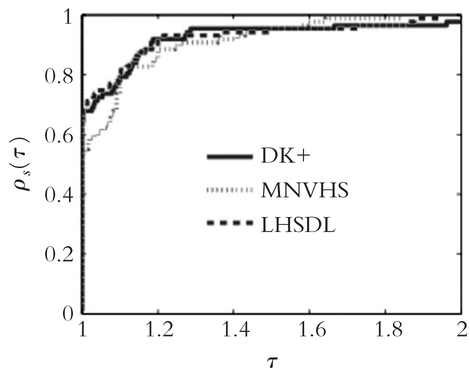

图 2 迭代次数性能曲线

Fig. 2 Iterative number performance curve

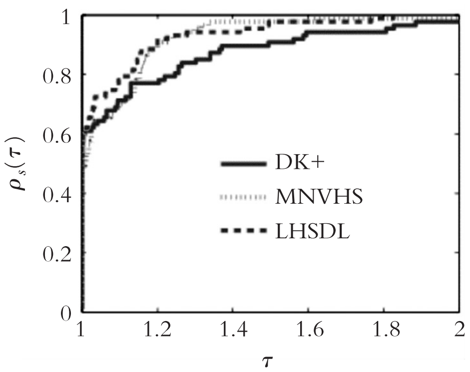

图 3　函数计算次数性能曲线

Fig. 3　Function calculation number performance curve

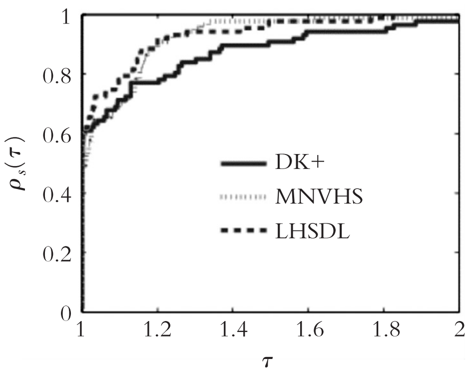

图 4　梯度计算次数性能曲线

Fig. 4　Gradient calculation number performance curve

5　结束语

通过将 DL 共轭梯度方法中的第一项$\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}$修正为 WYL 方法中的一个计算效果较好的式子,得到一个理论性质和数值计算均良好的混合 DL-WYL 共轭梯度法,简称 LHSDL 方法。理论证明了 LHSDL 方法在强 Wolfe 线搜索条件下满足充分下降性,且对一般函数是全局收敛的。由 Dolan 和 Moré 的性能曲线图表明 LHSDL 方法略优于 DK+方法、MNVHS 方法。

参考文献(References):

[1] 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):99—147

[2] POLYAK B T. The Conjugate Gradient Method in Extremal Problems[J]. USSR Computational Mathematics and Mathematical Physics,1969,9(4):94—112

[3] 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

[4] 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

[5] WEI Z X, YAO S W, LIU L Y. The Convergence Properties of Some New Conjugate Gradient Methods[J]. Applied Mathematics and Computation, 2006, 183(2): 1341—1350

[6] DAI Z F, WEN F H. Another Improved Wei-Yao-Liu Nonlinear Conjugate Gradient Method with Sufficient Descent Property[J]. Applied Mathematics and Computation,2012,218(14):7421—7430

[7] DU X W,ZHANG P,MA W Y. Some Modified Conjugate Gradient Methods for Unconstrained Optimization[J]. Journal of Computational and Applied Mathematics,2016, 305(1):92—114

[8] 陈恩. WYL 和 HZ 共轭梯度算法的改进和推广[D]. 重庆:重庆师范大学,2018
CHEN E. Improvement and Generalization of WYL and HZ Conjugate Gradient Algorithms[D]. Chongqing: Chongqing Normal University,2018(in Chinese)

[9] 马文亚. 基于韦增欣等的共轭梯度参数的修正共轭梯度法[D]. 重庆:重庆师范大学,2015
MA W Y. Conjugate Gradient Algorithms Based on the Conjugate Gradient Parameter of Wei et al[D]. Chongqing: Chongqing Normal University, 2015(in Chinese)

[10] YAO S W, QIN B. A Hybrid of DL and WYL Nonlinear Conjugate Gradient Methods[J]. Abstract and Applied Analysis,2014(2014):1—9

[11] YAO S W, LU X W, WEI Z X. A Conjugate Gradient Method with Global Convergence for Large-Scale Unconstrained Optimization Problems[J]. Journal of Applied Mathematics,2013(9):1—9

[12] CHENG Y L, MOU Q, PAN X B, et al. A Sufficient Descent Conjugate Gradient Method and Its Global Convergence[J]. Optimization Methods and Software,

2016,31(3):577—590

[13] 谢丽. 一类修正的 DL 共轭梯度法[J]. 周口师范学院学报,2019,36(5):31—35
XIE L. A Modified DL Conjugate Gradient Method[J]. Journal of Zhoukou Normal University,2019,36(5):31—35(in Chinese)

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

[15] DAI Y H, HAN J Y, LIU G H, et al. Convergence Properties of Nonlinear Conjugate Gradient Methods[J]. SIAM Journal on Optimization,2000,10(2):345—358

[16] DAI Y H, KOU C X. A Nonlinear Conjugate Gradient Algorithm with an Optimal Property and an Improved Wolf Line Search[J]. SIAM Journal on Optimization. 2013, 23(1):296—320

[17] GOULD N I M, ORBAN D, TOINT P L. CUTEr and SifDec: A Constained and Unconstrained and Testing Environment Revisited[J]. ACM Transactions on Mathematical Software,2003,29(4):373—394

[18] DOLAN E D, MORÉ J J. Benchmarking Optimization Software with Performance Profiles[J]. Mathematical Programming,2002,91(2):201—213

A Hybrid DL-WYL Conjugate Gradient Method

LI Yue

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

Abstract: Conjugate gradient method is widely used in many practical fields such as engineering problems, financial models and so on because of its simple iteration and low storage. For the large-scale unconstrained optimization problems, a hybrid DL-WYL conjugate gradient method is proposed—LHSDL method. It can be regarded as a modified DL conjugate gradient method,the first term of DL conjugate gradient method is modified by the conjugate parameter of Wei-Yao-Liu type conjugate gradient method which has good numerical and theoretical results. It can also be regarded as a modified WYL conjugate gradient method. By adding the second term of DL conjugate gradient method, the method may contain some Hessian information. The LHSDL method has a better property than DL method, i. e. , under the condition of strong Wolfe line search, it has sufficient descent property, and it is theoretically proved that LHSDL method has global convergence for general functions. Numerical experiments are carried out on a set of unconstrained optimization test problems of CUTEr collection. The performance profile of Dolan and Moré shows that LHSDL method is slightly superior to DK+method and MNVHS method.

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

责任编辑:李翠薇

引用本文/Cite this paper:

李月. 一类混合的 DL-WYL 共轭梯度法[J]. 重庆工商大学学报(自然科学版),2021,38(2):28—34

LI Y. A Hybrid DL-WYL Conjugate Gradient Method[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2021,38(2):28—34