

一种基于正则化模型的 Dai-Liao 共轭梯度法

倪 艳, 刘泽显, 陈炫睿

(贵州大学 数学与统计学院, 贵阳 550025)

摘要: 给出一种基于正则化模型的 Dai-Liao 共轭梯度法. 首先, 通过极小化 3 次正则化模型, 得到新的 Dai-Liao 参数 t , 并在此基础上根据函数在迭代点附近的性质, 产生一个自适应的 Dai-Liao 参数; 其次, 结合改进的 Wolfe 线搜索, 提出一种基于正则化模型的 Dai-Liao 共轭梯度法; 最后, 证明该算法的搜索方向满足充分下降性, 并在一般假设下建立该算法的全局收敛性. 数值结果表明该算法有效.

关键词: 共轭梯度法; 正则化模型; Dai-Liao 共轭参数; 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:** 1671-5489(2024)03-0529-09

A Dai-Liao Conjugate Gradient Method Based on Regularization Model

NI Yan, LIU Zexian, CHEN Xuanrui

(School of Mathematics and Statistics, Guizhou University, Guiyang 550025, China)

Abstract: We gave a Dai-Liao conjugate gradient method based on regularization model. Firstly, a new Dai-Liao parameter t was obtained by minimizing the 3-degree regularization model, and based on this, an adaptive Dai-Liao parameter was generated according to the properties of the function near the iterative point. Secondly, combined with improved Wolfe line search, we proposed a Dai-Liao conjugate gradient method based on regularization model. Finally, we proved that the search direction of the proposed method satisfied sufficient descent, and established the global convergence of the proposed algorithm under the general assumption. Numerical results show that the proposed algorithm is effective.

Keywords: conjugate gradient method; regularization model; Dai-Liao conjugate parameter; sufficient descent; global convergence

0 引 言

考虑如下无约束优化问题:

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

其中 $f(x)$ 是连续可微函数. 共轭梯度法是求解问题(1)的有效迭代方法, 该方法的一般迭代格式为

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

收稿日期: 2023-10-07.

第一作者简介: 倪 艳(1995—), 女, 汉族, 硕士研究生, 从事最优化方法与应用的研究, E-mail: ny20222202@163.com. 通信

作者简介: 刘泽显(1984—), 男, 汉族, 博士, 副教授, 从事最优化方法与应用的研究, E-mail: liuzexian2008@163.com.

基金项目: 国家自然科学基金(批准号: 12261029)和贵州省自然科学基金一般项目(批准号: 黔科合基础-ZK[2022]—般 084).

其中: α_k 是步长; \mathbf{d}_k 是搜索方向, 由下式确定:

$$\mathbf{d}_k = \begin{cases} -\mathbf{g}_k, & k=0, \\ -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}, & k \geq 1, \end{cases} \quad (3)$$

这里 $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$ 是目标函数 $f(\mathbf{x})$ 在 \mathbf{x}_k 处的梯度, β_k 是共轭参数. 在共轭梯度法中, β_k 的选取方法至关重要, 目前已有很多研究结果, 如 HS^[1], FR^[2], PRP^[3-4] 和 DY^[5], 其数学表达式分别为

$$\beta^{\text{HS}} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}, \quad \beta^{\text{FR}} = \frac{\|\mathbf{g}_k\|^2}{\|\mathbf{g}_{k-1}\|^2}, \quad \beta^{\text{PRP}} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\|\mathbf{g}_{k-1}\|^2}, \quad \beta^{\text{DY}} = \frac{\|\mathbf{g}_k\|^2}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}},$$

其中 $\mathbf{y}_{k-1} = \mathbf{g}_k - \mathbf{g}_{k-1}$, $\|\cdot\|$ 为 Euclid 范数.

共轭梯度法由于具有迭代格式简单、所需存储少和数值效果好等优点, 因此备受关注. Dai 等^[6]提出了一种新的共轭条件 $\mathbf{d}_k^T \mathbf{y}_{k-1} = -t \mathbf{g}_k^T \mathbf{s}_{k-1}$, 其中 t 称为 Dai-Liao 参数, 并利用该条件给出了 Dai-Liao 共轭梯度法, 其共轭参数为

$$\beta_k^{\text{DL}} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - t \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}. \quad (4)$$

Hager 等^[7]在无记忆 BFGS(Broyden-Fletcher-Goldfarb-Shanno)方法的启发下, 提出了共轭参数

$$\beta_k^{\text{HZ}} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - 2 \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}. \quad (5)$$

Dai 等^[8]将 Perry^[9]和 Shanno^[10]提出的无记忆 BFGS 方向投影到一维子流形 $\text{Span}(-\mathbf{g}_k, \mathbf{d}_{k-1})$ 上, 提出了共轭参数

$$\beta_k^{\text{DK}} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - \left(\tau_k + \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} - \frac{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}{\|\mathbf{s}_{k-1}\|^2} \right) \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}, \quad (6)$$

其中 τ_k 是无记忆 BFGS 中的自调比参数, $\tau_k = \frac{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}{\|\mathbf{s}_{k-1}\|^2}$ 是 β_k^{DK} 最有效的选择. 共轭参数式(5)和式(6)

分别对应共轭梯度软件包 CG_DESCENT^[7]和 CGOPT^[8], 它们均可视为 $t=2 \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}$ 和 $t=\tau_k +$

$\frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} - \frac{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}{\|\mathbf{s}_{k-1}\|^2}$ 的 Dai-Liao 共轭梯度法.

对任意的 k , 若存在常数 $c>0$, 使得方向 \mathbf{d}_k 满足

$$\mathbf{d}_k^T \mathbf{g}_k \leq -c \|\mathbf{g}_k\|^2, \quad (7)$$

则称该方向 \mathbf{d}_k 满足充分下降性.

共轭参数对共轭梯度法的数值效果有很大影响, 合适的参数不仅使算法在理论上能取得良好的性质, 而且数值性能也较好. 因此, 如何寻找最佳的共轭参数是一个重要课题.

正则化算法是求解无约束优化问题非常有效的算法. 本文首先通过极小化正则化模型得到新的 Dai-Liao 参数, 在此基础上根据目标函数在迭代点附近的性质设计自适应的 Dai-Liao 参数; 其次结合改进的 Wolfe 线搜索, 提出一种基于正则化模型的 Dai-Liao 共轭梯度法; 最后分析算法搜索方向的充分下降性, 并在一般假设下建立算法的全局收敛性. 数值结果表明, 针对 CUTEst 库^[11]中的测试函数, 该算法有效.

1 算法设计及其充分下降性证明

下面通过极小化 3 次正则化模型, 给出一种新的 Dai-Liao 参数, 并根据函数的性质设计自适应的 Dai-Liao 参数, 证明搜索方向的充分下降性.

1.1 算法设计

考虑如下 3 次正则化模型:

$$\min_{\mathbf{d}_k \in \mathbb{R}^n} m(\mathbf{d}_k) = f_k + \mathbf{g}_k^T \mathbf{d}_k + \frac{1}{2} \mathbf{d}_k^T \mathbf{B}_k \mathbf{d}_k + \frac{\sigma_k}{3} \|\mathbf{d}_k\|_{A_k}^3, \quad (8)$$

其中 $\|\mathbf{d}_k\|_{A_k} = \sqrt{\mathbf{d}_k^T \mathbf{A}_k \mathbf{d}_k}$, \mathbf{A}_k 是对称正定矩阵, \mathbf{B}_k 是实对称矩阵, σ_k 是正则化参数.

将 $\mathbf{d}_k = -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}$ 代入问题(8), 对 β_k 进行求导并令其导数为 0, 可得

$$\beta_k = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1} - \mathbf{g}_k^T \mathbf{s}_{k-1} + \gamma_k \mathbf{g}_k^T \mathbf{A}_k \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1} + \gamma_k \mathbf{d}_{k-1}^T \mathbf{A}_k \mathbf{s}_{k-1}}, \tag{9}$$

其中 $\gamma_k = \sigma_k \|\mathbf{d}_k\|_{\mathbf{A}_k}$. 若令 $\mathbf{A}_k = \mathbf{B}_k$, 且 \mathbf{B}_k 满足割线方程 $\mathbf{B}_k \mathbf{s}_{k-1} = \mathbf{y}_{k-1}$, 则式(9)可表示为

$$\beta_k = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - \frac{1}{1 + \gamma_k} \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}, \tag{10}$$

可将式(10)视为 $t = \frac{1}{1 + \gamma_k}$ 的 Dai-Liao 共轭参数, 而 γ_k 是正则化参数 σ_k 与 $\|\mathbf{d}_k\|_{\mathbf{A}_k}$ 的乘积, 下面分别对这两个值进行选取和分析.

1.1.1 正则化参数 σ_k 的选取

参考文献[12]中的 σ_k 值, 当 $\mathbf{A}_k = \mathbf{B}_k$ 时,

$$\sigma_k = \frac{3 \left| f_{k-1} - f_k + \mathbf{g}_k^T \mathbf{s}_{k-1} - \frac{1}{2} \mathbf{s}_{k-1}^T \mathbf{y}_{k-1} \right|}{(\mathbf{s}_{k-1}^T \mathbf{y}_{k-1})^{3/2}}. \tag{11}$$

1.1.2 $\|\mathbf{d}_k\|_{\mathbf{A}_k}$ 的分析与计算

首先考虑 2-范数下的 3 次正则化子问题:

$$\min_{\mathbf{p}_k \in \mathbb{R}^n} m(\mathbf{p}_k) = f_k + \mathbf{g}_k^T \mathbf{p}_k + \frac{1}{2} \mathbf{p}_k^T \mathbf{B}_k \mathbf{p}_k + \frac{\sigma_k}{3} \|\mathbf{p}_k\|^3, \tag{12}$$

文献[13]中定理 1.1 已给出问题(12)的解.

定理 1^[13] 点 \mathbf{p}_k^* 是问题(12)的一个全局极小值点, 当且仅当

$$(\mathbf{B}_k + \sigma_k \|\mathbf{p}_k^*\| \mathbf{I}) \mathbf{p}_k^* = -\mathbf{g}_k,$$

且矩阵 $\mathbf{B}_k + \sigma_k \|\mathbf{p}_k^*\| \mathbf{I}$ 是半正定的.

令 $\mathbf{p}_k = \mathbf{A}_k^{-1/2} \mathbf{d}_k$, 即 $\mathbf{d}_k = \mathbf{A}_k^{-1/2} \mathbf{p}_k$, 则问题(8)可表示为

$$\min_{\mathbf{p}_k \in \mathbb{R}^n} m(\mathbf{p}_k) = f_k + \mathbf{g}_k^T (\mathbf{A}_k^{-1/2} \mathbf{p}_k) + \frac{1}{2} \mathbf{p}_k^T \mathbf{A}_k^{-1/2} \mathbf{B}_k \mathbf{A}_k^{-1/2} \mathbf{p}_k + \frac{\sigma_k}{3} \|\mathbf{p}_k\|^3. \tag{13}$$

由定理 1 可知, \mathbf{p}_k^* 是问题(13)的一个全局极小值点当且仅当

$$(\mathbf{A}_k^{-1/2} \mathbf{B}_k \mathbf{A}_k^{-1/2} + \sigma_k \|\mathbf{p}_k^*\| \mathbf{I}) \mathbf{p}_k^* = -\mathbf{A}_k^{-1/2} \mathbf{g}_k,$$

且矩阵 $\mathbf{A}_k^{-1/2} \mathbf{B}_k \mathbf{A}_k^{-1/2} + \sigma_k \|\mathbf{p}_k^*\| \mathbf{I}$ 是半正定的. 可知矩阵 $\mathbf{A}_k^{-1/2} \mathbf{B}_k \mathbf{A}_k^{-1/2}$ 是一个实对称矩阵, 从而存在一个正交矩阵 $\mathbf{U} \in \mathbb{R}^{n \times n}$, 使得 $\mathbf{U}^T (\mathbf{A}_k^{-1/2} \mathbf{B}_k \mathbf{A}_k^{-1/2}) \mathbf{U} = \mathbf{V}$, 其中 $\mathbf{V} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 且 $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ 为矩阵 $\mathbf{A}_k^{-1/2} \mathbf{B}_k \mathbf{A}_k^{-1/2}$ 的特征值, 于是有

$$\begin{aligned} \mathbf{p}_k^* &= -(\mathbf{A}_k^{-1/2} \mathbf{B}_k \mathbf{A}_k^{-1/2} + \sigma_k \|\mathbf{p}_k^*\| \mathbf{I})^{-1} \mathbf{A}_k^{-1/2} \mathbf{g}_k = -(\mathbf{U} \mathbf{V} \mathbf{U}^T + \sigma_k \|\mathbf{p}_k^*\| \mathbf{I})^{-1} \mathbf{A}_k^{-1/2} \mathbf{g}_k = \\ &= -[\mathbf{U} (\mathbf{V} + \sigma_k \|\mathbf{p}_k^*\| \mathbf{I}) \mathbf{U}^T]^{-1} \mathbf{A}_k^{-1/2} \mathbf{g}_k = -\mathbf{U} (\mathbf{V} + \sigma_k \|\mathbf{p}_k^*\| \mathbf{I})^{-1} \mathbf{U}^T \mathbf{A}_k^{-1/2} \mathbf{g}_k. \end{aligned} \tag{14}$$

引入一个向量 $\mathbf{a} \in \mathbb{R}^n$, 使得 $\mathbf{p}_k^* = \mathbf{U} \mathbf{a}$. 令 $z = \|\mathbf{p}_k^*\|$, $\mathbf{b} = \mathbf{U}^T \mathbf{A}_k^{-1/2} \mathbf{g}_k$, 若在式(14)两端同时左乘 \mathbf{U}^T , 则可推出 $a_i = \frac{-b_i}{\lambda_i + \sigma_k z}$, 其中 a_i, b_i 分别是向量 \mathbf{a}, \mathbf{b} 的分量, 因此

$$z^2 = \mathbf{p}_k^T \mathbf{p}_k = \mathbf{a}^T \mathbf{a} = \sum_{i=1}^n \frac{b_i^2}{(\lambda_i + \sigma_k z)^2}. \tag{15}$$

定义函数 $s(z) = \sum_{i=1}^n \frac{b_i^2}{(\lambda_i + \sigma_k z)^2} - z^2$, 则 $s'(z) = \sum_{i=1}^n \frac{-2\sigma_k b_i^2 (\lambda_i + \sigma_k z)}{(\lambda_i + \sigma_k z)^4} - 2z$, 由于 $z > 0$ 且 $\sigma_k > 0$,

易推出 $s'(z) < 0$, 则函数 $s(z)$ 在区间 $[0, +\infty)$ 上是单调递减的. 当 $b \neq 0$ 时, 有 $s(0) = \sum_{i=1}^n \frac{b_i^2}{\lambda_i^2} > 0$, 而 $\lim_{z \rightarrow +\infty} s(z) = -\infty$, 所以函数 $s(z)$ 在区间 $[0, +\infty)$ 上必有一个正解, 从而方程(15)在区间 $[0, +\infty)$ 上必有一个正根; 当 $b = 0$ 时, $z = 0$ 是式(14)的唯一解.

由上述内容可知 $\|\mathbf{d}_k\|_{\mathbf{A}_k} = \sqrt{\mathbf{d}_k^T \mathbf{A}_k \mathbf{d}_k} = \|\mathbf{A}_k^{1/2} \mathbf{d}_k\| = \|\mathbf{p}_k\| = z$ 在区间 $[0, +\infty)$ 上必存在一个非负解. 对于问题(8), 若直接对 \mathbf{d}_k 求导, 并令其导数为 0, 则 \mathbf{d}_k^* 是问题(8)的全局极小点当且仅当

$(\mathbf{B}_k + \sigma_k \|\mathbf{d}_k^*\|_{\mathbf{A}_k} \mathbf{I})\mathbf{d}_k^* = -\mathbf{g}_k$, 且 $\mathbf{B}_k + \sigma_k \|\mathbf{d}_k^*\|_{\mathbf{A}_k} \mathbf{I}$ 是半正定的.

由于 $\|\mathbf{d}_k^*\|_{\mathbf{A}_k} = z^*$, 因此上述结论可表述为: \mathbf{d}_k^* 是问题(8)的全局极小点当且仅当 $(\mathbf{B}_k + \sigma_k z^* \mathbf{I})\mathbf{d}_k^* = -\mathbf{g}_k$, 且 $\mathbf{B}_k + \sigma_k z^* \mathbf{I}$ 是半正定的, 其中 z^* 是方程(15)的唯一非负解.

本文主要考虑一种特殊情况, 即 \mathbf{B}_k 是一个对称正定矩阵, 且 $\mathbf{A}_k = \mathbf{B}_k$, $\mathbf{B}_k \mathbf{s}_{k-1} = \mathbf{y}_{k-1}$. 显然 $\mathbf{B}_k + \sigma_k z^* \mathbf{I}$ 是一个正定矩阵, 将 $\mathbf{d}_k = -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}$ 代入问题(8), 则问题(8)可表示为

$$\min_{\beta_k \in \mathbb{R}} m(\beta_k) = f_k + \left(\begin{array}{c} \|\mathbf{g}_k\|^2 \\ \mathbf{g}_k^T \mathbf{s}_{k-1} \end{array} \right)^T \begin{pmatrix} -1 \\ \beta_k \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -1 \\ \beta_k \end{pmatrix}^T \mathbf{H}_k \begin{pmatrix} -1 \\ \beta_k \end{pmatrix} + \frac{\sigma_k}{3} \left\| \begin{pmatrix} -1 \\ \beta_k \end{pmatrix} \right\|_{\mathbf{H}_k}^3, \quad (16)$$

其中 $\beta_k = \frac{\beta_k}{\alpha_{k-1}}$, $\mathbf{H}_k = \begin{pmatrix} \rho_k & \mathbf{g}_k^T \mathbf{y}_{k-1} \\ \mathbf{g}_k^T \mathbf{y}_{k-1} & \mathbf{s}_{k-1}^T \mathbf{y}_{k-1} \end{pmatrix}$, $\rho_k = \mathbf{g}_k^T \mathbf{B}_k \mathbf{g}_k$.

推论 1^[12] 若 \mathbf{B}_k 是一个正定矩阵, $\mathbf{A}_k = \mathbf{B}_k$, 则 $\mathbf{d}_k^* = \frac{-1}{1 + \sigma_k z^*} \mathbf{B}_k^{-1} \mathbf{g}_k$ 是问题(8)的全局极小点, z^* 是方程 $\sigma_k z^2 + z - \sqrt{\mathbf{g}_k^T \mathbf{B}_k^{-1} \mathbf{g}_k} = 0$ 的唯一非负解.

根据推论 1 可知, 在问题(16)中, z^* 是方程 $\sigma_k z^2 + z - \sqrt{\left(\begin{array}{c} \|\mathbf{g}_k\|^2 \\ \mathbf{g}_k^T \mathbf{s}_{k-1} \end{array} \right)^T \mathbf{H}_k^{-1} \left(\begin{array}{c} \|\mathbf{g}_k\|^2 \\ \mathbf{g}_k^T \mathbf{s}_{k-1} \end{array} \right)} = 0$ 的唯一非

负解, 令 $q = \sqrt{\left(\begin{array}{c} \|\mathbf{g}_k\|^2 \\ \mathbf{g}_k^T \mathbf{s}_{k-1} \end{array} \right)^T \mathbf{H}_k^{-1} \left(\begin{array}{c} \|\mathbf{g}_k\|^2 \\ \mathbf{g}_k^T \mathbf{s}_{k-1} \end{array} \right)}$, 则有

$$z^* = \frac{2q}{1 + \sqrt{1 + 4\sigma_k q}}, \quad (17)$$

其中选取 ρ_k 为文献[14]中的 $\rho_k = \rho_k^{\text{BBCG3}} = \frac{3}{2} \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \|\mathbf{g}_k\|^2$, 于是可求出 Dai-Liao 参数

$$t_k = \frac{1}{1 + \gamma_k},$$

这里 $\gamma_k = \sigma_k \|\mathbf{d}_k\|_{\mathbf{A}_k}$, σ_k 和 $\|\mathbf{d}_k\|_{\mathbf{A}_k}$ 的值分别通过式(11)和式(17)求解.

为提升数值效果, 将参数 t_k 限制在 $\frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}$ 和 $2 \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}$ 之间, 即

$$\bar{t}_k = \min \left\{ \max \left\{ t_k, \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \right\}, 2 \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \right\}. \quad (18)$$

文献[15]引入了参数 θ_k , 用于描述函数 $f(\mathbf{x})$ 与二次函数的接近程度, θ_k 定义为

$$\theta_k = \left| \frac{2(f_k - f_{k-1} + \mathbf{g}_k^T \mathbf{s}_{k-1})}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} - 1 \right|, \quad (19)$$

如果参数 θ_k 满足

$$\theta_k \leq c_1 \text{ 或 } (\theta_k \leq c_2 \text{ 且 } \theta_{k-1} \leq c_2), \quad (20)$$

则表示目标函数 $f(\mathbf{x})$ 在 \mathbf{x}_k 附近可能非常接近二次函数, 此时可选择 Dai-Liao 参数 $t_k = \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}$,

否则选择参数 \bar{t}_k , 于是本文产生了一个参数 t_k 的自适应选择, 即

$$\check{t}_k = \begin{cases} \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}, & \text{满足式(20),} \\ \bar{t}_k, & \text{其他.} \end{cases} \quad (21)$$

结合式(18)和式(21), 不妨令 $\check{t}_k = a \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}$, 其中 $a \in [1, 2]$, 则有

$$\beta_k = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - a \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}. \quad (22)$$

与 Dai 等^[8]提出的 CGOPT 类似, 给出以下共轭参数 β_k 的截断:

$$\bar{\beta}_k = \max \left\{ \beta_k, \eta \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\|\mathbf{d}_{k-1}\|^2} \right\}, \quad \eta \in [0, 1), \quad (23)$$

其中 β_k 由式(22)确定.

1.2 搜索方向的充分下降性证明

引理 1 如果 $\mathbf{s}_{k-1}^T \mathbf{y}_{k-1} > 0$, 则对共轭梯度法(2)-(3)-(23), 存在 $c > 0$, 使得

$$\mathbf{d}_k^T \mathbf{g}_k \leq -c \|\mathbf{g}_k\|^2. \tag{24}$$

证明: 由文献[8]可知, $\mathbf{g}_k^T(-\mathbf{g}_k + \beta_k^{\text{DK}} \mathbf{d}_{k-1}) \leq -\frac{3}{4} \|\mathbf{g}_k\|^2$, 由式(22)可推出

$$\begin{aligned} \mathbf{g}_k^T(-\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}) &= \mathbf{g}_k^T \left[-\mathbf{g}_k + \left(\frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - a \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} \right) \mathbf{d}_{k-1} \right] \leq \\ &= -\|\mathbf{g}_k\|^2 + \frac{\mathbf{g}_k^T \mathbf{y}_{k-1} \mathbf{g}_k^T \mathbf{d}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - \frac{\|\mathbf{y}_{k-1}\|^2 (\mathbf{g}_k^T \mathbf{d}_{k-1})^2}{(\mathbf{d}_{k-1}^T \mathbf{y}_{k-1})^2} \leq \\ &= \mathbf{g}_k^T(-\mathbf{g}_k + \beta_k^{\text{DK}} \mathbf{d}_{k-1}) \leq -\frac{3}{4} \|\mathbf{g}_k\|^2. \end{aligned}$$

此外, 对于 $\hat{\beta}_k = \eta \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\|\mathbf{d}_{k-1}\|^2}$, 由 Cauchy-Schwarz 不等式, 有

$$\mathbf{g}_k^T \mathbf{d}_k = -\|\mathbf{g}_k\|^2 + \eta \frac{(\mathbf{g}_k^T \mathbf{d}_{k-1})^2}{\|\mathbf{d}_{k-1}\|^2} \leq -\|\mathbf{g}_k\|^2 + \eta \frac{\|\mathbf{g}_k\|^2 \|\mathbf{d}_{k-1}\|^2}{\|\mathbf{d}_{k-1}\|^2} = -(1-\eta) \|\mathbf{g}_k\|^2.$$

综上所述, 共轭梯度法(2)-(3)-(23)满足充分下降性, 其中 $c = \min\left\{\frac{3}{4}, 1-\eta\right\}$.

2 算法框架与收敛性分析

为避免数值误差, 保证该算法的全局收敛性, 本文使用文献[8]中提出的改进 Wolfe 线搜索, 找到一个步长, 使其满足

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + \min\{\epsilon |\mathbf{g}_k^T \mathbf{d}_k|, \delta \alpha_k \mathbf{g}_k^T \mathbf{d}_k + \eta_k\}, \tag{25}$$

$$\mathbf{g}(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k \geq \sigma \mathbf{g}_k^T \mathbf{d}_k, \tag{26}$$

其中 $0 < \delta < \sigma < 1$, $\epsilon > 0$, $\eta_k > 0$ 满足 $\sum_{k \geq 1} \eta_k < +\infty$.

为加快算法的速度, 本文采用一种自适应重启技术^[8]:

$$r_{k-1} = \frac{2(f_k - f_{k-1})}{\mathbf{g}_k^T \mathbf{s}_{k-1} + \mathbf{g}_{k-1}^T \mathbf{s}_{k-1}}, \tag{27}$$

如果 r_{k-1} 接近 1, 则说明线搜索函数接近某个二次函数. 若在连续的多次迭代中都有 r_k 接近 1, 则以最速下降方向重启算法.

算法 1 基于正则化模型的 Dai-Liao 共轭梯度法框架.

步骤 1) 给定初始值 $\mathbf{x}_0 \in \mathbb{R}^n$, $\epsilon > 0$, ξ, δ, σ 满足 $0 < \delta < \sigma < 1$;

步骤 2) 如果 $\|\mathbf{g}_k\| \leq \epsilon$, 则算法停止; 否则, $\mathbf{d}_0 = -\mathbf{g}_0$, 令 $\text{IterRestart} := 0, \text{IterQuad} := 0$;

步骤 3) 利用改进的 Wolfe 线搜索(25),(26)计算步长 α_k ;

步骤 4) 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$, 若 $\|\mathbf{g}_k\|_\infty \leq \epsilon$, 则算法停止; 令 $\text{IterRestart} := \text{IterRestart} + 1$, 利用式(27)计算 r_k , 若 $|r_k - 1| \leq \xi$, 则 $\text{IterQuad} := \text{IterQuad} + 1$; 否则, $\text{IterQuad} := 0$;

步骤 5) 如果 $\text{IterRestart} = \text{MaxRestart}$ 或者 $\text{IterQuad} = \text{MinQuad}$ 且 $\text{IterQuad} \neq \text{IterRestart}$, 则令 $\mathbf{d}_k = -\mathbf{g}_k$, $\text{IterRestart} := 0, \text{IterQuad} := 0, k := k + 1$, 转步骤 3);

步骤 6) 利用式(23)计算 $\bar{\beta}_k$, 利用式(2),(3)计算搜索方向 \mathbf{d}_k , 然后令 $k := k + 1$, 转步骤 3).

假设:

(H₁) 目标函数 f 在 \mathbb{R}^n 上连续可微;

(H₂) 水平集 $\{\mathbf{x} \mid f(\mathbf{x}) \leq f(\mathbf{x}_0) + \sum_{k \geq 0} \eta_k\}$ 有界;

(H₃) 梯度 \mathbf{g} 在 \mathbb{R}^n 上是 Lipschitz 连续的, 即存在常数 $L > 0$, 使得

$$\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

定理 2 设 f 满足假设 $(H_1) \sim (H_3)$, 考虑共轭梯度法(2)-(3)-(23), 利用改进的 Wolfe 线搜索(25),(26)计算步长 α_k , 如果函数 f 是一致凸的, 即存在常数 $\mu > 0$, 使得 $(\mathbf{g}(x) - \mathbf{g}(y))^T(x - y) \geq \mu \|\mathbf{x} - \mathbf{y}\|^2 (\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n)$, 则有

$$\lim_{k \rightarrow +\infty} \|\mathbf{g}_k\| = 0. \quad (28)$$

证明: 由式(25),(26)和假设 $(H_1) \sim (H_3)$ 可知

$$\sum_{k=0}^{+\infty} -\alpha_k \mathbf{g}_k^T \mathbf{d}_k < +\infty, \quad \alpha_k \geq \frac{-(1-\sigma) \mathbf{g}_k^T \mathbf{d}_k}{L \|\mathbf{d}_k\|^2}, \quad (29)$$

结合充分下降性和式(29)可推出

$$\sum_{k=0}^{+\infty} \frac{\|\mathbf{g}_k\|^4}{\|\mathbf{d}_k\|^2} < +\infty. \quad (30)$$

由于函数是一致凸的, 结合式(23), 有

$$\begin{aligned} |\bar{\beta}_k| &= \left| \max \left\{ \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - a \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}}, \eta \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\|\mathbf{d}_{k-1}\|^2} \right\} \right| = \\ &= \max \left\{ \left| \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} \right| + a \left| \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} \right|, \eta \left| \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\|\mathbf{d}_{k-1}\|^2} \right| \right\} \leq \\ &= \max \left\{ \frac{\|\mathbf{g}_k\| \|\mathbf{y}_{k-1}\|}{\mu \alpha_{k-1} \|\mathbf{d}_{k-1}\|^2} + a \frac{\|\mathbf{y}_{k-1}\|^2 \|\mathbf{g}_k\| \|\mathbf{d}_{k-1}\|}{\mu^2 \alpha_{k-1}^2 \|\mathbf{d}_{k-1}\|^4}, \eta \frac{\|\mathbf{g}_k\| \|\mathbf{d}_{k-1}\|}{\|\mathbf{d}_{k-1}\|^2} \right\} \leq \\ &= \max \left\{ \frac{L \alpha_{k-1} \|\mathbf{g}_k\| \|\mathbf{d}_{k-1}\|}{\mu \alpha_{k-1} \|\mathbf{d}_{k-1}\|^2} + a \frac{L^2 \alpha_{k-1}^2 \|\mathbf{d}_{k-1}\|^2 \|\mathbf{g}_k\| \|\mathbf{d}_{k-1}\|}{\mu^2 \alpha_{k-1}^2 \|\mathbf{d}_{k-1}\|^4}, \eta \frac{\|\mathbf{g}_k\| \|\mathbf{d}_{k-1}\|}{\|\mathbf{d}_{k-1}\|^2} \right\} = \\ &= \max \left\{ \frac{L}{\mu} + \frac{aL^2}{\mu^2}, \eta \right\} \frac{\|\mathbf{g}_k\|}{\|\mathbf{d}_{k-1}\|}, \end{aligned}$$

从而可推出

$$\|\mathbf{d}_k\| \leq \|\mathbf{g}_k\| + \max \left\{ \frac{L}{\mu} + \frac{aL^2}{\mu^2}, \eta \right\} \frac{\|\mathbf{g}_k\|}{\|\mathbf{d}_{k-1}\|} \|\mathbf{d}_{k-1}\| = \left[1 + \max \left\{ \frac{L}{\mu} + \frac{aL^2}{\mu^2}, \eta \right\} \right] \|\mathbf{g}_k\|.$$

令 $M = 1 + \max \left\{ \frac{L}{\mu} + \frac{aL^2}{\mu^2}, \eta \right\}$, 结合式(30), 可推出 $\sum_{k=0}^{+\infty} \|\mathbf{g}_k\|^2 < +\infty$. 因此式(28)成立. 证毕.

引理 2 设 f 满足假设 $(H_1) \sim (H_3)$, 考虑共轭梯度法(2)-(3)-(23), 利用改进的 Wolfe 线搜索式(25)和式(26)计算步长 α_k , 如果对任意的 $k \geq 1$, 有 $\|\mathbf{g}_k\| \geq \gamma_1$, 令 $\mathbf{u}_k = \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|}$, 则有

$$\sum_{k=0}^{+\infty} \|\mathbf{u}_k - \mathbf{u}_{k-1}\|^2 < +\infty.$$

证明: 将 $\bar{\beta}_k$ 分解成如下两部分:

$$\beta_k^{(1)} = \max \left\{ \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - a \frac{\|\mathbf{y}_{k-1}\|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - \eta \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\|\mathbf{d}_{k-1}\|^2}, 0 \right\}, \quad (31)$$

$$\beta_k^{(2)} = \eta \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\|\mathbf{d}_{k-1}\|^2}. \quad (32)$$

定义

$$\omega_k^{(1)} = \frac{-\mathbf{g}_k + \beta_k^{(2)} \mathbf{d}_{k-1}}{\|\mathbf{d}_k\|}, \quad \omega_k^{(2)} = \frac{\beta_k^{(1)} \mathbf{d}_{k-1}}{\|\mathbf{d}_k\|}, \quad (33)$$

则有

$$\mathbf{u}_k = \frac{\mathbf{d}_k}{\|\mathbf{d}_k\|} = \frac{-\mathbf{g}_k + (\beta_k^{(1)} + \beta_k^{(2)}) \mathbf{d}_{k-1}}{\|\mathbf{d}_k\|} = \omega_k^{(1)} + \omega_k^{(2)} \mathbf{u}_{k-1},$$

再利用 $\|\mathbf{u}_k\| = \|\mathbf{u}_{k-1}\| = 1$, 可推出

$$\|\omega_k^{(1)}\| = \|\mathbf{u}_k - \omega_k^{(2)} \mathbf{u}_{k-1}\| = \|\omega_k^{(2)} \mathbf{u}_k - \mathbf{u}_{k-1}\|.$$

于是

$$\|\mathbf{u}_k - \mathbf{u}_{k-1}\| \leq \|(1 + \omega_k^{(2)}) \mathbf{u}_k - (1 + \omega_k^{(2)}) \mathbf{u}_{k-1}\| \leq$$

$$\| \mathbf{u}_k - \boldsymbol{\omega}_k^{(2)} \mathbf{u}_{k-1} \| + \| \boldsymbol{\omega}_k^{(2)} \mathbf{u}_k - \mathbf{u}_{k-1} \| = 2 \| \boldsymbol{\omega}_k^{(1)} \| . \tag{34}$$

由式(32)可知

$$\| \boldsymbol{\omega}_k^{(1)} \| \| \mathbf{d}_k \| = \| -\mathbf{g}_k + \beta_k^{(2)} \mathbf{d}_{k-1} \| \leq \| \mathbf{g}_k \| + |\beta_k^{(2)}| \| \mathbf{d}_{k-1} \| \leq (1 + \eta) \| \mathbf{g}_k \| . \tag{35}$$

综合不等式(34),(35), 有

$$\| \mathbf{u}_k - \mathbf{u}_{k-1} \| \leq 2 \| \boldsymbol{\omega}_k^{(1)} \| \leq 2(1 + \eta) \frac{\| \mathbf{g}_k \|}{\| \mathbf{d}_k \|} .$$

由于 $\| \mathbf{g}_k \| \geq \gamma_1$, 故结合式(30)可得

$$\begin{aligned} \sum_{k=0}^{+\infty} \| \mathbf{u}_k - \mathbf{u}_{k-1} \|^2 &\leq 4(1 + \eta)^2 \sum_{k=0}^{+\infty} \frac{\| \mathbf{g}_k \|^2}{\| \mathbf{d}_k \|^2} = 4(1 + \eta)^2 \sum_{k=0}^{+\infty} \frac{\| \mathbf{g}_k \|^4}{\| \mathbf{d}_k \|^2} \frac{1}{\| \mathbf{g}_k \|^2} \leq \\ &\frac{4(1 + \eta)^2}{\gamma_1^2} \sum_{k=0}^{+\infty} \frac{\| \mathbf{g}_k \|^4}{\| \mathbf{d}_k \|^2} < +\infty, \end{aligned} \tag{36}$$

证毕.

定理 3 如果 f 满足假设(H₁)~(H₃), $\{\mathbf{x}_k\}$ 由算法 1 生成, 则有

$$\liminf_{k \rightarrow +\infty} \| \mathbf{g}_k \| = 0. \tag{37}$$

证明: 假设对任意的 $k \geq 1$, 有 $\| \mathbf{g}_k \| \geq \gamma_1$, 其中 $\gamma_1 > 0$. 根据假设(H₂)中水平集的有界性可知, 存在正常数 m , 使得 $\| \mathbf{x}_k \| \leq m (\forall k \geq 0)$, 因此有 $\| \mathbf{s}_k \| \leq 2m (\forall k \geq 0)$. 由 Wolfe 线搜索条件(26)和充分下降性可推出

$$\mathbf{d}_k^T \mathbf{y}_k = \mathbf{g}_{k+1}^T \mathbf{d}_k - \mathbf{g}_k^T \mathbf{d}_k \geq -(1 - \sigma) \mathbf{g}_k^T \mathbf{d}_k \geq c(1 - \sigma) \| \mathbf{g}_k \|^2 \geq c(1 - \sigma) \gamma_1^2, \tag{38}$$

且有

$$\mathbf{g}_{k+1}^T \mathbf{d}_k \geq \sigma \mathbf{g}_k^T \mathbf{d}_k = \sigma \mathbf{g}_{k+1}^T \mathbf{d}_k - \sigma \mathbf{d}_k^T \mathbf{y}_k, \tag{39}$$

$$\mathbf{g}_{k+1}^T \mathbf{d}_k = \mathbf{d}_k^T \mathbf{y}_k + \mathbf{g}_k^T \mathbf{d}_k < \mathbf{d}_k^T \mathbf{y}_k. \tag{40}$$

由式(39),(40), 可得

$$\frac{|\mathbf{g}_k^T \mathbf{d}_{k-1}|}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} \leq \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\}. \tag{41}$$

由假设(H₃)知存在一个正常数 γ_2 , 使得 $\| \mathbf{g}_k \| \leq \gamma_2$, 从而可得

$$\begin{aligned} |\beta_k| &= \left| \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} - a \frac{\| \mathbf{y}_{k-1} \|^2}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} \right| \leq \frac{1}{|\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}|} \left(|\mathbf{g}_k^T \mathbf{y}_{k-1}| + a \| \mathbf{y}_{k-1} \|^2 \left| \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} \right| \right) \leq \\ &\frac{1}{c\gamma_1^2(1 - \sigma)} \left(L \| \mathbf{g}_k \| \| \mathbf{s}_{k-1} \| + aL^2 \| \mathbf{s}_{k-1} \|^2 \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\} \right) \leq \\ &\frac{1}{c\gamma_1^2(1 - \sigma)} \left(L\gamma_2 + 2amL^2 \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\} \right) \| \mathbf{s}_{k-1} \| = N \| \mathbf{s}_{k-1} \|, \end{aligned}$$

其中 $N = \frac{1}{c\gamma_1^2(1 - \sigma)} \left(L\gamma_2 + 2amL^2 \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\} \right)$, 可知 $N > 0$. 令 $b = 2Nm$, $\tau = \frac{1}{2N^2m}$, 则有

$$|\beta_k| \leq b, \quad \| \mathbf{s}_{k-1} \| \leq 2Nm = b, \quad \| \mathbf{s}_{k-1} \| \leq \tau,$$

于是 $|\beta_k| \leq N \frac{1}{2N^2m} = \frac{1}{2Nm} = b$, 说明 β_k 具有文献[16]中的性质(*). 根据 $\| \mathbf{g}_k \| \geq \gamma_1$, $\sum_{k=0}^{+\infty} \frac{(\mathbf{g}_k^T \mathbf{d}_k)^2}{\| \mathbf{d}_k \|^2} < +\infty$ 和充分下降性可知

$$\| \mathbf{d}_k \| \rightarrow +\infty, \tag{42}$$

从而可推出

$$\| \mathbf{d}_k \| = \| -\mathbf{g}_k + \beta_k^{(2)} \mathbf{d}_{k-1} \| \leq (1 + \eta) \| \mathbf{g}_k \| \leq (1 + \eta) \gamma_2. \tag{43}$$

显然式(43)与式(42)矛盾, 故式(37)成立. 证毕.

3 数值实验

本文提出的算法 1 是通过修改 CGOPT 软件包(http://coa.amss.ac.cn/wordpress/?page_id1/421)实现的. 为检验算法 1 的实际应用效果, 本文对 CUTEst^[11]中的 113 个问题进行测试, 并将测试

的结果与 NEWCG^[17] 和 CGOPT 进行比较. 实验过程中的参数选取如下: $\epsilon=10^{-6}$, $\delta=0.1$, $\sigma=0.9$, $\xi=10^{-3}$, $c_1=10^{-4}$, $c_2=1.080$. 当迭代次数超过 200 000 或者 $\|g_k\|_\infty \leq 10^{-6}$ 时, 所有测试算法终止. 实验中, 分别对迭代次数、目标函数的计算次数、梯度的计算次数和 CPU 运行时间这 4 个重要指标进行测试, 并使用性能图^[18]对测试结果进行比较.

下面以迭代次数的性能图为例对性能图的横轴和纵轴做出解释. 记 \bar{P} 为 n_p 个测试问题组成的集合, S 为测试的优化算法组成的集合, $I_{p,s}$ 为优化算法 $s \in S$ 求解问题 $p \in \bar{P}$ 所需的迭代次数. 定义性能比 $r_{p,s} = \frac{I_{p,s}}{I_p^*}$, 其中 $I_p^* = \min\{I_{p,s} : s \in S\}$, 显然有 $r_{p,s} \geq 1$. 若算法 s 求解问题 p 失败, 则将 $r_{p,s}$ 设为一个很大的正数. 定义

$$P(\tau) = \frac{\text{size}\{p \in \bar{P} : r_{p,s} \leq \tau\}}{n_p},$$

其中 $\tau \geq 1$, $\text{size}\{\cdot\}$ 表示集合元素的个数. 记 $I^* = \{I_1^*, I_2^*, \dots, I_{n_p}^*\}$ 为 S 中的优化算法求解每个问题所需的最少迭代次数组成的集合, 则 $P(\tau)$ 表示算法 s 在集合 I^* 的 τ 倍内能解的问题数占问题总数的比率, 于是在性能图中最高的曲线表示该测试算法在最少迭代次数集的 τ 倍内能求解的问题最多.

图 1~图 4 分别为测试算法的迭代次数、函数值计算次数、梯度计算次数和 CPU 时间的性能图. 由图 1~图 4 可见, 在迭代次数、目标函数的计算次数、梯度的计算次数和 CPU 运行时间指标上, 算法 1 都比 CGOPT 和 NEWCG 更有优势. 特别地, 由图 1~图 3 可见算法 1 的明显优势. 数值实验结果表明, 针对 CUTEst 库^[11]中的测试函数, 算法 1 优于 CGOPT 和 NEWCG.

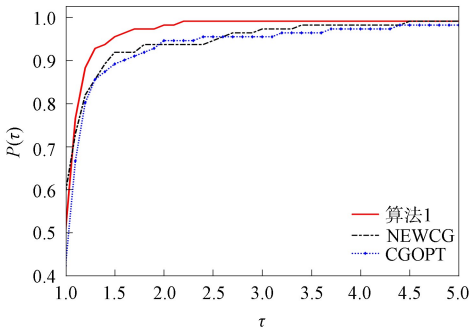


图 1 不同方法的迭代次数比较
Fig. 1 Comparison of iterations for different methods

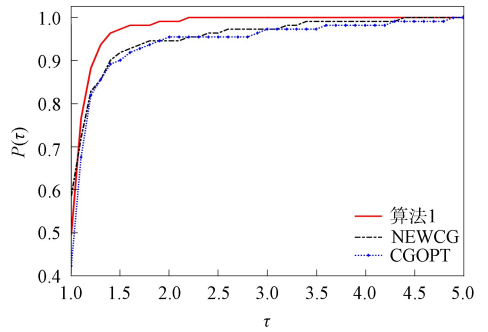


图 2 不同方法的函数计算次数比较
Fig. 2 Comparison of function calculation times for different methods

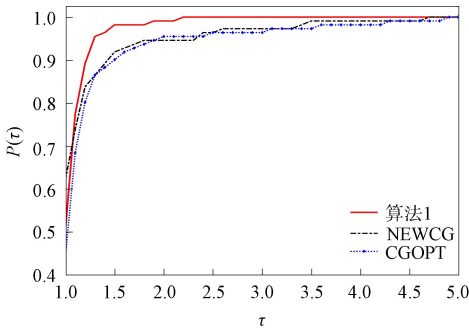


图 3 不同方法的梯度计算次数比较
Fig. 3 Comparison of gradient calculation times for different methods

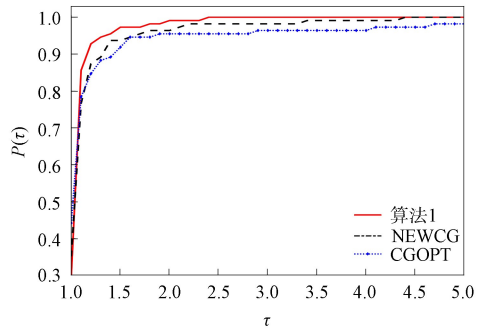


图 4 不同方法的 CPU 时间比较
Fig. 4 Comparison of CPU time for different methods

综上所述, 本文通过极小化 3 次正则化模型, 提出了一种新的 Dai-Liao 参数, 通过对函数性质的分析得到一个参数 t 的自适应选择. 并结合改进的 Wolfe 线搜索, 提出一种基于正则化模型的 Dai-Liao 共轭梯度法. 该算法的搜索方向满足充分下降性, 从而在一般假设条件下, 建立了该算法的全局收敛

性. 数值实验结果表明, 针对 CUTEst 库^[11]中的测试函数, 该算法优于 CGOPT 和 NEWCG 两种方法.

参 考 文 献

- [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): 409-436.
- [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 Extreme Problems [J]. USSR Computational Mathematical Physics, 1969, 9(4): 94-112.
- [5] DAI Y H, YUAN Y X. A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property [J]. SIAM Journal on Optimization, 1999, 10(1): 177-182.
- [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] HAGER W W, ZHANG H C. A New Conjugate Gradient with Guaranteed Descent and an Efficient Line Search [J]. SIAM Journal on Optimization, 2005, 16(1): 170-192.
- [8] DAI Y H, KOU C X. A Nonlinear Conjugate Gradient Algorithm with an Optimal Property and an Improved Wolfe Line Search [J]. SIAM Journal on Optimization, 2013, 23(1): 296-320.
- [9] PERRY A. A Class of Conjugate Gradient Algorithms with a Two-Step Variable Metric Memory [EB/OL]. (1977-01-01)[2023-09-23]. <https://hdl.handle.net/10409/220629>.
- [10] SHANNO D F. On the Convergence of a New Conjugate Gradient Algorithm [J]. SIAM Journal on Numerical Analysis, 1978, 15(6): 1247-1257.
- [11] GOULD N I M, ORBAN D, TOINT P L. CUTEst: A Constrained and Unconstrained Testing Environment with Safe Threads for Mathematical Optimization [J]. Computational Optimization and Applications, 2015, 60: 545-557.
- [12] ZHAO T, LIU H W, LIU Z X. New Subspace Minimization Conjugate Gradient Methods Based on Regularization Model for Unconstrained Optimization [J]. Numerical Algorithms, 2021, 87: 1501-1534.
- [13] HSIA Y, SHEU R L, YUAN Y X. Theory and Application of p -Regularized Subproblems for $p > 2$ [J]. Optimization Methods and Software, 2017, 32(5): 1059-1077.
- [14] DAI Y H, KOU C X. A Barzilai-Borwein Conjugate Gradient Method [J]. Science China Mathematics, 2016, 59: 1511-1524.
- [15] YUAN Y X. A Modified BFGS Algorithm for Unconstrained Optimization [J]. IMA Journal Numerical Analysis, 1991, 11(3): 325-332.
- [16] GILBERT J C, NOCEDAL J. Global Convergence Properties of Conjugate Gradient Methods for Optimization [J]. SIAM Journal on Optimization, 1992, 2(1): 21-42.
- [17] ZHANG K K, LIU H W, LIU Z X. A New Dai-Liao Conjugate Gradient Method with Optimal Parameter Choice [J]. Numerical Functional Analysis and Optimization, 2019, 40(2): 194-215.
- [18] DOLAN E D, MORÉ J J. Benchmarking Optimization Software with Performance Profiles [J]. Mathematical Programming, 2002, 91(2): 201-213.

(责任编辑: 李 琦)