doi：10.13413/j.cnki.jdxblxb.2023283

基于 Zhang-Hager 线搜索的改进近似最优梯度法

李 瑶[1]，刘红卫[1]，吕佳敏[1]，游海龙[2]
（1. 西安电子科技大学 数学与统计学院，西安 710126；2. 西安电子科技大学 微电子学院，西安 710071）

摘要：提出一种改进的近似最优梯度法，求解图划分问题中的无约束目标函数. 先用修正的 BFGS 更新公式及选取 BB 类步长的线性组合作为标量矩阵得到近似最优步长，再引入参数对经典的 Zhang-Hager 线搜索形式进行改进，构建算法框架并给出 R 线性收敛性证明. 实验结果表明，改进算法提高了原算法的性能.

关键词：修正的 BFGS 更新公式；近似最优步长；Zhang-Hager 线搜索；R 线性收敛性；图划分问题

中图分类号：O221.7 **文献标志码**：A **文章编号**：1671-5489(2024)02-0263-10

Improved Approximate Optimal Gradient Method Based on Zhang-Hager Line Search

LI Yao[1], LIU Hongwei[1], LÜ Jiamin[1], YOU Hailong[2]
(1. *School of Mathematics and Statistics, Xidian University, Xi'an* 710126, *China*;
2. *School of Microelectronics, Xidian University, Xi'an* 710071, *China*)

Abstract: We proposed an improved approximate optimal gradient method to solve the unconstrained objective function in the graph partition problem. We first used the modified BFGS updating formula and selected the linear combination of BB class step sizes as scalar matrices to obtain the approximate optimal step sizes, then we introduced parameters to improve the classical Zhang-Hager line search form, construced the algorithm framework and gave the proof of R-linear convergence. The experimental results show that the improved algorithm improves the performance of the original algorithm.

Keywords: modified BFGS updating formula; approximate optimal step size; Zhang-Hager line search; R-linear convergence; graph partition problem

最优化问题广泛应用于计划生产、设计工程、航空航天、企业运营和交通运输等领域. 其中无约束最优化问题作为最优化方法的基础已得到深入研究，无约束最优化问题可统一描述为以下数学模型：

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

式中：f 为待研究问题的目标函数；$\boldsymbol{x}$ 为影响问题的决策变量，定义在全体实数集 $\mathbb{R}^n$ 上.

无约束最优化模型(1)的求解方法主要分为线搜索法和信赖域法两大类. 线搜索法包括拟牛顿法、梯度法和共轭梯度法等，其中梯度法实现较简单且具有很好的收敛性，因此应用广泛. 但梯度法的数

收稿日期：2023-07-06.

第一作者简介：李 瑶(1997—)，女，汉族，硕士研究生，从事图划分问题的研究，E-mail：ly818400@163.com.

基金项目：国家自然科学基金(批准号：12261019).

值性能很大程度上依赖于步长，故步长的选取至关重要. 针对二次凸函数，Cauchy[1]设计了一种精确步长选取策略 $\lambda^k=\underset{\lambda>0}{\operatorname{argmin}} f(\boldsymbol{x}^k-\lambda\boldsymbol{g}^k)$，该策略称为 Cauchy 步长，对应的算法称为最速下降法. 但在目标函数病态时，该方法收敛慢且可能产生锯齿现象. Barzilai 和 Borwein[2]对梯度法进行改进，建立了新的步长选取规则，提出了 Barzilar-Borwein(BB)算法；Raydan[3]在目标函数为二次函数且严格凸时证明了 BB 算法的全局收敛性；Dai 等[4]给出了 BB 算法 R 收敛性的证明. 文献[2]数值实验结果表明，BB 算法的效果很多情况下优于多种经典梯度法；在此基础上，许多研究者利用修正的 BFGS(Broyden-Fletcher-Goldfarb-Shanno)更新公式提出了改进的 BB 算法，得到了多种 BB 类步长[5-8]，其中BFGS 方法广泛应用于近似 Hesse 矩阵的更新.

特别地，Liu 等[9]受 BB 算法的启发，基于二次模型和锥模型并结合 Zhang-Hager(ZH)非单调线搜索[10]提出了近似最优梯度法(GM_AOS)，得到的步长称为近似最优步长. 研究表明，不同标量矩阵的选取、$\boldsymbol{B}^k$ 的更新策略及搜索规则的选取对近似最优梯度法的算法性能有较大影响. 故本文基于 Liu 等[9,11]的研究，对原近似最优梯度法进行改进. 在构建模型得出步长的过程中，采用线性组合的形式选取标量矩阵，并基于修正的切割方程[12-13]更新 $\boldsymbol{B}^k$，进一步得到改进的近似最优步长. 此外，ZH 非单调线搜索是解决无约束优化问题非常有效的一种方法. 本文考虑对经典的 ZH 搜索法进行改进，结合上述新的近似最优步长，构建改进的近似最优梯度法(GM_IAOS)，并将改进的近似最优梯度法应用于电路划分问题中[14-15]. 通过电路划分实际的数值实验结果可见，本文算法性能优于原近似最优梯度法.

1 预备知识

定义 1(梯度法)[2] 在所有的迭代算法中，梯度法具有最简单的形式：

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k-\lambda^k\boldsymbol{g}^k,$$

这里 λ^k 表示步长，$\boldsymbol{g}^k$ 表示 $f(\boldsymbol{x}^k)$的梯度.

定义 2(BB 算法)[2] 对于无约束问题(1)，通过问题 $\min\limits_{\lambda>0}\left\|\dfrac{\boldsymbol{s}^{k-1}}{\lambda}-\boldsymbol{y}^{k-1}\right\|_2^2$ 和 $\min\limits_{\lambda>0}\|\lambda\boldsymbol{y}^{k-1}-\boldsymbol{s}^{k-1}\|_2^2$ 的求解得到经典 BB 步长：

$$\lambda_{\mathrm{BB1}}^k=\frac{\|\boldsymbol{s}^{k-1}\|^2}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}},\qquad \lambda_{\mathrm{BB2}}^k=\frac{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}}{\|\boldsymbol{y}^{k-1}\|^2},\tag{2}$$

这里 $\boldsymbol{s}^{k-1}=\boldsymbol{x}^k-\boldsymbol{x}^{k-1}$，$\boldsymbol{y}^{k-1}=\boldsymbol{g}^k-\boldsymbol{g}^{k-1}$.

此外对于梯度算法，构建目标函数的近似模型为

$$f(\boldsymbol{x}^k-\lambda^k\boldsymbol{g}^k)\approx f(\boldsymbol{x}^k)-\lambda(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k+\frac{1}{2}\lambda^2(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k,\tag{3}$$

步长 λ_{BB1}^k 也可通过极小化近似模型(3)得到. 这里的近似模型为二次函数，$\boldsymbol{B}^k$ 为 $\nabla^2 f(\boldsymbol{x}^k)$的某种近似.

定义 3(近似最优步长)[9] 在目标函数 f 可微且连续的条件下，对梯度法的某一近似模型 $\Theta(\lambda)$，若 $\lambda^*=\underset{\lambda>0}{\operatorname{argmin}}\,\Theta(\lambda)$成立，则近似模型的极小值点 λ^* 称为近似最优步长.

特别地，若 $\Theta^k(\lambda)$取式(3)，则解出的步长即为 BB 步长，对应的算法即为 BB 算法. 若 $\Theta^k(\lambda)=f(\boldsymbol{x}^k-\lambda^k\boldsymbol{g}^k)$，则得到的步长为 Cauchy 步长，此时的近似最优梯度法退化为最速下降法.

2 改进的近似最优梯度法

2.1 近似最优步长

近似最优步长的有效性与近似模型密切相关. 对于无约束优化问题，目前已基于二次模型得到了很多有效算法. 但当目标函数在迭代点有剧烈的曲率变化时，用二次函数近似对迭代结果会有较大偏差. 故当迭代点处目标函数的二次性态(即目标函数关于二次函数的接近程度)未知时，可分为如下两种情形考虑：

1) 当 $f(\boldsymbol{x}^k-\lambda\boldsymbol{g}^k)$在迭代点的二次性态弱时，考虑采用锥模型；

2) 当 $f(\boldsymbol{x}^k-\lambda\boldsymbol{g}^k)$在迭代点处二次性态强或锥模型不适用时，考虑采用二次模型及特殊的二次模型近似.

记 $\vartheta^k=\left|1-\dfrac{2(f^{k-1}-f^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{s}^{k-1})}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}}\right|$，用 ϑ^k 衡量目标函数在迭代点 $\boldsymbol{x}^k$ 和 $\boldsymbol{x}^{k-1}$ 之间二次性态的强弱，规定

$$\vartheta^k\leqslant m_1 \text{ 或 } \max\{\vartheta^k,\vartheta^{k-1}\}\leqslant m_2. \tag{4}$$

当满足式(4)时，认为目标函数二次性态强，其中 $0<m_1<m_2$.

情形 1) 当目标函数二次性态弱时，构建如下的锥模型[16]：

$$\Theta^k(\boldsymbol{p})=f^k+\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{p}}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{p}+1}+\frac{1}{2}\frac{\boldsymbol{p}^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{p}}{[(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{p}+1]^2}, \tag{5}$$

其中

$$\boldsymbol{\beta}^k=\frac{1-d^k}{d^k(\boldsymbol{g}^{k-1})^{\mathrm{T}}\boldsymbol{s}^{k-1}}\boldsymbol{g}^{k-1},\quad d^k=\frac{-(\boldsymbol{g}^{k-1})^{\mathrm{T}}\boldsymbol{s}^{k-1}}{\alpha^k+f^{k-1}-f^k},\quad \alpha^k=\sqrt{\rho^k},$$

$$\rho^k=(f^{k-1}-f^k)^2-((\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{s}^{k-1})((\boldsymbol{g}^{k-1})^{\mathrm{T}}\boldsymbol{s}^{k-1}).$$

许多 BB 类算法都是经典的 BB 算法基于拟牛顿方程的一种变形. 由于这里选取的 $\tilde{\boldsymbol{y}}^{k-1}$ 相对于 $\boldsymbol{y}^{k-1}=\boldsymbol{g}^k-\boldsymbol{g}^{k-1}$ 是关于 $\nabla^2 f(\boldsymbol{x}^k)\boldsymbol{s}^{k-1}$ 更优的近似，故 BB 类算法产生的步长 $\dfrac{(\tilde{\boldsymbol{s}}^{k-1})^{\mathrm{T}}\tilde{\boldsymbol{y}}^{k-1}}{\|\tilde{\boldsymbol{y}}^{k-1}\|^2}$ 比经典的步长 $\dfrac{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}}{\|\boldsymbol{y}^{k-1}\|^2}$ 更高效.

考虑修正的拟牛顿方程：

$$\boldsymbol{B}^k\boldsymbol{\Delta}^{k-1}=\hat{\boldsymbol{\omega}}^{k-1},$$

这里

$$\boldsymbol{\Delta}^{k-1}=d^k\boldsymbol{s}^{k-1},\quad \hat{\boldsymbol{\omega}}^{k-1}=\frac{1}{d^k}\boldsymbol{w}^{k-1},\quad \boldsymbol{w}^{k-1}=d^k\boldsymbol{g}^k-\frac{1}{d^k}\boldsymbol{g}^{k-1}.$$

记得到的 BB 类步长为

$$\lambda^k_{\mathrm{NBB1}}=\frac{(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\boldsymbol{\Delta}^{k-1}}{(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}},\qquad \lambda^k_{\mathrm{NBB2}}=\frac{(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}}{(\hat{\boldsymbol{\omega}}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}}.$$

由于对不同问题两种步长选取的更优性未知，且在构造标量矩阵时本文不愿放弃两者中的任一种，故本文考虑两种步长的线性组合，通过调整参数控制 $\dfrac{1}{\lambda^k_{\mathrm{NBB1}}}\boldsymbol{I}$ 和 $\dfrac{1}{\lambda^k_{\mathrm{NBB2}}}\boldsymbol{I}$ 对标量矩阵的贡献程度.

构造的标量矩阵为

$$\boldsymbol{\chi}^k=\left(a_1\frac{(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}}{(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\boldsymbol{\Delta}^{k-1}}+a_2\frac{(\hat{\boldsymbol{\omega}}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}}{(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}}\right)\boldsymbol{I}, \tag{6}$$

这里用 a_1,a_2 作为贡献参数. 由广义 BFGS 公式[16]对标量矩阵更新得

$$\boldsymbol{B}^k=\boldsymbol{\chi}^k-\frac{\boldsymbol{\chi}^k\boldsymbol{\Delta}^{k-1}(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\boldsymbol{\chi}^k}{(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\boldsymbol{\chi}^k\boldsymbol{\Delta}^{k-1}}+\frac{\boldsymbol{\gamma}^{k-1}(\boldsymbol{\gamma}^{k-1})^{\mathrm{T}}}{(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\boldsymbol{\gamma}^{k-1}}, \tag{7}$$

结合式(6)和式(7)易知，当 $(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}>0$ 时，$\boldsymbol{\chi}^k$ 正定，进一步有 $\boldsymbol{B}^k$ 对称正定.

将 $\boldsymbol{p}=-\lambda\boldsymbol{g}^k$ 代入式(5)，得到 $f(\boldsymbol{x}^k-\lambda\boldsymbol{g}^k)$的锥近似

$$\Theta^k_1(\lambda)=f^k-\frac{\lambda(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{1-\lambda(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}+\frac{\lambda^2}{2}\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k}{[1-\lambda(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k]^2}. \tag{8}$$

分析式(8)可得 $\Theta^k_1(\lambda)$的奇异点 $\hat{\lambda}^k=\dfrac{1}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}$. 由于函数 $\Theta^k_1(\lambda)$在除奇异点外是连续可微的，且关于 λ 求导有

$$\frac{\mathrm{d}(\Theta^k_1(\lambda))}{\mathrm{d}\lambda}=\frac{\lambda((\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)-(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{(1-\lambda(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)^3}. \tag{9}$$

在 $\rho^k>0$ 且 $(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}>0$ 成立的条件下，当 $(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k\neq0$ 时，解 $\dfrac{\mathrm{d}(\Theta^k_1(\lambda))}{\mathrm{d}\lambda}=0$ 得

$\Theta_1^k(\lambda)$的唯一驻点为

$$\bar{\lambda}^k=\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}.$$

下面分两种情形考虑驻点$\bar{\lambda}^k$.

① $\hat{\lambda}^k<0$. 设$(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k<0$, 则

$$\bar{\lambda}^k=\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}=\frac{1}{\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k}{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}+(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}<\frac{1}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}.$$

由式(9)可知

$$\lim_{\lambda\to((\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)^{-1}}\Theta_1^k(\lambda)=+\infty,\qquad\lim_{\lambda\to+\infty}\Theta_1^k(\lambda)=f^k+\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}+\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k}{2((\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)^2}.\tag{10}$$

当$\lambda>\frac{1}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}$时, 有$\frac{\mathrm{d}(\Theta_1^k(\lambda))}{\mathrm{d}\lambda}<0$, 则不存在$\lambda^*>0$满足$\lambda^*=\underset{\lambda>0}{\operatorname{argmin}}\,\Theta_1^k(\lambda)$. 反之, 当$(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k>0$时, 显然有$\bar{\lambda}^k>0$. 对于$\lambda>\bar{\lambda}^k$, 有

$$\lambda((\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)-(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k>0,\qquad1-\lambda(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k>0.$$

故此时

$$\frac{\mathrm{d}(\Theta_1^k(\lambda))}{\mathrm{d}\lambda}=\frac{\lambda((\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)-(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{(1-\lambda(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)^3}>0,$$

函数$\Theta_1^k(\lambda)$在$[\bar{\lambda}^k,+\infty)$上单调增.

由于函数$\Theta_1^k(\lambda)$在除奇异点外是连续可微的, 且$\left.\frac{\mathrm{d}(\Theta_1^k(\lambda))}{\mathrm{d}\lambda}\right|_{\lambda=0}=-\|\boldsymbol{g}^k\|^2<0$, 故函数$\Theta_1^k(\lambda)$在$[0,\bar{\lambda}^k)$上单调减. 从而当$(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k>0$时, 有$\bar{\lambda}^k=\underset{\lambda>0}{\operatorname{argmin}}\,\Theta_1^k(\lambda)$是关于近似模型$\Theta_1^k(\lambda)$的近似最优步长.

② $\hat{\lambda}^k>0$. 此时易见$0<\bar{\lambda}^k<\frac{1}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}$. 在区间$\left(\bar{\lambda}^k,\frac{1}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}\right)$上, 有$1-\lambda(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k>0$, 且

$$\lambda((\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)-(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k>0.$$

故$\frac{\mathrm{d}(\Theta_1^k(\lambda))}{\mathrm{d}\lambda}>0$, 函数$\Theta_1^k(\lambda)$在$[\bar{\lambda}^k,+\infty)$上单调增. 由于函数$\Theta_1^k(\lambda)$在除奇异点外是连续可微的, 且$\left.\frac{\mathrm{d}(\Theta_1^k(\lambda))}{\mathrm{d}\lambda}\right|_{\lambda=0}=-\|\boldsymbol{g}^k\|^2<0$, 故函数$\Theta_1^k(\lambda)$在$[0,\bar{\lambda}^k)$上单调减. 结合式(10), 当$\lambda>\frac{1}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}$时, 有

$$\Theta_1^k(\bar{\lambda}^k)-\lim_{\lambda\to+\infty}\Theta_1^k(\lambda)=-\frac{((\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k)^2}{2(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k}-\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k}-\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k}{2((\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k)^2}<0.$$

故$\bar{\lambda}^k=\underset{\lambda>0}{\operatorname{argmin}}\,\Theta_1^k(\lambda)$是关于函数$\Theta_1^k(\lambda)$的近似最优步长.

综上分析, 当式(4)不成立且

$$\rho^k>0,\quad(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}>0,\quad(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k>0$$

成立时, 可得近似最优步长$\bar{\lambda}^k$. 考虑到实际计算时的数值效果, 采用以下截断式:

$$\bar{\lambda}^k=\begin{cases}\max\{\min\{\bar{\lambda}^k,\lambda_{\mathrm{BB1}}^k\},\lambda_{\mathrm{BB2}}^k\}, & (\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}>0,\\ \bar{\lambda}^k, & \text{其他}.\end{cases}$$

情形 2) 分两种情况构造情形 1)不满足时的二次近似模型.

① $(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}>0$. 此时目标函数有较强的二次性态或锥模型不适用, 本文考虑使用二次近似, 构建如下近似模型:

$$\Theta_2^k(\lambda)=f(\boldsymbol{x}^k)-\lambda\|\boldsymbol{g}^k\|^2+\frac{1}{2}\lambda^2(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k.$$

基于修正的切割方程[12]

$$\boldsymbol{B}^k\boldsymbol{s}^{k-1}=\hat{\boldsymbol{y}}^{k-1},\qquad\hat{\boldsymbol{y}}^{k-1}=\boldsymbol{y}^{k-1}+\frac{\varphi^k}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{u}}\boldsymbol{u},$$

选取 $\boldsymbol{u}=\boldsymbol{y}^{k-1}$，$\varphi^k=3(\boldsymbol{g}^k+\boldsymbol{g}^{k-1})^{\mathrm{T}}\boldsymbol{s}^{k-1}+6(f^{k-1}-f^k)$. 由此推导出的 BB 类步长为

$$\lambda_{\mathrm{MBB1}}^k=\frac{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{s}^{k-1}}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}},\qquad \lambda_{\mathrm{MBB2}}^k=\frac{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}{(\hat{\boldsymbol{y}}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}.$$

与锥模型类似，取两者的线性组合作为标量矩阵：

$$\boldsymbol{\chi}^k=\left(a_3\frac{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{s}^{k-1}}+a_4\frac{(\hat{\boldsymbol{y}}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}\right)\boldsymbol{I}.$$

对应使用修正的 BFGS 更新公式[13]，有

$$\boldsymbol{B}^k=\boldsymbol{\chi}^k-\frac{\boldsymbol{\chi}^k\boldsymbol{s}^{k-1}(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{\chi}^k}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{\chi}^k\boldsymbol{s}^{k-1}}+\frac{\hat{\boldsymbol{y}}^{k-1}(\hat{\boldsymbol{y}}^{k-1})^{\mathrm{T}}}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}},\tag{11}$$

这里 $\hat{\boldsymbol{y}}^{k-1}=\boldsymbol{y}^{k-1}+\frac{\varphi^k}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}}\boldsymbol{y}^{k-1}$，$\varphi^k=3(\boldsymbol{g}^k+\boldsymbol{g}^{k-1})^{\mathrm{T}}\boldsymbol{s}^{k-1}+6(f^{k-1}-f^k)$.

对于上述 φ^k，$\exists t\in[0,1]$，使得

$$\varphi^k=3((\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}-(\boldsymbol{s}^{k-1})^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^{k-1}+t\boldsymbol{s}^{k-1})\boldsymbol{s}^{k-1}).$$

参考数值结果，这里限制

$$\varphi^k=\min\{\max\{\varphi^k,-h(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}\},h(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}\},\qquad 0<h<0.1.$$

由该限制可知，若$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}>0$，则

$$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}=(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}+\varphi^k\geqslant(1-h)(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}.$$

于是可得如下引理.

引理 1[9]　若$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}>0$，则

$$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}=(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}+\varphi^k\geqslant(1-k)(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}>0,$$

且对应的矩阵 $\boldsymbol{B}^k$ 是对称正定的.

令$\frac{\mathrm{d}(\Theta_2^k(\lambda))}{\mathrm{d}\lambda}=0$，由 $\Theta_2^k(\lambda)$是二次近似，可求得唯一的驻点为

$$\bar{\lambda}^k=\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k}=\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{\alpha\left(\|\boldsymbol{g}^k\|^2-\frac{((\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{s}^{k-1})^2}{\|\boldsymbol{s}^{k-1}\|^2}\right)+\frac{\left((\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{y}^{k-1}+\frac{\varphi^k(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{y}^{k-1}}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}}\right)^2}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}},$$

这里 $\alpha=a_3\frac{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{s}^{k-1}}+a_4\frac{(\hat{\boldsymbol{y}}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}{(\boldsymbol{s}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{y}}^{k-1}}$. 故当$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}>0$ 时，若式(4)成立或 $\rho^k>0$，$(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}>0$，$(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k>0$ 三者不同时成立，则采取上述模型近似可得对应近似最优步长. 与情形 1)类似，实际计算时采用以下形式：

$$\bar{\lambda}^k=\max\{\min\{\bar{\lambda}^k,\lambda_{\mathrm{BB1}}^k\},\lambda_{\mathrm{BB2}}^k\}.$$

② $(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}\leqslant 0$. 文献[17]介绍了一种特殊的二次近似模型，其主要思想是利用 $f(x)$的二阶 Taylor 展式得到近似模型：

$$\Theta_3^k(\lambda)=f(\boldsymbol{x}^k)-\lambda(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k+\frac{1}{2}\lambda^2\left|\frac{(\boldsymbol{g}^k)^{\mathrm{T}}(\boldsymbol{g}(\boldsymbol{x}^k-\mu^k\boldsymbol{g}^k)-\boldsymbol{g}(\boldsymbol{x}^k))}{\mu^k}\right|,$$

这里 $\left|\frac{(\boldsymbol{g}^k)^{\mathrm{T}}(\boldsymbol{g}(\boldsymbol{x}^k-\mu^k\boldsymbol{g}^k)-\boldsymbol{g}(\boldsymbol{x}^k))}{\mu^k}\right|$ 是 $(\boldsymbol{g}^k)^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^k)\boldsymbol{g}^k$ 的近似，$\mu^k>0$. 当 $\left|\frac{(\boldsymbol{g}^k)^{\mathrm{T}}(\boldsymbol{g}(\boldsymbol{x}^k-\mu^k\boldsymbol{g}^k)-\boldsymbol{g}(\boldsymbol{x}^k))}{\mu^k}\right|\neq 0$ 时，令$\frac{\mathrm{d}(\Theta_3^k(\lambda))}{\mathrm{d}\lambda}=0$ 可得近似最优步长

$$\tilde{\lambda}^k=\frac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{\left|\frac{(\boldsymbol{g}^k)^{\mathrm{T}}(\boldsymbol{g}(\boldsymbol{x}^k-\mu^k\boldsymbol{g}^k)-\boldsymbol{g}(\boldsymbol{x}^k))}{\mu^k}\right|};$$

当 $\left|\frac{(\boldsymbol{g}^k)^{\mathrm{T}}(\boldsymbol{g}(\boldsymbol{x}^k-\mu^k\boldsymbol{g}^k)-\boldsymbol{g}(\boldsymbol{x}^k))}{\mu^k}\right|=0$ 时，令 $\tilde{\lambda}^k=l\lambda^{k-1}$，$l>0$.

Liu 等[18]为降低计算成本提出了一种共线策略：当$\frac{\|\boldsymbol{g}^{k-1}\|^2}{\|\boldsymbol{g}^k\|^2}\geqslant\xi$时，认为 $\boldsymbol{g}^{k-1}$和 $\boldsymbol{g}^k$ 接近共线，取 $\xi>0$且为接近于 1 的常数，则 $\|\boldsymbol{g}^{k-1}\|$ 和 $\|\boldsymbol{g}^k\|$ 近似相等. 此时有

$$(\boldsymbol{g}^k)^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^k)\boldsymbol{g}^k \approx (\boldsymbol{g}^{k-1})^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^k)\boldsymbol{g}^{k-1} \approx \left|\frac{(\boldsymbol{g}(\boldsymbol{x}^k+\lambda^{k-1}\boldsymbol{g}^{k-1})-\boldsymbol{g}(\boldsymbol{x}^k))^{\mathrm{T}}\boldsymbol{g}^{k-1}}{\lambda^{k-1}}\right| = \frac{|(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}|}{(\lambda^{k-1})^2}.$$

由此可得新模型为

$$\Theta_4^k(\lambda) = f(\boldsymbol{x}^k) - \lambda(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k + \frac{1}{2}\lambda^2\frac{|(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}|}{(\lambda^{k-1})^2}.$$

当$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}\neq 0$时，令$\frac{\mathrm{d}(\Theta_4^k(\lambda))}{\mathrm{d}\lambda}=0$可得近似最优步长$\tilde{\lambda}^k=\frac{\|\boldsymbol{g}^k\|^2}{|(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}|}(\lambda^{k-1})^2$；当$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}=0$时，仍取$\tilde{\lambda}^k=l\lambda^{k-1}$，$l>0$.

故当$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}\leqslant 0$时，若式(4)成立或$\rho^k>0$，$(\boldsymbol{\Delta}^{k-1})^{\mathrm{T}}\hat{\boldsymbol{\omega}}^{k-1}>0$，$(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{B}^k\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k(\boldsymbol{\beta}^k)^{\mathrm{T}}\boldsymbol{g}^k>0$三者不同时成立，则采用以下步长选取策略：

$$\tilde{\lambda}^k=\begin{cases}\dfrac{(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k}{|\boldsymbol{\gamma}^k|}, & \dfrac{\|\boldsymbol{g}^{k-1}\|^2}{\|\boldsymbol{g}^k\|^2}<\xi \text{ 且 } \gamma^k\neq 0,\\ \dfrac{\|\boldsymbol{g}^k\|^2}{|(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}|}(\lambda^{k-1})^2, & \dfrac{\|\boldsymbol{g}^{k-1}\|^2}{\|\boldsymbol{g}^k\|^2}\geqslant\xi \text{ 且 } (\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}\neq 0,\\ l\lambda^{k-1}, & \text{其他},\end{cases}$$

这里$\gamma^k=\frac{(\boldsymbol{g}^k)^{\mathrm{T}}(\boldsymbol{g}(\boldsymbol{x}^k-\mu^k\boldsymbol{g}^k)-\boldsymbol{g}(\boldsymbol{x}^k))}{\mu^k}$.

2.2 改进的 Zhang-Hager 线搜索

经典 Zhang-Hager 线搜索的形式[13]为：给定常数$\delta\in(0,1)$，寻找$\lambda^k>0$，使得

$$f(\boldsymbol{x}^k+\lambda^k\boldsymbol{d}^k)\leqslant C^k+\delta\lambda^k\|\boldsymbol{d}^k\|^2, \tag{12}$$

其中

$$Q^0=1,\quad C^0=f(\boldsymbol{x}^0),\quad Q^{k+1}=\eta^kQ^k+1,\quad C^{k+1}=\frac{\eta^kQ^kC^k+f(\boldsymbol{x}^{k+1})}{Q^{k+1}}, \tag{13}$$

$\eta^k\in[\eta_{\min},\eta_{\max}]\subset[0,1]$. 这里参数$\eta^k$可控制线搜索的非单调程度.

为改善搜索性能，本文引入参数n^k对式(12)进行改进：

$$f(\boldsymbol{x}^k+\lambda^k\boldsymbol{d}^k)\leqslant C^k+n^k+\delta\lambda^k\|\boldsymbol{d}^k\|^2, \tag{14}$$

其中

$$n^k=\begin{cases}0, & k=0,\\ \min\left\{\dfrac{1}{k\lg(k/n+12)}, C^k-f(\boldsymbol{x}^k), \tau|\delta\lambda^k\|\boldsymbol{d}^k\|^2|\right\}, & k\geqslant 1,\end{cases} \tag{15}$$

$\tau\in(0,1)$，n为正整数.

2.3 改进的近似最优梯度法

根据上述构建的模型以及得到的对应近似最优步长，并结合改进的 Zhang-Hager 线搜索可得改进的近似最优梯度法如下.

算法 1 改进的近似最优梯度法.

给定参数：$\varepsilon>0$，$\lambda_{\min}$，$\lambda_{\max}$，η^k，δ，τ，l，h，ξ，μ^k，a_1，a_2，a_3，a_4，m_1，m_2，ϑ^0，$\dot{\lambda}^0$，选取初始点$\boldsymbol{x}^0$，设$Q^0=1$，$C^0=f^0$，$i=0$.

步骤 1) 当$\|\boldsymbol{g}^k\|_\infty\leqslant\varepsilon$时，停止迭代；

步骤 2) ① $k=0$时，设$\lambda=\dot{\lambda}^0$，转步骤 3)；

② 当满足情形 1)的条件时，取近似最优步长$\bar{\lambda}^k$，设$\dot{\lambda}^k=\max\{\min\{\bar{\lambda},\lambda_{\max}\},\lambda_{\min}\}$及$\lambda=\dot{\lambda}^k$，转步骤 3)；

③ 当满足情形 2)中的$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}>0$时，取近似最优步长$\bar{\bar{\lambda}}^k$，设$\dot{\lambda}^k=\max\{\min\{\bar{\bar{\lambda}}^k,\lambda_{\max}\},\lambda_{\min}\}$及$\lambda=\dot{\lambda}^k$，转步骤 3)；

④ 当满足情形 2)中的$(\boldsymbol{s}^{k-1})^{\mathrm{T}}\boldsymbol{y}^{k-1}\leqslant 0$时，取近似最优步长$\tilde{\lambda}^k$，设$\dot{\lambda}^k=\max\{\min\{\tilde{\lambda},\lambda_{\max}\},\lambda_{\min}\}$及$\lambda=\dot{\lambda}^k$，转步骤 3)；

步骤 3) 使用改进的 Zhang-Hager 线搜索：令 $\boldsymbol{d}^k=-\boldsymbol{g}^k$，若

$$f(\boldsymbol{x}^k-\lambda^k\boldsymbol{g}^k)\leqslant C^k+n^k+\delta\lambda^k\|\boldsymbol{g}^k\|^2,$$

则 $\lambda^k=\lambda$，转步骤 4)；否则，取 λ 为以下形式：

$$\lambda=\begin{cases}\hat{\lambda}, & \lambda>0.1\dot{\lambda}^k \text{ 且 } \hat{\lambda}\in[0.1\dot{\lambda}^k,0.9\lambda],\\ 0.5\lambda, & \text{其他},\end{cases}$$

这里 $\hat{\lambda}$ 通过二次插值选取，转步骤 3)；

步骤 4) 通过式(14)更新 Q^{k+1} 和 C^{k+1}；

步骤 5) 令 $\boldsymbol{x}^{k+1}=\boldsymbol{x}^k-\lambda^k\boldsymbol{g}^k$，$k=k+1$，转步骤 1).

2.4　收敛性证明

下面给出算法 1 的收敛性证明，涉及的函数全部定义在实数域 $\mathbb{R}^n$ 上. 假设：

(H_1) $f(\boldsymbol{x})$在定义域上连续可微且下有界；

(H_2) $g(\boldsymbol{x})$在定义域上是 Lipschitz 连续的，即 $\exists L>0$，使得

$$\|g(\boldsymbol{x}^1)-\boldsymbol{g}(\boldsymbol{x}^2)\|\leqslant L\|\boldsymbol{x}^1-\boldsymbol{x}^2\|,\qquad\forall\boldsymbol{x}^1,\boldsymbol{x}^2\in\mathbb{R}^n.$$

引理 2[17]　对算法 1 产生的序列$\{\boldsymbol{x}^k\}$，有 $f^k\leqslant C^k\leqslant Z^k$，这里 $Z^k=\dfrac{1}{k+1}\sum\limits_{i=0}^{k}f^i$.

证明：由式(14),(15)可知

$$f(\boldsymbol{x}^k-\lambda^k\boldsymbol{g}^k)\leqslant C^k+n^k-\delta\lambda^k\|\boldsymbol{g}^k\|^2\leqslant C^k-(1-\tau)\delta\lambda^k\|\boldsymbol{g}^k\|^2\leqslant C^k.\tag{16}$$

下面的证明与文献[10]中引理 1.1 的证明类似，故略.

引理 3　在假设(H_1)和(H_2)成立的条件下，对算法 1 产生的序列$\{\boldsymbol{x}^k\}$，$\exists c>0$，使得 $\lambda^k\geqslant c$.

证明：当 $\dot{\lambda}^k$ 满足搜索式(14)时，显然有 $\lambda=\dot{\lambda}^k\geqslant\lambda_{\min}$. 当 $\dot{\lambda}^k$ 不满足搜索式(14)时，分两种情形讨论：

1) 若 $\lambda\leqslant 0.01\dot{\lambda}^k$，则由引理 2 及 n^k 的定义可知 $\forall k\in\mathbb{N}^n$，$n^k\geqslant 0$. 再结合算法 1 中设计的步长选取规则可得

$$f(\boldsymbol{x}^k-2\lambda^k\boldsymbol{g}^k)>C^k+n^k-2\delta\lambda^k\|\boldsymbol{g}^k\|^2\geqslant C^k-2\delta\lambda^k\|\boldsymbol{g}^k\|^2\geqslant f^k-2\delta\lambda^k\|\boldsymbol{g}^k\|^2.$$

根据微分中值定理可知，$\exists\varepsilon^k\in(0,1)$，使得

$$\frac{f(\boldsymbol{x}^k-2\lambda^k\boldsymbol{g}^k)-f(\boldsymbol{x}^k)}{-2\lambda^k\boldsymbol{g}^k}=g(\boldsymbol{x}^k-2\varepsilon^k\lambda^k\boldsymbol{g}^k).\tag{17}$$

再根据假设(H_2)可得

$$\|g(\boldsymbol{x}^k-2\varepsilon^k\lambda^k\boldsymbol{g}^k)-\boldsymbol{g}(\boldsymbol{x}^k)\|\leqslant L\|-2\varepsilon^k\lambda^k\boldsymbol{g}^k\|=2L\varepsilon^k\lambda^k\boldsymbol{g}^k.\tag{18}$$

结合式(17),(18)有

$$2L\lambda^k\|\boldsymbol{g}^k\|_2^2\geqslant 2L\varepsilon^k\lambda^k\|\boldsymbol{g}^k\|_2^2\geqslant-\boldsymbol{g}(\boldsymbol{x}^k-2\varepsilon^k\lambda^k\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k+(\boldsymbol{g}^k)^{\mathrm{T}}\boldsymbol{g}^k\geqslant(1-\delta)\|\boldsymbol{g}^k\|_2^2,$$

即 $\lambda^k\geqslant\dfrac{(1-\delta)}{2L}$.

2) 若 $\lambda>0.01\dot{\lambda}^k$，则显然 $\lambda^k>0.01\dot{\lambda}^k\geqslant 0.01\lambda_{\min}$. 令 $c=\min\left\{0.01\lambda_{\min},\dfrac{(1-\delta)}{2L}\right\}$，即可得要证明的结果.

定理 1　在假设(H_1)和(H_2)成立的条件下，若$\{\boldsymbol{x}^k\}$为算法 1 产生的序列，则可得梯度的一些性质如下：

$$\liminf_{k\to\infty}\|\boldsymbol{g}^k\|=0.$$

特别地，当 $\eta_{\max}<1$ 时，有

$$\lim_{k\to\infty}\|\boldsymbol{g}^k\|=0.$$

因此$\{\boldsymbol{x}^k\}$的任何一个聚点都收敛于驻点 $\bar{\boldsymbol{x}}$.

证明：结合式(16)及引理 3，可得

$$f(\boldsymbol{x}^k-\lambda^k\boldsymbol{g}^k)\leqslant C^k+n^k-\delta\lambda^k\|\boldsymbol{g}^k\|^2\leqslant C^k-(1-\tau)\delta\lambda^k\|\boldsymbol{g}^k\|^2\leqslant C^k-c(1-\tau)\delta\|\boldsymbol{g}^k\|^2.$$

利用文献[10]中定理2.2的证明，令其证明过程中的$\beta=c(1-\tau)\delta$，可得结论.

定理2[10] 若增加假设f为一致凸函数，则f有全局极小值点$\bar{\boldsymbol{x}}$，且唯一.

对于算法1产生的序列$\{\boldsymbol{x}^k\}$，若$\eta_{\max}<1$，则$\exists\,\zeta\in(0,1)$，使得

$$f(\boldsymbol{x}^k)-f(\bar{\boldsymbol{x}})\leqslant\zeta^k(f(\boldsymbol{x}^0)-f(\bar{\boldsymbol{x}})).$$

3 数值实验

目前，解决电路划分问题的一种有效策略是将超图转化为图模型，再使用数学方法对其进行抽象. 王晓瑜等[19]对电路划分问题进行数学建模，将其松驰为无约束优化问题，并设计了求解图划分问题的启发式框架. IBM公司Austin实验室设计转换了ISPD98电路测试基准，是目前公认的电路测试样例[20]. 本文在文献[19]算法框架的基础上，将其中求解无约束优化问题的内点法替换为近似最优梯度法[19]以及本文改进的近似最优梯度法，解决其图划分问题. 通过选取文献[20]的两组数据(每组18个电路划分问题)，对比文献[9]中的近似最优梯度法和本文改进的近似最优梯度法的性能. 由于图多分问题是在图二分问题基础上多次迭代重复使用算法得出，涉及随机数据的选取及调整策略，故数据结果不具有说服力. 为此，本文仅使用图二分问题进行对比实验，但本文的改进算法也适用于图多分问题.

使用两组数据关于目标函数值、迭代次数、函数值计算次数、算法运行时间对两种算法进行对比，结果分别列于表1和表2. 选取初始点$\boldsymbol{x}^0$为圆上的均匀分布，其他参数选取如下(其中共有参数选取策略相同)：

$$\lambda_{\min}=10^{-30},\quad \lambda_{\max}=10^{30},\quad \eta^k=1,\quad \delta=10^{-4},\quad \tau=0.9,\quad l=10,$$

$$h=5/3\times10^{-5},\quad \xi=0.9,\quad \mu^k=\min\{0.1\lambda^{k-1},0.01\},$$

$$a_1=1.87,\quad a_2=1.85,\quad a_3=0.3,\quad a_4=0.7,$$

$$m_1=10^{-8},\quad m_2=0.07,\quad \vartheta^0=1,\quad \dot{\lambda}^0=1.$$

表1 在文献[19]框架基础上文献[9]算法(GM_AOS)与本文改进算法(算法1)在第1组数据下的实验结果

Table 1 Experimental results of algorithm (GM_AOS) in reference [9] and proposed improvement algorithm (algorithm 1) based on framework of reference [19] under the first set of data

电路划分问题	目标函数值		迭代次数		函数值计算次数		运行时间/s	
	文献[9]算法	算法1	文献[9]算法	算法1	文献[9]算法	算法1	文献[9]算法	算法1
1**	390.26	389.27	400	403	401	405	0.50	0.55
2	613.39	613.38	365	502	367	504	0.92	1.29
3*	1 698.01	1 850.54	1 220	752	1 223	755	2.69	1.70
4*	989.94	1 000.29	3 112	2 196	3 121	2 199	8.02	5.53
5*	4 073.22	4 173.17	338	230	342	234	1.05	0.78
6***	1 957.90	1 955.60	1 117	690	1 122	691	3.52	2.16
7	1 271.35	1 318.40	1 460	3 117	1 462	3 127	6.04	12.79
8***	2 362.27	2 210.72	4 262	3 948	4 279	3 952	29.50	26.19
9	876.58	1 282.71	1 495	3 895	1 500	3 900	7.66	19.71
10**	1 762.86	1 532.78	2 812	3 011	2 823	3 023	21.64	30.47
11	1 315.05	1 322.46	1 670	2 132	1 675	2 137	10.62	17.18
12*	3 351.97	3 367.10	5 247	3 362	5 267	3 370	45.92	28.72
13***	1 555.02	1 550.09	1 797	1 722	1 808	1 725	18.58	20.15
14	3 263.17	3 883.86	2 145	3 999	2 150	4 007	45.28	77.10
15**	4 021.12	3 883.86	1 616	3 518	1 623	3 526	51.06	106.65
16***	2 506.13	2 322.69	3 166	2 378	3 172	2 383	108.51	80.88
17**	3 622.68	3 557.78	1 069	1 925	1 076	1 928	42.03	73.43
18*	3 373.33	3 584.01	2 060	1 823	2 067	1 835	92.81	79.49

表 2　在文献[19]框架基础上文献[9]算法(GM_AOS)与本文改进算法(算法 1)在第 2 组数据下的实验结果

Table 2　Experimental results of algorithm (GM_AOS) in reference [9] and proposed improvement algorithm (alogorithm 1) based on framework of reference [19] under the second set of data

电路划分问题	目标函数值		迭代次数		函数值计算次数		运行时间/s	
	文献[9]算法	算法 1	文献[9]算法	算法 1	文献[9]算法	算法 1	文献[9]算法	算法 1
1***	428.72	379.68	652	452	658	453	0.87	0.58
2***	663.91	649.67	707	516	712	517	1.85	1.33
3***	1 850.62	1 824.87	1 167	1 158	1 170	1 159	2.53	2.52
4	993.25	1 154.71	1 650	2 888	1 657	2 891	4.22	7.27
5*	3 615.52	4 195.35	456	367	462	368	1.45	1.20
6***	1 993.42	1 992.46	967	756	969	757	3.09	2.40
7	1 270.00	1 332.71	1 703	1 751	1 715	1 753	6.98	7.23
8**	2 344.76	2 313.19	4 349	5 731	4 356	5 738	29.84	39.01
9	1 047.59	1 450.95	2 362	2 014	2 371	2 017	12.33	10.26
10***	1 775.31	1 768.59	3 315	2 121	3 324	2 126	26.98	16.96
11***	1 304.60	1 306.62	2 285	1 895	2 296	1 901	14.82	12.54
12***	3 493.99	3 419.91	2 900	2 649	2 917	2 655	24.55	22.36
13**	1 537.24	1 482.33	2 059	2 420	2 063	2 431	20.89	25.28
14**	3 255.90	2 591.85	2 320	5 813	2 330	5 828	50.31	124.25
15***	4 063.99	3 825.17	3 532	2 428	3 547	2 437	104.26	74.03
16	2 314.10	2 329.33	2 062	2 739	2 067	2 747	71.82	98.69
17**	3 608.90	3 575.50	1 121	1 482	1 124	1 486	47.46	57.47
18**	3 551.61	3 517.39	2 180	2 306	2 191	2 321	98.33	104.27

由表 1 和表 2 可见：标 *** 的电路划分问题改进算法的函数值和迭代次数均优于文献[9]算法的结果，占问题总数的 33.33%，标 ** 的电路划分问题改进算法的函数值结果优于文献[9]算法的结果，占问题总数的 25%；此外由于本文在图划分问题中处理的是大规模问题，故本文改进算法在函数值未改善的情况下减少了迭代次数，也节省了总划分问题的运行时间，用 * 标记这部分问题的结果，其数量占问题总数的 16.7%. 未标记的数据表示未改善的问题，占问题总数的 25%. 结合以上实验结果和数据分析可见，本文改进的近似最优梯度法的效率总体优于文献[9]近似最优梯度法.

综上所述，本文对近似最优梯度法进行了改进，并将其应用于图划分问题中. 首先，选取二次模型或锥模型对目标函数进行近似，使用修正的 BFGS 更新公式更新 Hesse 矩阵的近似矩阵，同时选取线性组合的形式作为标量矩阵，分析得到了对应的近似最优步长；其次，提出了一种改进的 ZH 非单调线搜索，结合得到的近似最优步长构建改进的近似最优梯度法；再次，对改进的近似最优梯度算法进行收敛性分析，并给出了收敛性证明；最后，选取图划分问题的两组数据进行对比实验，结果表明，改进的近似最优梯度法的性能优于原近似最优梯度法.

参　考　文　献

[1] CAUCHY A. Méthode Générale Pour la Résolution des Systéms Déquations Simultanées [J]. Comptes Rendus, 1847, 25: 536-538.

[2] BARZILAI J, BORWEIN J M. Two-Point Step Size Gradient Methods [J]. IMA J Numer Anal, 1988, 8(1): 141-148.

[3] RAYDAN M. The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem [J]. SIAM J Optimiz, 1997, 7(1): 26-33.

[4] DAI Y H, LIAO L Z. R-Linear Convergence of the Barzilai and Borwein Gradient Method [J]. IMA J Numer Anal, 2002, 22(1): 1-10.

[5] DAI Y H, YUAN J Y, YUAN Y X. Modified Two-Point Stepsize Gradient Methods for Unconstrained Optimization [J]. Comput Optim Appl, 2002, 22(1): 103-109.

[6] XIAO Y H, WANG Q Y, WANG D. Notes on the Dai-Yuan-Yuan Modified Spectral Gradient Method [J].

J Comput Appl Math, 2010, 234(10): 2986-2992.

[7] BIGLARI F, SOLIMANPUR M. Scaling on the Spectral Gradient Method [J]. J Optim Theory Appl, 2013, 158(2): 626-635.

[8] MILADINOVIĆ M, STANIMIROVIĆ P, MILJKOVIĆ S. Scalar Correction Method for Solving Large Scale Unconstrained Minimization Problems [J]. J Optim Theory Appl, 2011, 151(2): 304-320.

[9] LIU Z X, LIU H W. An Efficient Gradient Method with Approximate Optimal Stepsize for Large-Scale Unconstrained Optimization [J]. Numer Algor, 2018, 78(1): 21-39.

[10] ZHANG H C, HAGER W W. A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization [J]. SIAM J Optim, 2004, 14(4): 1043-1056.

[11] LIU Z X, LIU H W, DONG X L. An Efficient Gradient Method with Approximate Optimal Stepsize for the Strictly Convex Quadratic Minimization Problem [J]. Optimization, 2018, 67(3): 427-440.

[12] ZHANG J Z, DENG N Y, CHEN L H. New Quasi-Newton Equation and Related Methods for Unconstrained Optimization [J]. J Optimiz Theory Appl, 1999, 102(1): 147-167.

[13] WEI Z X, LI G Y, QI L Q. New Quasi-Newton Methods for Unconstrained Optimization Problems [J]. Appl Math Comput, 2006, 175(2): 1156-1188.

[14] HELMBERG C, RENDL F, WEISMANTEL R. Quadratic Knapsack Relaxations Using Cutting Planes and Semidefinite Programming [C]//International Conference on Integer Programming and Combinatorial Optimization. Berlin: Springer, 1996: 175-189.

[15] WIEGELE A, ZHAO S D. SDP-Based Bounds for Graph Partition via Extended ADMM [J]. Comp Optim Appl, 2022, 82(1): 251-291.

[16] 孙文瑜, 徐东. 解无约束最优化的基于锥模型的过滤集-信赖域方法 [J]. 中国科学: 数学, 2012, 42(5): 527-543. (SUN W Y, XU D. A Filter Trust-Region Method Based on Conic Model for Unconstrained Optimization [J]. Scientia Sinica: Mathematica, 2012, 42(5): 527-543.)

[17] LIU H W, LIU Z X, DONG X L. A New Adaptive Barzilai and Borwein Method for Unconstrained Optimization [J]. Optim Lett, 2018, 12(4): 845-873.

[18] LIU Z X, LIU H W. Several Efficient Gradient Methods with Approximate Optimal Stepsizes for Large Scale Unconstrained Optimization [J]. J Comput Appl Math, 2018, 328: 400-413.

[19] 王晓瑜, 刘红卫, 王婷, 等. 基于半定规划的多约束图划分问题 [J]. 吉林大学学报(理学版), 2023, 61(3): 540-546. (WANG X Y, LIU H W, WANG T, et al. Multi-constraint Graph Partitioning Problem Based on Semidefinite Programming [J]. Journal of Jilin University (Science Edition), 2023, 61(3): 540-546.)

[20] BUSTANY I, KAHNG A B, KOUTIS I, et al. SpecPart: A Supervised Spectral Framework for Hypergraph Partitioning Solution Improvement [C]//Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design. New York: Association for Computing Machinery, 2022: 13-1-13-9.

(责任编辑: 李 琦)