

一类充分下降的混合 CD-LS 共轭梯度法^{*}

尹玉玲, 彭再云, 梁仁莉, 王 鹏
(重庆交通大学 数学与统计学院, 重庆 400074)

摘要:为了结合共轭下降(conjugate descent, CD)法良好的理论性质和 Liu-Storey (LS)法较好的数值效果,以降低小步长对迭代的不良影响,以及使搜索方向的下降性独立于线搜索的选择。通过混合 CD 法和 LS 法的分子,对梯度函数进行了相应的修正。方向的充分下降性独立于线搜索的选取,可应用于多种线搜索;基于 Wolfe 线搜索,证明了算法的全局收敛性。42 类无约束测试函数和图像去噪的结果表明,基于相同的终止条件所提出的算法的迭代次数和迭代时间均少于之前的 3 类共轭梯度算法。

关键词:混合共轭梯度法;充分下降性;无约束优化;图像去噪

中图分类号:O224

文献标志码:A

文章编号:1672-6693(2024)02-0026-10

本文主要讨论无约束最优化问题:

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

此时, $f: \mathbf{R}^n \rightarrow \mathbf{R}$ 是连续可微的, 梯度 $g(x) = \nabla f(x)$ 是 Lipschitz 连续的。

共轭梯度法的一般迭代形式为:

$$x_{k+1} = x_k + \alpha_k d_k, \quad (2)$$

其中: x_k 是一个解的第 k 次逼近; d_k 为搜索方向, 定义如下:

$$d_k = \begin{cases} -g_k, & k=0, \\ -g_k + \beta_k d_{k-1}, & k>0. \end{cases}$$

其中: $g_k := g(x_k)$, $\alpha_k > 0$ 是步长因子, β_k 为共轭梯度参数。

不同的 β_k 决定了不同的共轭梯度法, 一些经典的 β_k 包括 Fletcher-Reeves (FR)^[1]、Dai-Yuan (DY)^[2]、Conjugate Descent (CD)^[3]、Hestenes-Stiefel (HS)^[4]、Polak-Ribière Polak (PRP)^[5] 和 Liu-Storey (LS)^[6], 它们的参数分别为:

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

其中: $\|\cdot\|$ 是欧氏范数。

为了改进上述 6 种经典的共轭梯度方法, 许多学者引入了混合共轭梯度法。例如, Andrei^[7] 提出了 THREECG 法求解式(1), 该方法通过拟牛顿公式得到搜索方向满足共轭条件和下降条件; 同时, 基于水平集的有界性和梯度的 Lipschitz 连续性证明了该方法的全局收敛性定理。Jian 等人^[8] 提出了基于 DY 法和 FR 法的 JHJ 法求解式(1), 在每次迭代时不需要使用任何线搜索也可以生成下降方向。Abubakar 等人^[9] 提出了基于

* 收稿日期: 2023-09-28 修回日期: 2024-03-06 网络出版时间: 2024-05-17T12:04

资助项目: 重庆市研究生联合培养基地建设项目 (No. JDLHPYJD2021016); 重庆交通大学研究生科研创新项目 (No. 2022ST005); 国家自然科学基金面上项目 (No. 12271067); 重庆市教育委员会科学技术研究计划——重点项目 (No. KJZD-202200704), 青年项目 (No. KJQN202000710); 重庆市高校创新研究群体项目 (No. CXQT21021); 重庆市自然科学基金面上项目 (No. cstc2021jcyjmsxmX0080, No. CSTB2022NSCQ-MSX1498)

第一作者简介: 尹玉玲, 女, 研究方向为最优化理论与算法, E-mail: Yinyul@163.com; 通信作者: 彭再云, 男, 教授, 博士生导师, E-mail: pengzaiyun@126.com

网络出版地址: <https://link.cnki.net/urlid/50.1165.n.20240516.0459.002>

PRP 法和 LS 法的 TTPRLS 法求解式(1),并证明了该方法的全局收敛性定理。

文献[3]中 CD 法的方向下降性独立于线搜索,但是对步长的要求较高,如果在迭代过程中出现了一个较小的步长,就可能导致迭代过程中连续出现很多很小的步长。文献[6]中的 LS 方法具有较好的数值效果,但需要基于一些线搜索才能得到下降条件。因此,本文考虑将 CD 方法和 LS 方法相结合,并进行相应的修正,得到一类用于求解式(1)的新混合共轭梯度法(记为 HCDLS)。

1 一类充分下降的共轭梯度法及算法

本节通过混合 CD 法和 LS 法,提出了充分下降共轭梯度法, β_k 的更新公式定义如下:

$$\beta_k = \frac{v_k}{\omega_k} - \frac{v_k \mathbf{g}_k^T \mathbf{d}_{k-1}}{\omega_k^2}, \mathbf{g}_k := \mathbf{g}(\mathbf{x}_k), \quad (3)$$

$$v_k = \max\{\|\mathbf{g}_k\|^2, \mathbf{g}_k^T \mathbf{y}_{k-1}\}, \quad (4)$$

$$\gamma_k = 1 - t_k \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\omega_k}, \mathbf{y}_k = \mathbf{g}_k - \gamma_k \mathbf{g}_{k-1}, \quad (5)$$

$$t_k = \min\left\{\bar{t}, \max\left\{0, \frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{s}_{k-1})}{\|\mathbf{g}_k\|^2}\right\}\right\}, 0 \leq t_k \leq \bar{t} < \sqrt{2} - 1, \quad (6)$$

$$\omega_k = -2\mathbf{d}_{k-1}^T \mathbf{g}_{k-1}. \quad (7)$$

基于 Wolfe 线搜索,给出如下算法步骤。

算法 1 (HCDLS 法)初始化: $\mathbf{x}_0 \in \mathbf{R}^n, \varepsilon \in (0, 1), \sigma_1 \in (0, 1), \sigma_2 \in (0, 1), \sigma_1 < \sigma_2, \bar{t} \geq 0$, 使 $\mathbf{d}_0 = -\mathbf{g}_0, k := 0$ 。

步骤 1, 若 $\|\mathbf{g}_0\| \leq \varepsilon$, 则停止; 否则, 进入步骤 2。

步骤 2, 使用 Wolfe 步长规则计算步长 α_k :

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + \sigma_1 \alpha_k \mathbf{g}_k^T \mathbf{d}_k, \quad (8)$$

$$\nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k \geq \sigma_2 \mathbf{g}_k^T \mathbf{d}_k. \quad (9)$$

步骤 3, 更新迭代 $\mathbf{x}_{k+1} := \mathbf{x}_k + \alpha_k \mathbf{d}_k$ 。

步骤 4, 根据式(3)计算 β_{k+1} , 并计算搜索方向:

$$\mathbf{d}_{k+1} = -\gamma_{k+1} \mathbf{g}_{k+1} + \beta_{k+1} \mathbf{d}_k. \quad (10)$$

步骤 5, 使 $k := k+1$, 并回到步骤 1。

下面的引理表明, HCDLS 在不依赖线搜索的情况下, 满足充分下降条件。

引理 1 考虑 HCDLS 法中的 β_{k+1} 由式(3)生成, 则搜索方向 $\mathbf{d}_{k+1}$ 满足:

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} \leq -c \|\mathbf{g}_{k+1}\|^2, c > 0. \quad (11)$$

证明 当 $v_{k+1} = \|\mathbf{g}_{k+1}\|^2$ 时, 在式(10)两边同时乘 $\mathbf{g}_{k+1}^T$, 并将式(5)和式(7)代入可得:

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} &= -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k - \frac{\|\gamma_{k+1}\|^2}{\omega_{k+1}^2} (\mathbf{g}_{k+1}^T \mathbf{d}_k)^2 + t_{k+1} \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{g}_{k+1} = \\ &= -\|\mathbf{g}_{k+1}\|^2 + (1+t_{k+1}) \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} (\mathbf{g}_{k+1}^T \mathbf{d}_k)^2. \end{aligned}$$

根据 $2\mathbf{u}^T \mathbf{v} \leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$, 其中 $\mathbf{u} = \frac{(t_{k+1}+1)\mathbf{g}_{k+1}}{2}, \mathbf{v} = \frac{\mathbf{g}_{k+1}}{\omega_{k+1}} (\mathbf{g}_{k+1}^T \mathbf{d}_k)$, 可得:

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} &\leq -\|\mathbf{g}_{k+1}\|^2 + \frac{(1+t_{k+1})^2}{4} \|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} (\mathbf{g}_{k+1}^T \mathbf{d}_k)^2 - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} (\mathbf{g}_{k+1}^T \mathbf{d}_k)^2 = \\ &= -\|\mathbf{g}_{k+1}\|^2 + \frac{(1+t_{k+1})^2}{4} \|\mathbf{g}_{k+1}\|^2 \leq -\left(1 - \frac{(1+\bar{t})^2}{4}\right) \|\mathbf{g}_{k+1}\|^2. \end{aligned}$$

当 $v_{k+1} = \mathbf{g}_{k+1}^T \mathbf{y}_k$ 时, 在式(10)两边同时乘 $\mathbf{g}_{k+1}^T$, 可得:

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} = -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{y}_k^T \mathbf{d}_k - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} \mathbf{y}_k^T \mathbf{d}_k \mathbf{g}_{k+1}^T \mathbf{d}_k + t_{k+1} \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k.$$

将 $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ 代入上式, 可得:

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} = & -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_k^T \mathbf{d}_k + t_{k+1} \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} \mathbf{y}_k^T \mathbf{d}_k \mathbf{g}_{k+1}^T \mathbf{d}_k = \\ & -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k + t_{k+1} \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_k^T \mathbf{d}_k - \frac{\mathbf{g}_{k+1}^T \mathbf{y}_k \mathbf{g}_{k+1}^T \mathbf{d}_k}{\omega_{k+1}^2} \mathbf{g}_{k+1}^T \mathbf{d}_k, \end{aligned}$$

其中: $\omega_{k+1} = \mathbf{g}_{k+1}^T \mathbf{y}_k$, 即 $\|\mathbf{g}_{k+1}\|^2 \leq \mathbf{g}_{k+1}^T \mathbf{y}_k$, 于是有:

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} \leq & -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k + t_{k+1} \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} (\mathbf{g}_{k+1}^T \mathbf{d}_k)^2 - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_k^T \mathbf{d}_k = \\ & -\|\mathbf{g}_{k+1}\|^2 + (1+t_{k+1}) \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_{k+1}^T \mathbf{d}_k - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} (\mathbf{g}_{k+1}^T \mathbf{d}_k)^2 - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_k^T \mathbf{d}_k \leq \\ & -\|\mathbf{g}_{k+1}\|^2 + \frac{(1+t_{k+1})^2}{4} \|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} (\mathbf{g}_{k+1}^T \mathbf{d}_k)^2 - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}^2} (\mathbf{g}_{k+1}^T \mathbf{d}_k)^2 - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_k^T \mathbf{d}_k = \\ & -\|\mathbf{g}_{k+1}\|^2 + \frac{(1+t_{k+1})^2}{4} \|\mathbf{g}_{k+1}\|^2 - \frac{\|\mathbf{g}_{k+1}\|^2}{\omega_{k+1}} \mathbf{g}_k^T \mathbf{d}_k. \end{aligned}$$

其中: $\omega_{k+1} = -2\mathbf{d}_k^T \mathbf{g}_k$, 于是有:

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} \leq -\|\mathbf{g}_{k+1}\|^2 + \frac{(1+t_{k+1})^2}{4} \|\mathbf{g}_{k+1}\|^2 + \frac{1}{2} \|\mathbf{g}_{k+1}\|^2 \leq -\left(\frac{1}{2} - \frac{(1+\bar{t})^2}{4}\right) \|\mathbf{g}_{k+1}\|^2.$$

综上, $\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} \leq -c \|\mathbf{g}_{k+1}\|^2$, 其中 $c = \max\left\{1 - \frac{(1+\bar{t})^2}{4}, \frac{1}{2} - \frac{(1+\bar{t})^2}{4}\right\}$, $0 \leq \bar{t} < \sqrt{2} - 1$. 证毕

注 1 文献[10]中 HRMIL 法需在强 Wolfe 线搜索下产生下降方向, 文献[11]中 HFRBA 的下降方向也需基于 Wolfe 线搜索得到, 由引理 1 可得 HCDLS 法方向的充分下降性独立于线搜索的选择, HCDLS 相较于文献[10-11]所提出的共轭梯度法可应用于更多的线搜索。

2 全局收敛

为建立 HCDLS 法的全局收敛性, 对目标函数做以下 2 个假设。

假设 1 水平集 $H = \{\mathbf{x} : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ 有界, 其中 $\mathbf{x}_0$ 为初始点。

假设 2 目标函数 $f(\mathbf{x})$ 在水平集 H 的某一邻域 U 内连续可微, 梯度函数 $\mathbf{g}(\mathbf{x})$ 满足 Lipschitz 连续, 即存在常数 $L > 0$, 对任意 $\mathbf{x}, \mathbf{y} \in U$, 使得:

$$\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|. \quad (12)$$

引理 2 当假设 1 和假设 2 成立时, $f(\mathbf{x})$ 在水平集 H 上有下界。

证明 首先证明 $f(\mathbf{x})$ 在 U 上连续时, 水平集 $H = \{\mathbf{x} : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ 是闭集。任取 $\mathbf{x}' \in \overline{H}$, 则存在 $\{\mathbf{x}_n\} \subset H$, 使得 $\mathbf{x}_n \rightarrow \mathbf{x}'$ 成立。由于 $f(\mathbf{x})$ 在 U 上连续, 于是 $f(\mathbf{x}_n) \rightarrow f(\mathbf{x}')$, 又因为 $\{\mathbf{x}_n\} \subset H$, 所以 $f(\mathbf{x}_n) \leq f(\mathbf{x}_0)$ 。根据极限的保号性可得 $f(\mathbf{x}') \leq f(\mathbf{x}_0)$, 则 $\mathbf{x}' \in H$, 即 $\overline{H} \subset H$, 所以水平集 $H = \{\mathbf{x} : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ 是闭集。由于 $f(\mathbf{x})$ 在 U 上连续, 且 $H = \{\mathbf{x} : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ 是有界闭集, 可得 $f(\mathbf{x})$ 在 H 上有下界。即当假设 1 和假设 2 成立时, $f(\mathbf{x})$ 在水平集 H 上有下界。证毕

接下来, 将得到下面的定理。

定理 1 如果假设 1 和假设 2 成立, 在 HCDLS 法中使用 Wolfe 线搜索, 若:

$$\sum_{k=0}^{\infty} \frac{1}{\|\mathbf{d}_k\|^2} = +\infty, \quad (13)$$

则有:

$$\liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0. \quad (14)$$

证明 假设式(14)不成立, 那么存在一个非负常数 M , 使得对于任意 $k \geq 0$, 都有:

$$\|\mathbf{g}_k\| \geq M. \quad (15)$$

由引理 1 和式(8), 可得:

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + \sigma_1 \alpha_k \mathbf{g}_k^T \mathbf{d}_k \leq f(\mathbf{x}_k) - \sigma_1 c \alpha_k \|\mathbf{g}_k\|^2 \leq f(\mathbf{x}_k) \leq f(\mathbf{x}_{k-1}) \leq \dots \leq f(\mathbf{x}_0). \quad (16)$$

式(9)两边同时减 $\mathbf{g}_k^T \mathbf{d}_k$, 则有 $\sigma_2 \mathbf{g}_k^T \mathbf{d}_k - \mathbf{g}_k^T \mathbf{d}_k \leq \mathbf{g}_{k+1}^T \mathbf{d}_k - \mathbf{g}_k^T \mathbf{d}_k$, 即:

$$-(1-\sigma_2)\mathbf{g}_k^\top \mathbf{d}_k \leq (\mathbf{g}_{k+1} - \mathbf{g}_k)^\top \mathbf{d}_k. \quad (17)$$

那么,式(17)结合引理 1、假设 2 和式(2),有:

$$-(1-\sigma_2)\mathbf{g}_k^\top \mathbf{d}_k \leq (\mathbf{g}_{k+1} - \mathbf{g}_k)^\top \mathbf{d}_k \leq \|\mathbf{g}_{k+1} - \mathbf{g}_k\| \|\mathbf{d}_k\| \leq L\alpha_k \|\mathbf{d}_k\|^2. \quad (18)$$

根据式(16)、(18)可得 $f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) \leq \sigma_1 \alpha_k \mathbf{g}_k^\top \mathbf{d}_k$, $\frac{-(1-\sigma_2)\mathbf{g}_k^\top \mathbf{d}_k}{L\|\mathbf{d}_k\|^2} \leq \alpha_k$ 。于是,有 $\frac{\sigma_1(1-\sigma_2)}{L} \frac{(\mathbf{g}_k^\top \mathbf{d}_k)^2}{\|\mathbf{d}_k\|^2} \leq f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})$ 成立。即:

$$\frac{\sigma_1(1-\sigma_2)}{L} \sum_{k=0}^{\infty} \frac{(\mathbf{g}_k^\top \mathbf{d}_k)^2}{\|\mathbf{d}_k\|^2} \leq (f(\mathbf{x}_0) - f(\mathbf{x}_1)) + (f(\mathbf{x}_1) - f(\mathbf{x}_2)) + \cdots \leq f(\mathbf{x}_0) - f(\mathbf{x}_{n+1}).$$

由引理 1 可知搜索方向 $\mathbf{d}_k$ 满足下降条件,并从式(8)、(9)及引理 2 可以看出, $\{f(\mathbf{x}_k)\}$ 是有下界的递减数列。因此,数列 $\{f(\mathbf{x}_k)\}$ 是收敛的。为了证明级数 $\sum_{k=0}^{\infty} (f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}))$ 的收敛性,需先证明数列 $\{S_n\}$ 部分和的收敛性。将 $\{S_n\}$ 定义为:

$$\begin{aligned} S_n &= \sum_{k=0}^n (f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})) = \\ &= f(\mathbf{x}_0) - f(\mathbf{x}_1) + f(\mathbf{x}_1) - f(\mathbf{x}_2) + \cdots + f(\mathbf{x}_n) - f(\mathbf{x}_{n+1}) = \\ &= f(\mathbf{x}_0) - f(\mathbf{x}_{n+1}). \end{aligned}$$

于是,有 $\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} (f(\mathbf{x}_0) - f(\mathbf{x}_{n+1})) = f(\mathbf{x}_0) - \lim_{n \rightarrow \infty} f(\mathbf{x}_{n+1}) = \kappa \in \mathbf{R}$ 。这表明级数 $\sum_{k=0}^{\infty} (f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}))$ 收敛。则有:

$$\sum_{k=0}^{\infty} \frac{(\mathbf{g}_k^\top \mathbf{d}_k)^2}{\|\mathbf{d}_k\|^2} < +\infty \quad (19)$$

成立。根据引理 1 及式(15)可得 $\mathbf{g}_k^\top \mathbf{d}_k \leq -c\|\mathbf{g}_k\|^2 \leq -cM^2$, 两边同时乘 $\mathbf{g}_k^\top \mathbf{d}_k$, 且 $\mathbf{g}_k^\top \mathbf{d}_k \leq 0$, 则 $(\mathbf{g}_k^\top \mathbf{d}_k)^2 \geq -c(\mathbf{g}_k^\top \mathbf{d}_k)\|\mathbf{g}_k\|^2$ 。结合 $\mathbf{g}_k^\top \mathbf{d}_k \leq -c\|\mathbf{g}_k\|^2$, 于是有 $(\mathbf{g}_k^\top \mathbf{d}_k)^2 \geq c^2 M^4$, 即:

$$\sum_{k=0}^{\infty} (\mathbf{g}_k^\top \mathbf{d}_k)^2 \geq c^2 \sum_{k=0}^{\infty} M^4. \quad (20)$$

式(20)两边同时除以 $\|\mathbf{d}_k\|^2 \neq 0$, 便有:

$$\sum_{k=0}^{\infty} \frac{(\mathbf{g}_k^\top \mathbf{d}_k)^2}{\|\mathbf{d}_k\|^2} \geq c \sum_{k=0}^{\infty} \frac{M^4}{\|\mathbf{d}_k\|^2} = +\infty. \quad (21)$$

那么式(19)、(21)矛盾,故假设不成立,则 $\lim_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$ 。证毕

3 数值实验

3.1 测试函数

下面对本文提出的方法(HCDLS)进行数值实验,测试并将该方法的计算性能与一些有效的共轭梯度法进行比较。运用 HCDLS、TTPRLS^[9]、LLS2^[12] 和 MHZ^[13] 解决常用的无约束测试函数集^[14] 中的 42 类问题。

实验代码是用 MATLAB R2020 编写的(MACOS, M2, 8 GB RAM), 本文的参数设定如下: $\bar{t} = 0.3$, $\sigma_1 = 0.0001$ 和 $\sigma_2 = 0.009$ 。停止标准设置为: 1) $\|\mathbf{g}_k\| \leq \epsilon = 10^{-6}$; 2) 迭代次数 $I > 2000$ 。表 1 给出了测试函数名及维度。为了说明 HCDLS 法的有效性,选取了 50~12000 维的问题维度进行测试。

表 2 给出了测试函数的数值结果,其中 I 、 t_{CPU} 和 E 分别表示迭代次数、CPU 耗时和梯度精确值。

为了分析所有方法的计算性能,使用 Dolan 和 Moré^J 提出的性能曲线^[15]。对于求解器的数目 n_s 及问题的数目 n_p , 性能曲线 $\omega: \mathbf{R} \rightarrow [0, 1]$ 定义为如下形式。

设 P 是一组测试问题, S 是求解器的集合, n_s 是方法数, n_p 是测试问题数。性能曲线 $\omega: \mathbf{R} \rightarrow [0, 1]$ 表示为:

$$\omega_s(\tau) = \frac{1}{n_p} l_{\text{size}} \{p \in P : \log_2 r_{p,s} \leq \tau\}.$$

其中: $\tau > 0$; 性能比 $r_{p,s} = \frac{t_{\text{CPU}_{p,s}}}{\min\{b_{p,s} : p \in P, s \in S\}}$, $t_{\text{CPU}_{p,s}} > 0$ 是解决问题 p 所需要的 CPU 时间; $l_{\text{size}} \{p \in P :$

$\log_2 r_{p,s} \leq \tau$ 是在集合 $\{p \in P : \log_2 r_{p,s} \leq \tau\}$ 里的元素个数。因此, $\omega_s(\tau)$ 是方法 s 的概率, $\log_2 r_{p,s}$ 在最佳可能比率的因子 τ 内。具体图像见图 1、图 2。通常性能曲线位于顶部的方法将优于其余方法或代表最佳方法。

表 1 测试函数

Tab. 1 The test function

序号	函数名称	序号	函数名称	序号	函数名称
1	Randay 1(3 000)	15	Diadonal 4(5 000)	29	ARGLINB(300)
2	Randay 1(5 000)	16	Extended PSC1(3 000)	30	ARWHEAD(500)
3	Extended Rosenbrock(3 000)	17	Extended PSC1(8 000)	31	NONDIA(5 000)
4	Extended Rosenbrock(5 000)	18	Full Hessian FH2(50)	32	NONDIA(10 000)
5	Extended Rosenbrock(8 000)	19	Extended BD1(10 000)	33	Broyden Tridiagonal(5 000)
6	Extended Rosenbrock(12 000)	20	Extended Maratos(3 000)	34	LIARWHD(5 000)
7	Extended Beale(3 000)	21	Extended Maratos(8 000)	35	LIARWHD(8 000)
8	Extended Beale(8 000)	22	Extended Wood(12 000)	36	ENGVAL1(5 000)
9	Randay 2(8 000)	23	Extended Hiebert(1 000)	37	GENHUMPS(1 000)
10	Extended Freudenstein & Roth(3 000)	24	Extended Hiebert(5 000)	38	GENHUMPS(3 000)
11	Generalized Tridiagonal 1(3 000)	25	Extended quadratic penalty QP2(3 000)	39	NONSCOMP(5 000)
12	Generalized Tridiagonal 1(5 000)	26	Extended Tridiagonal 2(8 000)	40	LIAREHD(8 000)
13	Generalized Tridiagonal 1(8 000)	27	FLETCHBV3(100)	41	Diagonal 7(5 000)
14	Extended TET(1 000)	28	FLETCHCR(500)	42	HIMMELH(8 000)

3.2 图像去噪

下面将所提出的 HCDLS 法应用于含椒盐噪音的图像去噪问题。

Raymond 等人^[16]利用两相方案来恢复被脉冲噪声破坏的图像。在第 1 阶段,使用中值滤波器来检测噪声像素。设 X 大小为 $M \times N$ 且 $A = \{1, 2, \dots, M\} \times \{1, 2, \dots, N\}$ 为图像 X 的索引集。 $\mathcal{N} \subset A$ 是从第 1 阶段检测到噪声像素的索引集, $|\mathcal{N}|$ 表示在 $\mathcal{N}$ 中的元素个数。令 $\nu_{i,j}$ 是像素位置 $(i, j) \in A$ 处像素的 4 个最近邻的集合。例如: $\nu_{i,j} = \{(i, j-1), (i, j+1), (i-1, j), (i+1, j)\}$ 以及 $y_{i,j}$ 是在像素位置 (i, j) 处观察到的图像的像素值。在第 2 阶段,通过解决非光滑最小化问题来清理噪声像素:

$$\min_u \sum_{(i,j) \in \mathcal{N}} \left[|\nu_{i,j} - y_{i,j}| + \frac{\beta}{2} (2S_{i,j}^1 + S_{i,j}^2) \right]. \quad (22)$$

此时,有:

$$S_{i,j}^1 = \sum_{(m,n) \in \nu_{i,j}/\mathcal{N}} \psi_a(\mu_{i,j} - y_{m,n}), S_{i,j}^2 = \sum_{(m,n) \in \nu_{i,j} \cap \mathcal{N}} \psi_a(\mu_{i,j} - \mu_{m,n}).$$

其中: $\psi_a = \sqrt{t^2 + \alpha}$ 是一个边缘保留函数,参数 $\alpha > 0$, 并且 $\boldsymbol{\mu} = [\mu_{i,j}]_{(i,j) \in \mathcal{N}}$ 是长度为 $|\mathcal{N}|$ 的行向量。无论如何,精确地解决非光滑最小化问题(22)是耗时巨大的。Cai 等人^[17]删除了非平滑项,并获得了以下平滑无约束优化:

$$\min_{\boldsymbol{\mu}} F_a(\boldsymbol{\mu}) = \sum_{(i,j) \in \mathcal{N}} \left(2 \sum_{(m,n) \in \nu_{i,j}/\mathcal{N}} \psi_a(\mu_{i,j} - y_{m,n}) + \sum_{(m,n) \in \nu_{i,j} \cap \mathcal{N}} \psi_a(\mu_{i,j} - \mu_{m,n}) \right). \quad (23)$$

显然,噪声比越高,式(23)的刻度越大。文献[17]的作者发现,即使噪声很高,通过使用共轭梯度法可以有效地恢复问题(23)受污染的图像。

现将应用两相方案来解决椒盐噪声问题,这是脉冲噪声的一个特例。在第 1 阶段,本文使用自适应中值滤波器^[18]来检测噪声像素。在第 2 阶段,本文使用 TTCDLs 求解(23),并与 TTPRLS、JHJ、LLS2 进行比较。

测试图像 Boat、Hill、Lena 和 Man 的分辨率均为 512×512 。所有用于比较的方法停止标准均为 $I > 300$ 或者 $\frac{|F_a(\mu_k) - F_a(\mu_{k-1})|}{F_a(\mu_k)} \leq 10^{-4}$ 。

表 2 测试函数的数值结果
Tab. 2 The numerical results of the test functions

问题	MHZ			TTPRLS			MHZ			HCDLS		
	I	t_{CPU}/s	E	I	t_{CPU}/s	E	I	t_{CPU}/s	E	I	t_{CPU}/s	E
1	—	—	—	—	34.77	3.40E-02	—	38.15	5.59E-05	—	20.97	6.31E-04
2	—	—	—	—	12.61	1.10E-03	—	14.22	4.59E-06	1480	6.57	8.24E-06
3	1205	0.15	8.98E-07	142	0.04	8.86E-07	—	0.12	2.62E-05	71	0.02	1.32E-06
4	1646	4.81	9.58E-04	106	0.28	4.52E-07	984	0.39	9.36E-07	108	0.26	1.26E-06
5	1363	7.45	9.51E-07	123	0.71	3.19E-07	173	0.26	5.06E-07	100	0.52	2.83E-06
6	1140	14.33	9.64E-07	89	0.87	9.89E-07	—	2.13	5.25E-05	82	0.53	4.07E-07
7	186	7.39	9.01E-07	72	2.13	3.71E-07	42	0.25	6.51E-07	32	0.32	2.67E-07
8	252	14.75	9.06E-07	43	2.47	4.79E-07	51	0.58	5.44E-08	32	0.51	3.38E-07
9	—	—	—	203	3.64	9.85E-05	—	38.15	5.59E-05	7	0.02	1.72E-07
10	—	—	—	—	5.09	1.57E-07	1866	14.22	4.59E-06	554	1.01	4.43E-07
11	43	0.32	6.97E-07	32	0.94	9.68E-07	47	0.12	2.62E-05	56	0.52	4.30E-07
12	57	5.67	7.74E-07	54	3.29	8.75E-07	47	0.39	9.36E-07	57	0.95	9.18E-07
13	575	20.25	7.89E-07	42	1.89	9.12E-07	57	0.26	5.06E-07	37	0.83	8.41E-07
14	—	0.41	—	42	0.17	8.60E-07	115	2.13	5.25E-05	34	0.05	5.61E-07
15	227	0.18	9.68E-07	70	0.17	2.06E-07	1436	0.25	6.51E-07	52	0.13	6.31E-07
16	431	5.24	9.30E-07	—	18.44	3.98E-05	103	0.58	5.44E-08	67	0.55	9.37E-07
17	1951	78.04	6.84E-07	—	66.19	2.57E-07	43	0.87	1.71E-07	39	0.85	2.88E-07
18	1700	0.29	8.79E-07	644	0.06	9.63E-07	489	0.05	5.75E-07	507	0.03	7.60E-07
19	43	1.69	8.48E-07	27	3.24	3.57E-07	33	2.05	5.81E-07	24	2.95	9.87E-07
20	—	1.73	1.16E-07	235	0.19	9.47E-07	—	0.15	509.32	210	0.11	3.40E-07
21	—	5.92	1.77E-07	—	7.74	2.46E-07	—	—	1.20E+03	836	3.04	5.92E-07
22	1697	17.06	9.61E-07	333	3.28	5.52E-07	—	—	1.37E+03	209	1.64	5.13E-07
23	—	—	—	—	—	57.39	—	0.01	1.22E-07	620	0.01	6.79E-07
24	—	—	—	—	—	0.38	—	0.25	3.49E+06	—	2.49	7.96E-07
25	976	31.32	9.81E-07	—	21.08	4.47E-06	217	0.38	9.36E-07	65	0.28	9.71E-07
26	—	1.12	—	—	11.99	7.07E-06	—	12.88	3.41E-06	285	1.76	7.70E-07
27	—	0.14	1.10E-03	1227	0.97	7.67E-09	—	0.22	6.95E-04	1626	0.99	1.32E-04
28	—	0.27	1.20E-03	—	0.15	8.36E-05	—	0.05	3.59E-03	—	0.05	9.98E-07
29	—	—	—	—	33.65	1.69E-04	—	34.67	1.62E-04	—	35.11	2.19E-07
30	—	0.37	3.23E-04	—	0.43	1.61E-05	—	0.39	6.75E-05	55	0.01	8.49E-07
31	—	7.53	3.20E-03	1242	1.61	9.67E-07	—	0.48	2.40E-04	63	0.05	5.44E-07
32	—	21.13	1.62E-02	890	2.11	6.91E-07	—	19.34	2.64E-06	90	0.21	4.24E-07
33	46	1.09	9.96E-07	101	6.60	7.66E-07	693	3.08	9.87E-07	79	0.66	4.19E-07
34	—	8.63	1.09E-04	153	0.36	2.09E-07	—	0.68	7.70E-03	110	0.18	4.65E-07
35	—	—	1.27	93	0.29	5.78E-07	—	0.96	9.93E-04	107	0.21	1.02E-07
36	606	1.69	8.87E-07	870	2.54	7.56E-07	915	2.63	3.58E-07	402	1.15	7.10E-07
37	—	0.10	2.52	—	0.11	0.06	—	0.01	3.7	—	0.05	7.45E-07
38	119	1.27	9.94E-07	65	2.36	7.54E-07	894	1.28	9.64E-07	63	0.61	4.19E-07
39	64	0.49	5.73E-07	—	8.15	1.36E-07	1196	0.16	9.94E-07	39	0.03	6.42E-07
40	—	15.72	1.27	93	0.30	5.78E-05	—	0.94	9.93E-04	107	0.25	3.45E-07
41	34	2.60	8.24E-06	—	35.12	4.52E-07	41	0.50	3.35E-07	17	0.18	1.59E-07
42	—	4.44	—	—	67.08	1.34E-05	217	5.87	9.16E-07	140	3.63	4.58E-07

注:表中“—”表示 $I > 2\,000$ 时算法停止,或 E 在求解过程缺失,又或 $E \geq 10^{-3}$ 时 t_{CPU} 无法获得。

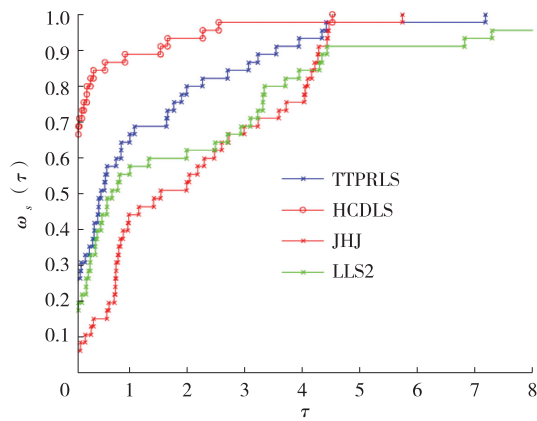


图 1 次数性能

Fig. 1 Iter performance chart

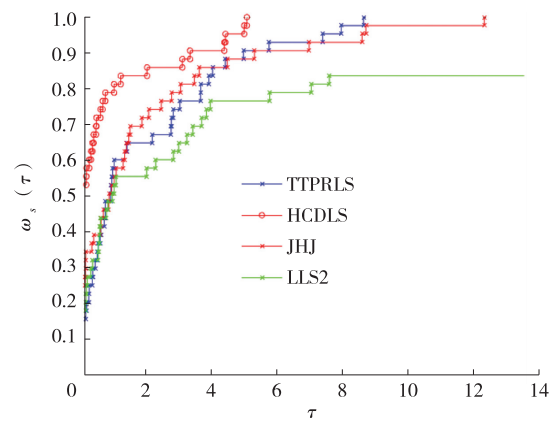


图 2 时间性能

Fig. 2 Time performance chart

本例的操作环境与之前测试函数的环境相同。本文采用峰值信噪比(R_{PSNR})对恢复性能进行定性评估,计算公式为 $R_{\text{PSNR}} = 10 \log \frac{255^2}{\sum_{i,j} (x_{i,j}^r - x_{i,j}^*)^2 / MN}$, 其中 $x_{i,j}^r$ 和 $x_{i,j}^*$ 分别表示恢复的图像和原始图像的像素值。

表 3 给出了不同算法下图像去噪结果,其中 I 、 t_{CPU} 和 R_{PSNR} 分别表示已恢复的迭代次数、CPU 时间和 PSNR 值。本文只给出了 90% 的噪声图像和恢复图像,见图 3、图 4。

表 3 图像去噪的运行时间、次数及峰值信噪比

Tab. 3 The running time, frequency and peak signal-to-nois ratio of image denoising

图像	噪声比	TTPRLS			LLS2			JHJ			HCDLS		
		I	t_{CPU}/s	R_{PSNR}	I	t_{CPU}/s	R_{PSNR}	I	t_{CPU}/s	R_{PSNR}	I	t_{CPU}/s	R_{PSNR}
Lena	30%	19	3.25	35.71	16	2.34	36.82	61	4.19	36.16	13	2.12	36.95
	50%	28	6.36	33.84	19	4.12	34.16	85	9.04	33.55	13	3.62	34.14
	70%	23	7.25	27.29	24	5.06	30.10	81	9.84	30.09	16	4.37	31.03
	90%	18	6.75	17.92	37	7.13	24.63	99	11.90	25.55	22	5.82	25.90
Baboon	30%	20	2.66	26.51	13	2.12	26.52	43	3.36	26.43	11	2.02	26.55
	50%	18	4.89	24.25	16	3.86	24.54	48	6.10	24.44	13	3.51	24.55
	70%	17	5.41	21.74	23	4.98	22.45	62	8.12	22.34	16	4.39	22.46
	90%	15	5.85	16.89	32	6.73	19.92	85	10.72	20.03	20	5.66	20.10
Man	30%	28	1.05	29.49	18	0.64	29.58	80	1.37	29.34	14	0.57	29.62
	50%	28	1.77	26.63	20	1.06	27.17	68	2.00	26.94	14	0.92	27.30
	70%	19	1.91	20.92	29	1.43	24.28	90	2.78	24.46	18	1.15	24.70
	90%	17	1.61	15.22	45	1.94	20.22	77	2.49	20.85	26	1.45	21.09
Barbara	30%	23	3.15	28.61	16	2.29	28.66	59	4.11	28.49	11	2.01	28.66
	50%	12	3.82	22.65	18	4.04	26.66	62	7.19	26.42	13	3.59	26.70
	70%	27	8.10	24.11	28	5.57	24.59	78	9.58	24.37	17	4.52	24.63
	90%	13	5.46	15.72	38	7.17	21.59	96	11.85	22.19	23	5.93	22.47

注:加粗的数字表示不同算法中的最优结果。



图 3 不同图像在 90%椒盐噪音下的恢复情况

Fig. 3 Numerical results of image denoising (90% salt and pepper noise)



图 4 不同算法在 90%椒盐噪音下的图像恢复情况

Fig. 4 Numerical results of image denoising (90% salt and pepper noise)

从表 3 可以看出,HCDLS 法可以获得与其余 3 种方法相同的峰值信噪比,但迭代次数和运行时间均更少。

4 结论

本文提出了一种混合共轭梯度法(HCDLS),该方法不仅可以避免连续小步长,并具有较好的全局收敛性。其中搜索方向始终满足下降方向,并且与线搜索的选择无关。在一些合适的条件下,可以得到 HCDLS 的全局收敛性,并应用于处理大中型无约束优化测试函数和图像去噪问题中,数值实验表明:1) 求解无约束优化测试函数,在达到同一精度下,HCDLS 法的运行时间以及迭代次数整体低于 MHZ、TTPRLS 和 LLS2 法。并且对于“Extended Hiebert”问题,只有 HCDLS 可以达到“ $\|g_k\| \leq 10^{-6}$ ”。2) 求解图像去噪问题,HCDLS 法可以获得与 MHZ、TTPRLS 和 LLS2 法相同的峰值信噪比,但迭代次数和运行时间均更少。

因此,从整体优化效率看,HCDLS 法优于 MHZ 法、TTPRLS 法和 LLS2 法。

参考文献:

- [1] FLETCHER R,REEVES C M. Function minimization by conjugate gradients[J]. Computer Journal,1969,2:149-154.
- [2] 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.
- [3] FLETCHER R. Practical methods of optimization[J]. SIAM Review,1984,26(1):143-144.
- [4] HESTENES M R,STIEFEL E L. Methods of conjugate gradients for solving linear systems[J]. Journal of Research of the National Bureau of Standards,1952,49:403-436.
- [5] POLAK B T. The conjugate gradient method in extreme problems[J]. USSR Computational Mathematics and Mathematical Physics,1969,9(4):94-112.
- [6] LIU Y,STOREY C. Efficient generalized conjugate gradient algorithms,part 1:theory[J]. Journal of Optimization Theory and Applications,1991,69(1):129-137.
- [7] ANDREI N. A simple three-term conjugate gradient algorithm for unconstrained optimization[J]. Journal of Computational and Applied Mathematics,2013,241:19-9.
- [8] JIAN J B,LIN H,JIANG X Z. A hybrid conjugate gradient method with descent property for unconstrained optimization[J]. Applied Mathematical Modelling,2014,39(3/4):1281-1290.
- [9] ABUBAKAR A B,KUMAM P,MALIK M,IBRAHIM A H. A hybrid conjugate gradient based approach for solving unconstrained optimization and motion control problems[J]. Mathematics and Computers in Simulation,2021,201:640-657.
- [10] SULAIMAN I M,MAMAT M,WAZIRI M Y,et al. The convergence properties of a new hybrid conjugate gradient parameter for unconstrained optimization models[C]//International Conference on Recent Trends in Applied Research (ICoRTAR), August 14-15,2020,Nigeria. Journal of physics:Conference series. Bristol:IOP Publishing,2021,1734(1):012012.
- [11] DELLADJIS,BELLOUFI M,SELLAMI B. New hybrid conjugate gradient method as a convex combination of FR and BA methods[J]. Journal of Information and Optimization Sciences,2021,42(3):591-602.
- [12] LIU Y F,ZHU Z B,ZHANG B X. Two sufficient descent three-term conjugate gradient methods for unconstrained optimization problems with applications in compressive sensing[J]. Journal of Applied Mathematics and Computing,2022,68(3):1787-1816.
- [13] YUAN G L,JIAN A,ZHANG M. A modified HZ conjugate gradient algorithm without gradient Lipschitz continuous condition for non convex functions[J]. Journal of Applied Mathematics and Computing,2022,68,4691-4712.
- [14] ANDREI N. An unconstrained optimization test functions collection[J]. Environmental Science & Technology,2008,10(1): 6552-6558.
- [15] DOLAN E D,MORJ J. Benchmarking optimization software with performance profiles[J]. Mathematical Programming,2002, 91(2):201-213.
- [16] CHAN R H,HO C W,NIKOLOVA M. Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization[J]. IEEE Transactions on Image Processing,2005,14(10):1479-1485.
- [17] CAI J F,CHAN R,Morini B. Minimization of an edge-preserving regularization functional by conjugate gradient type methods [C]//Image Processing Based on Partial Differential Equations:Proceedings of the International Conference on PDE-Based Image Processing and Related Inverse Problems,CMA,Oslo,August 8-12,2005. Heidelberg:Springer,2007:109-122.
- [18] WANG H,HADDAD R A. Adaptive median filters[J]. IEEE Transactions on Image Processing,1995,4(4):499-502.

Special Topics on Optimization Theory and Algorithm**A Class of Sufficiently Descending Hybrid CD-LS Conjugate Gradient Methods**

YIN Yuling, PENG Zaiyun, LIANG Renli, WANG Peng

(College of Mathematics and Statistics, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: In order to combine the good theoretical properties of Conjugate Descent (CD) method and the good numerical effect of Liu-Storey (LS) method, it is necessary to reduce the adverse effect of small steps on iteration, and make the decline of search direction independent of the selection of line search. By mixing the molecules of the CD method and the LS method, the gradient function was modified accordingly. The sufficient descent of direction is independent of the selection of line search and can be applied to a variety of line searches. Based on the Wolfe line search, the global convergence of the algorithm is proved. The results of 42 classes of unconstrained test function and image denoising show that based on the same termination conditions, the number of iterations and iteration time of the proposed algorithm are lower than those of the previous three types of conjugate gradient algorithms.

Keywords: mixed conjugate gradient method; sufficient descent; unconstrained optimization; image denoising

(责任编辑 黄 颖)