arXiv:2602.01152v1 [math.OC] 1 Feb 2026

弹性型张量极值M-特征值的高效记忆梯度法*

Zhuolin Du，Yisheng Song†重庆师范大学数学科学学院，中国重庆，401331。电子邮箱：duzhuolin728@163.com（Du）；yisheng.song@cqnu.edu.cn（Song）

摘要

四阶分层对称张量的M-特征值在非线性弹性材料分析和量子纠缠问题中发挥着重要作用。本文重点研究此类张量极值M-特征值的计算。为此，我们首先通过引入移位参数，将M-特征值问题重新表述为一系列无约束优化问题。随后，我们提出一种专门用于近似这些极值M-特征值的记忆梯度法。在此框架下，我们证明了所提方法的全局收敛性。最后，全面的数值实验验证了该方法的有效性和稳定性。

关键词：层次对称张量；M-特征值；记忆梯度法；全局收敛

AMS 2010 数学学科分类：90C23、65H17、15A18、90C30。

*本研究得到中国国家自然科学基金（项目编号：12171064）资助。

†通讯作者电子邮箱：yisheng.song@cqnu.edu.cn

1 引言

张量特征值问题在数值多重线性代数中发挥着关键作用 [6, 12,27, 32, 43, 50], ，并在磁共振成像（MRI） [5], 、谱超图理论 [34], 和自动控制 [16–18]等领域有重要应用。近年来，四阶分层对称张量的 M-特征值问题尤其受到广泛关注，这主要归因于其广泛的应用，包括非线性弹性材料分析 [12, 19, 20, 27, 29, 40, 42, 49] 以及物理学中的量子纠缠问题 [8, 10, 14, 45]。为进一步研究四阶分层对称张量 M-特征值的相关性质，Han-Dai-Qi [20] 和 Qi-Dai-Han [42] 提出了这类张量的 M-特征值和 M-特征向量的定义。

定义1.1 设 $\mathcal{A} = (a_{ijkl}) \in \mathbb{R}^{m\times n\times m\times n}$ 或 $\mathcal{A} = (a_{ikjl}) \in \mathbb{R}^{m\times m\times n\times n}$。若对所有$i$, $k \in [m]$, $j, l \in [n]$,

$$a_{ijkl} = a_{kjil} = a_{ilkj} = a_{klij},$$

其中 $[n] = \{1, 2, \ldots, n\}$，则称 $\mathcal{A}$为四阶分层对称张量（Helbig *[15, 26]*）。这种张量也称为弹性型张量（Backus *[4]*）。若对所有 $i, k \in [m]$, $j, l \in [n]$,

$$a_{ijkl} = a_{kjil} = a_{ilkj},$$

则称 $\mathcal{A}$ 为四阶部分对称张量（Qi-Dai-Han *[42]*）。

显然，当 a_{ijkl} =a_{ikjl} 时，四阶分层对称张量是对称张量；而一个 n维（$m = n$）四阶分层对称张量有$n(n + 1)(n^2 + n + 2)$ 8 个独立分量（Backus [4]）。当 $m = n = 3$ 时，四阶分层对称张量有 21 个独立分量，而四阶部分对称张量有 36 个独立分量。为简便起见，我们用 $\mathbb{H}^{m\times n\times m\times n}$ 表示所有四阶分层对称张量的集合。

定义 1.2（*[20, 42]*）设 $\mathcal{A} = (a_{ijkl}) \in \mathbb{H}^{m\times n\times m\times n}$。若（$\lambda$,$\mathbf{x}$, $\mathbf{y}$）$\in \mathbb{R} \times \mathbb{R}^m\backslash\{\mathbf{0}\} \times \mathbb{R}^n\backslash\{\mathbf{0}\}$ 满足以下方程组

$$\mathcal{A} \cdot \mathbf{yxy} = \lambda\mathbf{x}, \quad \mathcal{A}\mathbf{xyx}\cdot = \lambda\mathbf{y}, \quad \mathbf{x}^\top\mathbf{x} = 1, \quad \mathbf{y}^\top\mathbf{y} = 1, \tag{1.1}$$

其中

$$(\mathcal{A}\cdot \mathbf{yxy})_i = \sum_{k\in[m];\ j,l\in[n]} a_{ijkl}y_jx_ky_l \quad and \quad (\mathcal{A}\mathbf{xyx}\cdot)_l = \sum_{i,k\in[m];\ j\in[n]} a_{ijkl}x_iy_jx_k,$$

则称 λ 为 $\mathcal{A}$ 的 M-特征值，其左 M-特征向量为 x，右 M-特征向量为 y。

过去十年间，四阶分层对称张量的 M-特征值研究已成为一个活跃的研究领域 [7, 9, 24, 25, 35, 47, 48]。通过区间包含方法和谱近似技术，人们取得了显著进展 [11, 30]。然而，要准确刻画这些 M-特征值，仍面临重大的计算挑战。为此，Wang 等人 [47] 提出了一种计算方法，将最大 M-特征值问题重新表述为单位球面上的双二次齐次多项式优化问题；该方法还可用于验证强椭圆性。在此基础上，Li 等人 [35] 推导出了适用于四阶分层对称张量的两个专门的 M-特征值包含区间。随后，Che 等人 [7] 通过分析不同的 M-特征向量分量来估计 M-特征值，并为非负分层对称张量的 M-谱半径建立了上界。值得注意的是，这些上界被纳入 WQZ 算法 [47]，作为参数，展示了其在迭代方法中的实用性。M-特征值集的一项重要应用是验证正定性 [31–33]。然而，现有包含集 [7, 35]，的一项显著局限是其仅覆盖坐标轴附近的区域，因此无法有效判定 M-正定性。针对这一不足，近期研究 [24] 提出了新颖的 S 型 M-特征值包含定理。作为这些理论进展的补充，人们还开发了一种交替移位幂方法，用于计算四阶分层对称张量的极值 M-特征值 [46]。

除了交替移位幂方法等迭代方法外，近年来，针对张量特征值问题 [21–23, 28, 38]，还涌现出若干专门的优化策略。Han [21] 提出了一种无约束优化模型，用于计算对称张量的广义特征对；将 BFGS 方法应用于该模型，可得到超线性收敛序列，为此类问题提供了稳健的求解框架。在磁共振成像（MRI） [44], 等实际应用中，关注重点通常转向计算极端特征值或局部极大值，而非确定整个谱集，这凸显了对高效、目标明确的算法的需求。然而，随着张量阶数或维数增加，这些优化问题会变成大规模乃至超大规模问题，从而带来重大

计算挑战。为解决这一问题，Li 等人 [34] 提出了一种线性收敛自适应梯度法。这自然引出一个问题：自适应梯度法能否扩展到四阶分层对称张量？我们注意到，最速下降法 [13] 仅依赖当前梯度信息来确定下一个迭代点，可能未能充分利用历史数据。相比之下，Narushima等人 [37] 提出了记忆梯度法，通过纳入过去的信息来改进经典梯度下降法，同时避免海森矩阵计算。受这些见解启发，特别是考虑到历史梯度数据在结构化张量问题中的潜在价值尚未得到充分利用，我们提出了一种记忆梯度法（MGM），用于计算四阶分层对称张量的最大M特征值。我们证明了 MGM 的全局收敛性，并给出了全面的数值结果，展示其计算效率。

本文其余部分安排如下。第 2 节介绍一种变换，将 M-特征值问题重新表述为无约束优化框架。在此基础上，第 3 节提出用于计算极值 M-特征值的记忆梯度法（MGM），并证明其全局收敛性。第 4 节通过全面的数值实验，展示 MGM 相较于现有方法在计算效率和性能方面的优势。最后，第 5 节给出总结。

2 弹性型张量 M-特征值的无约束框架

设 $\mathbb{R}^n$ 为 n 维实欧几里得空间。通常，标量用小写字母表示，如 a、b；向量用小写粗体字母表示，如 $\mathbf{x}$、$\mathbf{y}$。此外，矩阵用大写字母表示，如 A、B；张量用花体大写字母表示，如 $\mathcal{A}$、$\mathcal{B}$。上标 $^\top$ 表示向量和矩阵的转置。设 $\mathbb{H}^{m\times n\times m\times n}$ 表示所有四阶层级对称张量的集合。对于任意 $\mathcal{A} = (a_{ijkl}) \in \mathbb{H}^{m\times n\times m\times n}$，相应的梯度如下：

$$\nabla_{\mathbf{x}}(\mathcal{A}\mathbf{xyxy}) = 2\mathcal{A}\cdot\mathbf{yxy}, \quad \nabla_{\mathbf{y}}(\mathcal{A}\mathbf{xyxy}) = 2\mathcal{A}\mathbf{xyx}\cdot.$$

为简便起见，$\mathcal{A}\cdot\mathbf{yxy}$ 和 $\mathcal{A}\mathbf{xyx}\cdot$ 也分别记作 $\mathcal{A}\mathbf{yxy}$ 和 $\mathcal{A}\mathbf{xyx}$。受 Auchmuty [1], 的工作启发，我们考虑以下无约束优化

问题，它与 M-特征值问题密切相关（见定理 2.1）：

$$\min_{\substack{\mathbf{x}\in\mathbb{R}^m\\ \mathbf{y}\in\mathbb{R}^n}} f(\mathbf{x},\mathbf{y}) = \frac{1}{4}(\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})^2 - \frac{1}{2}\mathcal{A}\mathbf{xyxy}. \tag{2.1}$$

于是，f(x, y) 关于 x 和 y 的梯度由下式给出

$$g_1(\mathbf{x},\mathbf{y}) = \nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{y}) = (\mathbf{x}^\top\mathbf{x})\mathbf{x}(\mathbf{y}^\top\mathbf{y})^2 - \mathcal{A}\mathbf{yxy}, \tag{2.2}$$

$$g_2(\mathbf{x},\mathbf{y}) = \nabla_{\mathbf{y}} f(\mathbf{x},\mathbf{y}) = (\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})\mathbf{y} - \mathcal{A}\mathbf{xyx}. \tag{2.3}$$

此外，我们有如下结果。

定理 2.1 设 $\mathcal{A} \in \mathbb{H}^{m\times n\times m\times n}$ 为非零张量。假设 λ^* 是 $\mathcal{A}$的最大M特征值。则我们有如下结果

(1) 问题 (2.1) 存在全局最小值。

(2) 设 $\mathbf{x} \in \mathbb{R}^m\backslash\{\mathbf{0}\}$与 $\mathbf{y} \in \mathbb{R}^n\backslash\{\mathbf{0}\}$为 (2.1) 的临界点。则 $\lambda = (\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y})$ 是 $\mathcal{A}$的一个正 M-特征值，其相应的 M-特征向量为 $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$ 与 $\mathbf{v} = \mathbf{y}/\|\mathbf{y}\|$。

(3) 若 $\lambda^* > 0$，则 f(x, y) 在非零临界点 (x,y) 处取得其全局最小值 $f_{\min} = -\frac{1}{4}(\lambda^*)^2$。

证明。(1) 由于单位球面是紧的，且函数 (u, v) $\mapsto |\mathcal{A}\mathbf{uvuv}|$ 连续，记 $C =$ 最大值$_{\|\mathbf{u}\|=1,\|\mathbf{v}\|=1}|\mathcal{A}\mathbf{uvuv}|$，并且最优值 C 可以取得，即存在单位向量 $\mathbf{u}_0$,$\mathbf{v}_0$ ，使得 $C = |\mathcal{A}\mathbf{u}_0\mathbf{v}_0\mathbf{u}_0\mathbf{v}_0|$。由于 $\mathcal{A}$ 是非零张量，因此有 $C > 0$。现在，我们根据 $\mathcal{A}\mathbf{uvuv}$ 的符号分两种情况讨论。令 $t = \|\mathbf{x}\|^2\|\mathbf{y}\|^2$，若 $t = 0$，则 f(x, y) $= 0$。否则，我们得到

$$f(\mathbf{x},\mathbf{y}) = \frac{1}{4}(\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})^2 - \frac{1}{2}\mathcal{A}\mathbf{xyxy} \geq \frac{1}{4}\|\mathbf{x}\|^4\|\mathbf{y}\|^4 - \frac{C}{2}\|\mathbf{x}\|^2\|\mathbf{y}\|^2 = \frac{1}{4}t^2 - \frac{C}{2}t.$$

通过直接计算可知，f(x, y)$\geq -\frac{1}{4}C^2$，且等号在 $t = C$时成立。(2) 根据 (2.2)-(2.3)，f (x, y) 的临界点满足

$$\mathcal{A}\mathbf{yxy} = (\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y})^2\mathbf{x} \text{ and } \mathcal{A}\mathbf{xyx} = (\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})\mathbf{y}. \tag{2.4}$$

记 $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$、$\mathbf{v} = \mathbf{y}/\|\mathbf{y}\|$。然后，结合（2.4），可得

$$\mathcal{A}\mathbf{vuv} = (\mathbf{x}^\top\mathbf{x})\frac{(\mathbf{y}^\top\mathbf{y})^2}{\|\mathbf{y}\|^2}\mathbf{u} \text{ and } \mathcal{A}\mathbf{uvu} = \frac{(\mathbf{x}^\top\mathbf{x})^2}{\|\mathbf{x}\|^2}(\mathbf{y}^\top\mathbf{y})\mathbf{v}.$$

根据定义 1.2，有 $\lambda = (\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y}) > 0$ 满足 $\mathcal{A}\mathbf{vuv} = \lambda\mathbf{u}$，$\mathcal{A}\mathbf{uvu} = \lambda\mathbf{v}$。因此，$\mathbf{u} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$ 和 $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ 是对应于正 M-特征值 λ的 M-特征向量。

(3) 设 λ 是 $\mathcal{A}$ 的任意一个正 M-特征值。根据定义 1.2，存在单位向量 u、v，使得 $\mathcal{A}\mathbf{vuv} = \lambda\mathbf{u}$ 且 $\mathcal{A}\mathbf{uvu} = \lambda\mathbf{v}$。令 $\mathbf{x} = p\mathbf{u}$，$\mathbf{y} = q\mathbf{v}$，其中 p、$q \in \mathbb{R} \setminus \{0\}$ 为某些标量。$f(\mathbf{x}, \mathbf{y})$ 的梯度推导如下

$$\begin{aligned}\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{y}) &= (\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y})^2\mathbf{x} - \mathcal{A}\mathbf{yxy} \\ &= p^2q^4p\mathbf{u} - \mathcal{A}(q\mathbf{v})(p\mathbf{u})(q\mathbf{v}) \\ &= p^3q^4\mathbf{u} - pq^2\mathcal{A}\mathbf{vuv} = p^3q^4\mathbf{u} - pq^2\lambda\mathbf{u} \\ &= pq^2(p^2q^2 - \lambda)\mathbf{u}\end{aligned}$$

$$\begin{aligned}\nabla_{\mathbf{y}} f(\mathbf{x},\mathbf{y}) &= (\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})\mathbf{y} - \mathcal{A}\mathbf{xyx} \\ &= p^4q^2q\mathbf{v} - \mathcal{A}(p\mathbf{u})(q\mathbf{v})(p\mathbf{u}) \\ &= p^4q^3\mathbf{v} - p^2q\mathcal{A}\mathbf{uvu} = p^4q^3\mathbf{v} - p^2q\lambda\mathbf{v} \\ &= p^2q(p^2q^2 - \lambda)\mathbf{v}.\end{aligned}$$

注意，$\mathbf{x}^\top\mathbf{x} = (p\mathbf{u})^\top(p\mathbf{u}) = p^2$，且 $\mathbf{y}^\top\mathbf{y} = (q\mathbf{v})^\top(q\mathbf{v}) = q^2$。选择满足$p^2q^2 = \lambda$的 p、q，可得 $\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) = 0$且 $\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) = 0$。因此，$(\mathbf{x}, \mathbf{y})$ 是 $f(\mathbf{x}, \mathbf{y})$ 的非零临界点，并满足 $(\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y}) = p^2q^2 = \lambda$。在这个临界点处，我们还有 $\mathcal{A}\mathbf{xyxy} = (\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})^2 = \lambda^2$。因此，

$$f(\mathbf{x},\mathbf{y}) = -\frac{1}{4}(\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})^2 = -\frac{1}{4}\lambda^2 \geq -\frac{1}{4}(\lambda^*)^2.$$

这意味着全局最小值是

$$f_{\min} = -\frac{1}{4}(\lambda^*)^2,$$

在对应于 $\lambda^* = (\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y})$的临界点处取得。 □

注 2.1 若 $\mathcal{A}$ 的所有 M-特征值均为非正，则 $f(\mathbf{x}, \mathbf{y})$ 不存在非零临界点。事实上，若存在非零临界点 (x,y)，则意味着 $\lambda = (\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y}) > 0$ 是一个 M-特征值。因此，在这种情况下，临界点集仅限于形如 (0,y) 或 (x,0) 的平凡解，而这些解不对应任何 M-特征值。

为处理 $\mathcal{A}$ 仅具有非正 M-特征值的情形，我们引入如下带参数 $t>0$ 的移位问题

$$\min_{\substack{\mathbf{x}\in\mathbb{R}^m\\ \mathbf{y}\in\mathbb{R}^n}} f_t(\mathbf{x},\mathbf{y})=\frac{1}{4}(\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})^2-\frac{1}{2}\mathcal{A}\mathbf{xyxy}-\frac{t}{2}(\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y}). \tag{2.5}$$

对于足够大的 t，任意非零向量 (x, y) 都满足 f_t(x, y) <0。另一方面，对于所有 x $\in\mathbb{R}^m$ 和 y $\in\mathbb{R}^n$，f_t(0, y) $=0$ 且 f_t(x, 0) $=0$。因此，形如 (0, y) 或 (x, 0) 的点都是 f_t(x, y) 的极大值点，而全局最小值严格为负，并在某个非零点处取得。f_t(x, y) 的梯度为

$$g_{t_1}(\mathbf{x},\mathbf{y})=\nabla_{\mathbf{x}}f_t(\mathbf{x},\mathbf{y})=(\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y})^2\mathbf{x}-\mathcal{A}\mathbf{yxy}-t(\mathbf{y}^\top\mathbf{y})\mathbf{x}, \tag{2.6}$$

$$g_{t_2}(\mathbf{x},\mathbf{y})=\nabla_{\mathbf{y}}f_t(\mathbf{x},\mathbf{y})=(\mathbf{x}^\top\mathbf{x})^2(\mathbf{y}^\top\mathbf{y})\mathbf{y}-\mathcal{A}\mathbf{xyx}-t(\mathbf{x}^\top\mathbf{x})\mathbf{y}. \tag{2.7}$$

显然，对于所有 x $\in\mathbb{R}^m$ 和 y $\in\mathbb{R}^n$，点 (0, y) 和 (x, 0) 都是 f_t(x, y) 的临界点。更重要的是，(2.5) 的任意非零临界点 (x, y) 都对应一个 M-特征值 $\lambda=$(x$^\top$x)(y$^\top$y) $-t$。由于 f_t(x, y) 在非零 (x, y) 处可取负值，因此适用于 (2.5) 的某种下降算法预计会收敛到一个非零临界点。

因此，通过求解原问题 (2.1) 或移位问题 (2.5)，可以有效地计算 M-特征值及其对应的 M-特征向量。以下算法概述了这一自适应过程。

算法1

步骤0。输入 $\mathcal{A}\in\mathbb{P}^{m\times n\times m\times n}$, $t\geq 1$, $\rho>1$ 和 $0<\epsilon\ll 1$, $k=0$。

步骤1。求解问题 (2.1)，得到 (x$_k$, y$_k$)。计算 $\lambda_k=$(x$_k{}^\top$x$_k$)(y$_k{}^\top$y$_k$)。

步骤2。若 $\min\{\|\mathbf{x}_k\|,\|\mathbf{y}_k\|\}>\epsilon$，则输出 (x$_k$, y$_k$) 和 λ_k，停止。否则转到步骤3。

步骤 3. 求解问题 (2.5)，得到 (x$_k$, y$_k$)。计算 $\lambda_k=$(x$_k{}^\top$x$_k$)(y$_k{}^\top$y$_k$) $-t$。

步骤 4. 若 $\min\{\|\mathbf{x}_k\|,\|\mathbf{y}_k\|\}>\epsilon$，则输出 (x$_k$, y$_k$) 和 λ_k，并停止。否则，更新 $t=\rho t$，并重复步骤 3。

注 2.2 步骤2与步骤3之间的迭代过程构成一个内循环。由于只要 t 足够大，对式 (2.5) 应用任何合适的下降算法都能保证收敛到一个非零临界点，因此该内循环必定会在有限步内终止。因此，算法1定义良好。

注 2.3 在实际应用中，应使用相同的无约束优化方法求解原问题 (2.1) 和移位问题 (2.5)。在下一节中，我们将介绍一种专为这类问题设计的记忆梯度法。

3 一种记忆梯度方法及其收敛性

本节介绍用于求解第2节所构建无约束优化问题的MGM。具体而言，在算法1的自适应框架中，MGM被用作步骤1和步骤3的主要求解器。

为简便起见，令 $\mathbf{z} = (\mathbf{x}^\top, \mathbf{y}^\top)^\top \in \mathbb{R}^{m+n}$ 和 $\Phi(\mathbf{z}) : \mathbb{R}^{m+n} \to \mathbb{R}$ 表示目标函数（即 $f(\mathbf{x}, \mathbf{y})$ 或 $f_t(\mathbf{x}, \mathbf{y})$）。相应地，令 $g(\mathbf{z}) = \nabla\Phi(\mathbf{z})$表示梯度。所提出方法的迭代更新方案如下：

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \alpha_k \mathbf{d}_k,$$

其中，$\alpha_k > 0$ 是通过线搜索过程确定的步长。对于 $k \geq 1$，搜索方向 $\mathbf{d}_k$ 按如下方式计算：

$$\mathbf{d}_k = -\gamma_k g(\mathbf{z}_k) + \frac{1}{N}\sum_{i=1}^{N} \beta_{k_i} \mathbf{d}_{k-i}, \tag{3.1}$$

其中，$\beta_{k_i} \in \mathbb{R}$ 适用于 $i = 1, \ldots, N$，而 $\gamma_k > 0$。对于 $k < N$，(3.1) 可解释为 $\mathbf{d}_k = -\gamma_k g(\mathbf{z}_k) + \frac{1}{N}\sum_{i=1}^{N} \beta_k \mathbf{d}_{k-i}$。初始搜索方向设为带有缩放参数 $\gamma_0 > 0$ 的最速下降方向，具体而言，$\mathbf{d}_0 = -\gamma_0 g(\mathbf{z}_0)$。请注意，这些参数不同于 Miele等人提出的参数 [36]。为确保搜索方向满足下降性质，我们将参数定义如下：

$$\beta_{k_i} = \|g(\mathbf{z}_k)\|^2 \varphi^\dagger_{k_i}, \quad \text{where } a^\dagger = \begin{cases} 0 & \text{if } a = 0, \\ \dfrac{1}{a} & \text{otherwise.} \end{cases} \tag{3.2}$$

这里，标量 φ_{k_i} 的选取满足

$$\begin{cases} \varphi_{k_1} > \max\left\{ \dfrac{g(\mathbf{z}_k)^\top \mathbf{d}_{k-1}}{\gamma_k},\ 0 \right\}, & i = 1, \\ \varphi_{k_i} \geq \max\left\{ \dfrac{g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}}{\gamma_k},\ 0 \right\}, & i = 2, \ldots, N. \end{cases} \tag{3.3}$$

需要注意的是， $a^\dagger a \leq 1$，由 (3.2) 和 (3.3) 可得，对于 $i = 2, \ldots, N$，有 $\beta_{k_1} > 0$ 和 $\beta_{k_i} \geq 0$。

基于这些性质，下面的定理证明了所提出的搜索方向具有下降性质。

定理 3.1 设搜索方向 $\mathbf{d}_k$ 由 (3.1) 定义。假设参数β_{k_i} 和 φ_{k_i} 的选取满足对所有 k均成立的条件 (3.2) 和 (3.3)。那么，对所有 k，下降条件均成立，即 $g(\mathbf{z}_k)^\top \mathbf{d}_k < 0$。

证明。对于 $k = 0$，有 $g(\mathbf{z}_0)^\top \mathbf{d}_0 = -\gamma_0 \|g(\mathbf{z}_0)\|^2 < 0$，这证明了初始步骤的下降性质。

对于 $k \geq 1$，我们考察以下表达式

$$-\gamma_k \|g(\mathbf{z}_k)\|^2 + \beta_{k_i} g(\mathbf{z}_k)^\top \mathbf{d}_{k-i},\ i = 1, \ldots, N.$$

情形 1： $i = 1$。由 (3.1) 和 (3.2)，并考虑到 β_{k_1}、 $\gamma_k > 0$，可得

$$\begin{aligned}
&-\gamma_k \|g(\mathbf{z}_k)\|^2 + \beta_{k_1} g(\mathbf{z}_k)^\top \mathbf{d}_{k-1} \\
&\leq -\gamma_k \|g(\mathbf{z}_k)\|^2 + \beta_{k_1} \max\left\{g(\mathbf{z}_k)^\top \mathbf{d}_{k-1}, 0\right\} \\
&= -\gamma_k \|g(\mathbf{z}_k)\|^2 + \|g(\mathbf{z}_k)\|^2 \varphi_{k_1}^\dagger \max\left\{g(\mathbf{z}_k)^\top \mathbf{d}_{k-1}, 0\right\} \\
&< -\gamma_k \|g(\mathbf{z}_k)\|^2 + \gamma_k \|g(\mathbf{z}_k)\|^2 \varphi_{k_1}^\dagger \varphi_{k_1} \\
&\leq 0.
\end{aligned}$$

严格不等式由 $i = 1$ 时的条件 (3.3) 得出，最后一个不等式则利用了性质 $a^\dagger a \leq 1$。

情况 2： $i = 2, \ldots, N$. 同理,

$$\begin{aligned}
&-\gamma_k \|g(\mathbf{z}_k)\|^2 + \beta_{k_i} g(\mathbf{z}_k)^\top \mathbf{d}_{k-i} \\
&\leq -\gamma_k \|g(\mathbf{z}_k)\|^2 + \beta_{k_i} \max\left\{g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}, 0\right\} \\
&= -\gamma_k \|g(\mathbf{z}_k)\|^2 + \|g(\mathbf{z}_k)\|^2 \varphi_{k_i}^\dagger \max\left\{g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}, 0\right\} \\
&\leq -\gamma_k \|g(\mathbf{z}_k)\|^2 + \gamma_k \|g(\mathbf{z}_k)\|^2 \varphi_{k_i}^\dagger \varphi_{k_i} \\
&\leq 0.
\end{aligned}$$

结合式（3.1），可得

$$\begin{aligned}
g(\mathbf{z}_k)^\top \mathbf{d}_k &= -\gamma_k \|g(\mathbf{z}_k)\|^2 + \frac{1}{N} \sum_{i=1}^{N} \beta_{k_i} g(\mathbf{z}_k)^\top \mathbf{d}_{k-i} \\
&= \frac{1}{N} \sum_{i=1}^{N} \left(-\gamma_k \|g(\mathbf{z}_k)\|^2 + \beta_{k_i} g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}\right) < 0.
\end{aligned} \tag{3.4}$$

因此，满足下降条件，所需结果成立。 □

值得注意的是，满足严格不等式 (3.3) 的指标不一定局限于 $i=1$。具体而言，只要存在至少一个指标 i 使得

$$\varphi_{k_i} > \max\left\{\frac{g(\mathbf{z}_k)\mathbf{d}_{k-i}}{\gamma_k}, 0\right\}.$$

How不过，采用最新的迭代信息是一个自然的选择。

接下来，我们着重证明充分下降条件，即

$$g(\mathbf{z}_k)^\top \mathbf{d}_k \le -c\|g(\mathbf{z}_k)\|^2, \quad \forall\, k \ge 1 \tag{3.5}$$

对于某个正常数 c。下面的定理表明，在对 γ_k、φ_{k_i}（$i=1$，.....，N）施加更强假设的情况下，所提出的方法满足条件 (3.5)。

定理 3.2 设搜索方向 $\mathbf{d}_k$ 由 (3.1) 定义。假设存在常数 $\bar{\gamma}>0$，使得对于所有 k，$\gamma_k \ge \bar{\gamma}$ 成立，并且选择 φ_{k_i}（$i=1$，.....，N）使其满足

$$\begin{cases} g(\mathbf{z}_k)^\top \mathbf{d}_{k-1} + \|g(\mathbf{z}_k)\|\|\mathbf{d}_{k-1}\| < \gamma_k \varphi_{k_1}, & (i=1), \\ g(\mathbf{z}_k)^\top \mathbf{d}_{k-i} + \|g(\mathbf{z}_k)\|\|\mathbf{d}_{k-i}\| \le \gamma_k \varphi_{k_i}, & (i=2,\ldots,N). \end{cases} \tag{3.6}$$

那么，所提出的方法满足充分下降条件 (3.5)。

证明。定义指标集 $\mathcal{I} = \{i \in \{1,\ldots,N\} \mid g(\mathbf{z}_k)^\top \mathbf{d}_{k-i} > 0\}$。令 $t = |\mathcal{I}|$。显然，$0 \le t \le N$。

注意，对于 $i \notin \mathcal{I}$，我们有 $g(\mathbf{z}_k)^\top \mathbf{d}_k \le 0$。由 (3.1)、(3.2) 和 (3.6) 可知，

$$\begin{aligned} g(\mathbf{z}_k)^\top \mathbf{d}_k &= -\gamma_k\|g(\mathbf{z}_k)\|^2 + \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}\varphi_{k_i}^{\dagger}\|g(\mathbf{z}_k)\|^2 \\ &\le -\gamma_k\|g(\mathbf{z}_k)\|^2 + \frac{1}{N}\sum_{i\in\mathcal{I}} \frac{g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}}{\varphi_{k_i}}\|g(\mathbf{z}_k)\|^2 \\ &\le -\gamma_k\|g(\mathbf{z}_k)\|^2 + \frac{1}{N}\sum_{i\in\mathcal{I}} \frac{\gamma_k g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}}{g(\mathbf{z}_k)^\top \mathbf{d}_{k-i} + \|g(\mathbf{z}_k)\|\|\mathbf{d}_{k-i}\|}\|g(\mathbf{z}_k)\|^2 \\ &\le -\gamma_k\|g(\mathbf{z}_k)\|^2 + \frac{1}{N}\sum_{i\in\mathcal{I}} \frac{\gamma_k g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}}{2g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}}\|g(\mathbf{z}_k)\|^2 \\ &= -\gamma_k\|g(\mathbf{z}_k)\|^2 + \frac{1}{N}\sum_{i\in\mathcal{I}} \frac{\gamma_k}{2}\|g(\mathbf{z}_k)\|^2 \\ &= -\gamma_k\|g(\mathbf{z}_k)\|^2\left(1 - \frac{t}{2N}\right). \end{aligned}$$

由于 $0 \le t \le N$，我们有 $1 - \frac{t}{2N} \ge \frac{1}{2}$。将此与 $\gamma_k \ge \bar{\gamma}$ 相结合，可得

$$g(\mathbf{z}_k)^\top \mathbf{d}_k \le -\frac{\gamma_k}{2}\|g(\mathbf{z}_k)\|^2 \le -\frac{\bar{\gamma}}{2}\|g(\mathbf{z}_k)\|^2,$$

这意味着充分下降条件 (3.5) 在 $c = \frac{\bar{\gamma}}{2}$ 下成立。 □

注意，条件 (3.5) 比 (3.4) 更强。现在，我们给出 MGM，其中步长 $\alpha_k > 0$ 由 Wolfe 线搜索条件确定

$$\Phi(\mathbf{z}_k + \alpha_k \mathbf{d}_k) - \Phi(\mathbf{z}_k) \le \rho \alpha_k g(\mathbf{z}_k)^\top \mathbf{d}_k, \tag{3.7}$$

$$g(\mathbf{z}_k + \alpha_k \mathbf{d}_k)^\top \mathbf{d}_k \ge \sigma g(\mathbf{z}_k)^\top \mathbf{d}_k, \tag{3.8}$$

其中 $0 < \rho < \sigma < 1$.

算法2. MGM

输入：函数 $\Phi(\mathbf{z})$（问题 (2.1) 或 (2.5)），移位参数 $t \ge 0$（设 $t = 0$ 为问题 (2.1)），$0 < \epsilon \ll 1$，$\gamma_0 > 0$，$N > 0$，$0 < \rho < \sigma < 1$。

步骤 0. 选取一个随机初始点 $\mathbf{z}_0 = (\mathbf{x}_0^\top, \mathbf{y}_0^\top)^\top \in \mathbb{R}^{m+n}$。计算 $g(\mathbf{z}_0)$。如果 $\|g(\mathbf{z}_0)\| \le \epsilon$，则停止；否则，令 $\mathbf{d}_0 = -\gamma_0 g(\mathbf{z}_0)$ 和 $k = 0$。

步骤 1. 确定步长 $\alpha_k > 0$，使其满足 (3.7) 和 (3.8)。

步骤 2. 更新 $\mathbf{z}_{k+1} = \mathbf{z}_k + \alpha_k \mathbf{d}_k$。如果 $\|g(\mathbf{z}_{k+1})\| \le \epsilon$，则令 $\mathbf{z}_k = \mathbf{z}_{k+1}$，停止。否则，计算 $\xi_k = \sqrt{\|\mathbf{y}_{k+1}\|/\|\mathbf{x}_{k+1}\|}$。更新 $\mathbf{x}_{k+1} \leftarrow \xi_k \mathbf{x}_{k+1}$ 和 $\mathbf{y}_{k+1} \leftarrow (1/\xi_k)\mathbf{y}_{k+1}$。前往步骤 3。

步骤 3. 计算满足式 (3.3) 的 γ_{k+1} 和 $\varphi_{(k+1)_i}$，然后根据式 (3.2) 计算 $\beta_{(k+1)_i}$。

步骤 4. 根据式 (3.1) 计算 $\mathbf{d}_{k+1}$。令 $k \leftarrow k+1$，并返回步骤 1。

输出：$\mathbf{z}_k = \left(\mathbf{x}_k{}^\top, \mathbf{y}_k{}^\top\right)^\top$。M-特征值 $\lambda = (\mathbf{x}_k{}^\top \mathbf{x}_k)(\mathbf{y}_K{}^\top \mathbf{y}_k) - t$；对应的左、右特征向量 $\mathbf{x}_k/\|\mathbf{x}_k\|$ 和 $\mathbf{y}_k/\|\mathbf{y}_k\|$。

注 3.1 值得注意的是，对于任意 $\xi \ne 0$，目标函数 $\Phi(\mathbf{z})$ 在变换 $\mathbf{x} \to \xi\mathbf{x}$ 和 $\mathbf{y} \to \xi^{-1}\mathbf{y}$ 下保持不变。因此，生成的序列$\{\mathbf{z}_k\}$ 理论上可能无界，其中 $\|\mathbf{x}\| \to \infty$ 且 $\|\mathbf{y}\| \to 0$。在算法2中，重新缩放步骤将迭代点限制为满足 $\|\mathbf{x}\| = \|\mathbf{y}\|$。此操作确保收敛性分析所需的紧性（见引理 3.1），同时不改变目标函数值或 M-特征向量的方向。

为证明所提方法的全局收敛性，我们首先建立若干关键引理。

引理 3.1 设 $\{\mathbf{z}_k\}$ 为 MGM 生成的序列。那么，序列 $\{\mathbf{z}_k\}$ 有界。具体而言，存在常数 η_1、$\eta_2 > 0$，使得对所有 $k \geq 0$，$\|\mathbf{x}_k\| \leq \eta_1$ 且 $\|\mathbf{y}_k\| \leq \eta_2$。

证明。由于 $\{\mathbf{z}_k\}$由算法2生成，下降性质保证对所有 k，$\Phi(\mathbf{z}_k) \leq \Phi(\mathbf{z}_0)$。此外，该算法确保 $\|\mathbf{x}_k\| = \|\mathbf{y}_k\|$。令 $s_k = \|\mathbf{x}_k\|\|\mathbf{y}_k\|$。由此可得 $\|\mathbf{x}_k\|^2 = \|\mathbf{y}_k\|^2 = s_k$，进而有 $\|\mathbf{z}_k\|^2 = 2s_k$。

回顾式 (2.5) 中定义的目标函数 $\Phi(\mathbf{z})$（当 $t = 0$ 时，该函数涵盖原问题）。

$$\Phi(\mathbf{z}_k) = \frac{1}{4}\|\mathbf{x}_k\|^4\|\mathbf{y}_k\|^4 - \frac{1}{2}\mathcal{A}\mathbf{y}_k\mathbf{x}_k\mathbf{y}_k - \frac{t}{2}\|\mathbf{x}_k\|^2\|\mathbf{y}_k\|^2.$$

由定理 2.1(1)，存在某个常数 $C > 0$，使得 $|\mathcal{A}\mathbf{x}_k\mathbf{y}_k\mathbf{x}_k\mathbf{y}_k| \leq C\|\mathbf{x}_k\|^2\|\mathbf{y}_k\|^2 = Cs_k^2$。因此，

$$\Phi(\mathbf{z}_k) \geq \frac{1}{4}s_k^4 - \frac{C}{2}s_k^2 - \frac{t}{2}s_k^2 = \frac{1}{4}s_k^4 - \frac{C+t}{2}s_k^2.$$

反设序列 $\{\mathbf{z}_k\}$ 无界。则存在一个子序列 $\{\mathbf{z}_{k_j}\}$，使得当 $j \to \infty$ 时，$\|\mathbf{z}_{k_j}\| \to \infty$，这等价于 $s_{k_j} \to \infty$。因此，我们有

$$\lim_{j\to\infty} \Phi(\mathbf{z}_{k_j}) \geq \lim_{s_{k_j}\to\infty} \left(\frac{1}{4}s_{k_j}^4 - \frac{C+t}{2}s_{k_j}^2\right) = +\infty.$$

这意味着，当 j, $\Phi(\mathbf{z}^{k_j}) > \Phi(\mathbf{z}_0)$ 充分大时，会导致矛盾。因此，序列 $\{\mathbf{z}_k\}$ 有界。于是，存在常数 $\eta_1, \eta_2 > 0$，使得对于所有 $k \geq 0$，$\|\mathbf{x}_k\| \leq \eta_1$ 且 $\|\mathbf{y}_k\| \leq \eta_2$。□

为进行后续的收敛性分析，定义

$$\mathcal{D} = \left\{\mathbf{z} = (\mathbf{x}^\top, \mathbf{y}^\top)^\top \in \mathbb{R}^{m+n} \mid \|\mathbf{x}\| \leq \eta_1, \|\mathbf{y}\| \leq \eta_2\right\}, \tag{3.9}$$

其中，η_1 和 η_2 是引理 3.1 中的常数。

引理3.2 若 $\mathcal{A} \in \mathbb{H}^{m\times n\times m\times n}$，则 $\mathcal{A}$yxy 和 $\mathcal{A}$xyx 在 $\mathcal{D}$ 上 Lipschitz 连续。

证明。令 $M = \max_{i,\ j,\ k,\ l} |a_{ijkl}|$。首先，我们证明 $\mathcal{A}\mathbf{yxy}$ 的利普希茨连续性。对于任意的 $(\mathbf{x}, \mathbf{y})$、$(\mathbf{x}', \mathbf{y}) \in \mathcal{D}$，有

$$\begin{aligned}
\|\mathcal{A}\mathbf{yxy} - \mathcal{A}\mathbf{yx}'\mathbf{y}\| &= \left\| \left(\sum_{k=1}^{m} \sum_{j,l=1}^{n} a_{ijkl}\, y_j\, (x_k - x'_k)\, y_l \right)_{i=1}^{m} \right\| \\
&\leq \sqrt{m} M \sum_{j=1}^{n} |y_j| \sum_{l=1}^{n} |y_l| \sum_{k=1}^{m} |x_k - x'_k| \\
&\leq \sqrt{m} M \left(\sqrt{n} \|\mathbf{y}\| \right) \left(\sqrt{n} \|\mathbf{y}\| \right) \left(\sqrt{m} \|\mathbf{x} - \mathbf{x}'\| \right) \\
&\leq mnM\eta_2^2 \|\mathbf{x} - \mathbf{x}'\| = P_1 \|\mathbf{x} - \mathbf{x}'\|,
\end{aligned} \tag{3.10}$$

其中 $P_1 = mnM\eta_2^2$ 是正常数。

接下来，对于 $(\mathbf{x}', \mathbf{y})$、$(\mathbf{x}', \mathbf{y}') \in \mathcal{D}$，利用不等式 $|ab - a'b'| \leq |a||b - b'| + |b'||a - a'|$，我们得到

$$\begin{aligned}
\|\mathcal{A}\mathbf{yx}'\mathbf{y} - \mathcal{A}\mathbf{y}'\mathbf{x}'\mathbf{y}'\| &= \left\| \left(\sum_{k=1}^{m} \sum_{j,l=1}^{n} a_{ijkl}\, x'_k\, (y_j y_l - y'_j y'_l) \right)_{i=1}^{m} \right\| \\
&\leq \sqrt{m} M \sum_{k=1}^{m} |x'_k| \sum_{j,l=1}^{n} \left(|y_j||y_l - y'_l| + |y'_l||y_j - y'_j| \right) \\
&\leq \sqrt{m} M \left(\sqrt{m} \eta_1 \right) \left(2\sqrt{n} \eta_2 \sqrt{n} \|\mathbf{y} - \mathbf{y}'\| \right) \\
&= 2mnM\eta_1\eta_2 \|\mathbf{y} - \mathbf{y}'\| = P_2 \|\mathbf{y} - \mathbf{y}'\|,
\end{aligned} \tag{3.11}$$

其中 $P_2 = 2mnM\eta_1\eta_2$ 是正常数。因此，$\mathcal{A}\mathbf{yxy}$ 关于 $\mathbf{x}$ 和 $\mathbf{y}$ 均为 Lipschitz 连续。

类似地，我们分析 $\mathcal{A}\mathbf{xyx}$。对于 $(\mathbf{x}, \mathbf{y}), (\mathbf{x}', \mathbf{y}) \in \mathcal{D}$，可得

$$\begin{aligned}
\|\mathcal{A}\mathbf{xyx} - \mathcal{A}\mathbf{x}'\mathbf{yx}'\| &= \left\| \left(\sum_{i,k=1}^{m} \sum_{j=1}^{n} a_{ijkl}\, y_j\, (x_i x_k - x'_i x'_k) \right)_{l=1}^{n} \right\| \\
&\leq \sqrt{n} M \sum_{j=1}^{n} |y_j| \sum_{i,k=1}^{m} \left(|x_k||x_i - x'_i| + |x'_i||x_k - x'_k| \right) \\
&\leq \sqrt{n} M \left(\sqrt{n} \eta_2 \right) \left(2\sqrt{m} \eta_1 \sqrt{m} \|\mathbf{x} - \mathbf{x}'\| \right) \\
&= P_2 \|\mathbf{x} - \mathbf{x}'\|.
\end{aligned} \tag{3.12}$$

对于 (x′, y), (x′, y′) $\in \mathcal{D}$，有

$$\begin{aligned}\|\mathcal{A}\mathbf{x}'\mathbf{y}\mathbf{x}' - \mathcal{A}\mathbf{x}'\mathbf{y}'\mathbf{x}'\| &= \left\| \left(\sum_{i,k=1}^{m} \sum_{j=1}^{n} a_{ijkl}\, x_i' \,(y_j - y_j')x_k' \right)_{l=1}^{n} \right\| \\ &\le \sqrt{n}M \sum_{i,k=1}^{m} |x_i' x_k'| \sum_{j=1}^{n} |y_j - y_j'| \\ &\le \sqrt{n}M \left(m\eta_1^2\right) \left(\sqrt{n}\|\mathbf{y} - \mathbf{y}'\|\right) \\ &= mnM\eta_1^2 \|\mathbf{y} - \mathbf{y}'\| = P_3\|\mathbf{y} - \mathbf{y}'\|,\end{aligned} \tag{3.13}$$

其中，$P_3 = mnM\eta_1^2$ 为正常数。由（3.10）—（3.13）可知，$\mathcal{A}$yxy 和 $\mathcal{A}$xyx 在 $\mathcal{D}$上均为 Lipschitz 连续函数。□

引理 3.3 设 $\mathcal{A} \in \mathbb{H}^{m\times n\times m\times n}$。则，移位问题（2.5）的梯度 g_t(z) 在 $\mathcal{D}$上是 Lipschitz 连续的。具体而言，存在正常数 L_t，使得对于所有 z，z′ $\in \mathcal{D}$，以下不等式成立

$$\|g_t(\mathbf{z}) - g_t(\mathbf{z}')\| \le L_t\|\mathbf{z} - \mathbf{z}'\|. \tag{3.14}$$

特别地，当 $t = 0$时，对于（2.1）同样有上述结论。

证明。首先，我们估计 g_{t_1}(x,y) 的 Lipschitz 常数，由 (2.6) 和引理 3.1 可知，对于所有 (x,y)，(x′,y′) $\in \Omega$，均有

$$\begin{aligned}&\|g_{t_1}(\mathbf{x},\mathbf{y}) - g_{t_1}(\mathbf{x}',\mathbf{y})\| \\ =&\|(\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y})^2\mathbf{x} - \mathcal{A}\mathbf{y}\mathbf{x}\mathbf{y} - t(\mathbf{y}^\top\mathbf{y})\mathbf{x} - (\mathbf{x}'^\top\mathbf{x}')(\mathbf{y}^\top\mathbf{y})^2\mathbf{x}' + \mathcal{A}\mathbf{y}\mathbf{x}'\mathbf{y} + t(\mathbf{y}^\top\mathbf{y})\mathbf{x}'\| \\ \le&\|\mathcal{A}\mathbf{y}\mathbf{x}\mathbf{y} - \mathcal{A}\mathbf{y}\mathbf{x}'\mathbf{y}\| + \|(\mathbf{x}^\top\mathbf{x})(\mathbf{y}^\top\mathbf{y})^2\mathbf{x} - (\mathbf{x}'^\top\mathbf{x}')(\mathbf{y}^\top\mathbf{y})^2\mathbf{x}'\| + \|t(\mathbf{y}^\top\mathbf{y})\mathbf{x}' - t(\mathbf{y}^\top\mathbf{y})\mathbf{x}'\| \\ \le&P_1\|\mathbf{x} - \mathbf{x}'\| + \|\mathbf{y}\|^4\|(\mathbf{x}^\top\mathbf{x})\mathbf{x} - (\mathbf{x}^\top\mathbf{x})\mathbf{x}' + (\mathbf{x}^\top\mathbf{x})\mathbf{x}' - (\mathbf{x}'^\top\mathbf{x}')\mathbf{x}'\| + |t|\|\mathbf{y}\|^2\|\mathbf{x} - \mathbf{x}'\| \\ \le&P_1\|\mathbf{x} - \mathbf{x}'\| + \|\mathbf{y}\|^4\left(\|\mathbf{x}\|^2\|\mathbf{x} - \mathbf{x}'\| + (\|\mathbf{x}\| + \|\mathbf{x}'\|)\|\mathbf{x}'\|(\|\mathbf{x}\| - \|\mathbf{x}'\|)\right) + |t|\|\mathbf{y}\|^2\|\mathbf{x} - \mathbf{x}'\| \\ \le&\left(P_1 + \eta_2^4(\eta_1^2 + \eta_1(\eta_1 + \eta_1)) + t\eta_2^2\right)\|\mathbf{x} - \mathbf{x}'\| \\ =&L_{11}\|\mathbf{x} - \mathbf{x}'\|\end{aligned}$$

其中 $L_{11} = P_1 + 3\eta_1^2\eta_2^4 + t\eta_2^2$ 是一个正常数。

$$
\begin{aligned}
&\|g_{t_1}(\mathbf{x}', \mathbf{y}) - g_{t_1}(\mathbf{x}', \mathbf{y}')\| \\
=&\|(\mathbf{x}'^{\top}\mathbf{x}')(\mathbf{y}^{\top}\mathbf{y})^2\mathbf{x}' - \mathcal{A}\mathbf{y}\mathbf{x}'\mathbf{y} - t(\mathbf{y}^{\top}\mathbf{y})\mathbf{x}' - (\mathbf{x}'^{\top}\mathbf{x}')(\mathbf{y}'^{\top}\mathbf{y}')^2\mathbf{x}' + \mathcal{A}\mathbf{y}'\mathbf{x}'\mathbf{y}' + t(\mathbf{y}'^{\top}\mathbf{y}')\mathbf{x}'\| \\
\leq&\|\mathcal{A}\mathbf{y}\mathbf{x}'\mathbf{y} - \mathcal{A}\mathbf{y}'\mathbf{x}'\mathbf{y}'\| + \|(\mathbf{x}'^{\top}\mathbf{x}')(\mathbf{y}^{\top}\mathbf{y})^2\mathbf{x}' - (\mathbf{x}'^{\top}\mathbf{x}')(\mathbf{y}'^{\top}\mathbf{y}')^2\mathbf{x}'\| + \|t(\mathbf{y}^{\top}\mathbf{y})\mathbf{x}' - t(\mathbf{y}'^{\top}\mathbf{y}')\mathbf{x}'\| \\
\leq&P_2\|\mathbf{y} - \mathbf{y}'\| + \|\mathbf{x}'\|^3\|(\mathbf{y}^{\top}\mathbf{y})^2 - (\mathbf{y}'^{\top}\mathbf{y}')^2\| + \|t\mathbf{x}'\|\|\mathbf{y}^{\top}\mathbf{y} - \mathbf{y}'^{\top}\mathbf{y}'\| \\
\leq&P_2\|\mathbf{y} - \mathbf{y}'\| + \|\mathbf{x}'\|^3(\|\mathbf{y}\|^2 + \|\mathbf{y}'\|^2)(\|\mathbf{y}\| + \|\mathbf{y}'\|)\|\mathbf{y} - \mathbf{y}'\| + t\|\mathbf{x}'\|\|\mathbf{y} + \mathbf{y}'\|\|\mathbf{y} - \mathbf{y}'\| \\
\leq&\left(P_2 + 4\eta_1^3\eta_2^3 + 2t\eta_1\eta_2\right)\|\mathbf{y} - \mathbf{y}'\| = L_{12}\|\mathbf{y} - \mathbf{y}'\|.
\end{aligned}
$$

因此，有

$$
\begin{aligned}
\|g_{t_1}(\mathbf{x}, \mathbf{y}) - g_{t_1}(\mathbf{x}', \mathbf{y}')\| &= \|g_{t_1}(\mathbf{x}, \mathbf{y}) - g_{t_1}(\mathbf{x}', \mathbf{y}) + g_{t_1}(\mathbf{x}', \mathbf{y}) - g_{t_1}(\mathbf{x}', \mathbf{y}')\| \\
&\leq \|g_{t_1}(\mathbf{x}, \mathbf{y}) - g_{t_1}(\mathbf{x}', \mathbf{y}')\| + \|g_{t_1}(\mathbf{x}', \mathbf{y}) - g_{t_1}(\mathbf{x}', \mathbf{y}')\| \qquad (3.15) \\
&\leq L_{11}\|\mathbf{x} - \mathbf{x}'\| + L_{12}\|\mathbf{y} - \mathbf{y}'\|.
\end{aligned}
$$

类似地，我们估计 $g_{t_2}(\mathbf{x}, \mathbf{y})$ 的 Lipschitz 常数。由 (2.7) 和引理 3.1 可得

$$
\begin{aligned}
&\|g_{t_2}(\mathbf{x}, \mathbf{y}) - g_{t_2}(\mathbf{x}', \mathbf{y})\| \\
=&\|(\mathbf{x}^{\top}\mathbf{x})^2(\mathbf{y}^{\top}\mathbf{y})\mathbf{y} - \mathcal{A}\mathbf{x}\mathbf{y}\mathbf{x} - t(\mathbf{x}^{\top}\mathbf{x})\mathbf{y} - (\mathbf{x}'^{\top}\mathbf{x}')^2(\mathbf{y}^{\top}\mathbf{y})\mathbf{y} + \mathcal{A}\mathbf{x}'\mathbf{y}\mathbf{x}' + t(\mathbf{x}'^{\top}\mathbf{x}')\mathbf{y}\| \\
\leq&\|\mathcal{A}\mathbf{x}\mathbf{y}\mathbf{x} - \mathcal{A}\mathbf{x}'\mathbf{y}\mathbf{x}'\| + \|(\mathbf{x}^{\top}\mathbf{x})^2(\mathbf{y}^{\top}\mathbf{y})\mathbf{y} - (\mathbf{x}'^{\top}\mathbf{x}')^2(\mathbf{y}^{\top}\mathbf{y})\mathbf{y}\| + \|t(\mathbf{x}^{\top}\mathbf{x})\mathbf{y} - t(\mathbf{x}'^{\top}\mathbf{x}')\mathbf{y}\| \\
\leq&P_2\|\mathbf{x} - \mathbf{x}'\| + \|\mathbf{y}\|^3\left(\|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2\right)(\|\mathbf{x}\| + \|\mathbf{x}'\|)\|\mathbf{x} - \mathbf{x}'\| + t\|\mathbf{y}\|(\|\mathbf{x}\| + \|\mathbf{x}'\|)\|\mathbf{x} - \mathbf{x}'\| \\
=&L_{21}\|\mathbf{x} - \mathbf{x}'\|,
\end{aligned}
$$

其中，$L_{21} = P_2 + 4\sqrt{n}\eta_1^3\eta_2^3 + 2t\eta_1\eta_2$ 是正常数。

$$
\begin{aligned}
&\|g_{t_2}(\mathbf{x}', \mathbf{y}) - g_{t_2}(\mathbf{x}', \mathbf{y}')\| = \|(\mathbf{x}'^{\top}\mathbf{x}')^2(\mathbf{y}^{\top}\mathbf{y})\mathbf{y} - \mathcal{A}\mathbf{x}'\mathbf{y}\mathbf{x}' - t(\mathbf{x}'^{\top}\mathbf{x}')\mathbf{y} - (\mathbf{x}'^{\top}\mathbf{x}')^2(\mathbf{y}'^{\top}\mathbf{y}')\mathbf{y}' + \mathcal{A} \\
&\mathbf{x}'\mathbf{y}'\mathbf{x}' + t(\mathbf{x}'^{\top}\mathbf{x}')\mathbf{y}'\| \leq \|\mathcal{A}\mathbf{x}'\mathbf{y}\mathbf{x}' - \mathcal{A}\mathbf{x}'\mathbf{y}'\mathbf{x}'\| + \|(\mathbf{x}'^{\top}\mathbf{x}')^2(\mathbf{y}^{\top}\mathbf{y})\mathbf{y} - (\mathbf{x}'^{\top}\mathbf{x}')^2(\mathbf{y}'^{\top}\mathbf{y}')\mathbf{y}'\| + \|t(\mathbf{x}'^{\top}\mathbf{x}') \\
&\mathbf{y} - t(\mathbf{x}'^{\top}\mathbf{x}')\mathbf{y}'\| \leq P_3\|\mathbf{y} - \mathbf{y}'\| + \|\mathbf{x}'\|^4\left(\|\mathbf{y}\|^2\|\mathbf{y} - \mathbf{y}'\| + (\|\mathbf{y}\| + \|\mathbf{y}'\|)\|\mathbf{y}'\|(\|\mathbf{y} - \mathbf{y}'\|)\right) + t\|\mathbf{x}'\|^2\| \\
&\mathbf{y} - \mathbf{y}'\| = L_{22}\|\mathbf{y} - \mathbf{y}'\|,
\end{aligned}
$$

其中，$L_{22} = P_3 + 3\eta_1^4\eta_2^2 + t\eta_1^2$ 是正常数。

因此，有

$$
\begin{aligned}
\|g_{t_2}(\mathbf{x}, \mathbf{y}) - g_{t_2}(\mathbf{x}', \mathbf{y}')\| &= \|g_{t_2}(\mathbf{x}, \mathbf{y}) - g_{t_2}(\mathbf{x}', \mathbf{y}) + g_{t_2}(\mathbf{x}', \mathbf{y}) - g_{t_2}(\mathbf{x}', \mathbf{y}')\| \\
&\leq \|g_{t_2}(\mathbf{x}, \mathbf{y}) - g_{t_2}(\mathbf{x}', \mathbf{y}')\| + \|g_{t_2}(\mathbf{x}', \mathbf{y}) - g_{t_2}(\mathbf{x}', \mathbf{y}')\| \qquad (3.16) \\
&\leq L_{21}\|\mathbf{x} - \mathbf{x}'\| + L_{22}\|\mathbf{y} - \mathbf{y}'\|.
\end{aligned}
$$

综合以上，令 $M_1 = \sqrt{2L_{11}^2 + 2L_{21}^2}$ 和 $M_2 = \sqrt{2L_{12}^2 + 2L_{22}^2}$。利用不等式 $(a+b)^2 \leq 2(a^2+b^2)$，可得

$$\begin{aligned}\|g_t(\mathbf{z}) - g_t(\mathbf{z}')\|^2 &= \|g_{t_1}(\mathbf{z}) - g_{t_1}(\mathbf{z}')\|^2 + \|g_{t_2}(\mathbf{z}) - g_{t_2}(\mathbf{z}')\|^2 \\ &\leq 2\left(L_{11}^2 + L_{21}^2\right)\|\mathbf{x} - \mathbf{x}'\|^2 + 2\left(L_{12}^2 + L_{22}^2\right)\|\mathbf{y} - \mathbf{y}'\|^2 \\ &\leq \max\left(M_1^2, M_2^2\right)\left(\|\mathbf{x} - \mathbf{x}'\|^2 + \|\mathbf{y} - \mathbf{y}'\|^2\right) \\ &= L_t^2\|\mathbf{z} - \mathbf{z}'\|^2.\end{aligned}$$

取平方根即可完成证明，其中 $L_t = \max(M_1, M_2)$。 □

基于上述引理，我们现在陈述以下著名引理，该引理由 Zoutendijk [53], 证明，适用于一般的迭代方法。

Le引理 3.4 设序列 $\{\mathbf{z}_k\}$ 和 $\{\mathbf{d}_k\}$ 由 MGM 生成。那么，

$$\sum_{k=0}^{\infty} \frac{\left(g(\mathbf{z}_k)^\top \mathbf{d}_k\right)^2}{\|\mathbf{d}_k\|^2} < \infty.$$

证明。 由 Wolfe 线搜索条件 (3.8) 和引理 3.3 可知,

$$(\sigma - 1)g(\mathbf{z}_k)^\top \mathbf{d}_k \leq \left(g(\mathbf{z}_{k+1}) - g(\mathbf{z}_k)\right)^\top \mathbf{d}_k \leq L_t \alpha_k \|\mathbf{d}_k\|^2,$$

这意味着

$$\alpha_k \geq \frac{(\sigma - 1)g(\mathbf{z}_k)^\top \mathbf{d}_k}{L\|\mathbf{d}_k\|^2}.$$

鉴于 $g(\mathbf{z}_k)^\top \mathbf{d}_k < 0$，由 (3.7) 可得

$$\Phi(\mathbf{z}_{k+1}) - \Phi(\mathbf{z}_k) \leq \rho\alpha_k g(\mathbf{z}_k)^\top \mathbf{d}_k \leq \frac{\rho(\sigma - 1)}{L_t} \frac{\left(g(\mathbf{z}_k)^\top \mathbf{d}_k\right)^2}{\|\mathbf{d}_k\|^2}. \tag{3.17}$$

由 (3.17) 可知，序列 $\{\Phi(\mathbf{z}_k)\}$ 是非增的。此外，根据定理 2.1， $\Phi(\mathbf{z}_k)$ 有下界。因此，$\{\Phi(\mathbf{z}_k)\}$ 收敛。对 (3.17) 两边取级数，可得

$$\sum_{k=0}^{\infty} \frac{\left(g(\mathbf{z}_k)^\top \mathbf{d}_k\right)^2}{\|\mathbf{d}_k\|^2} < \infty.$$

□

根据引理 3.4，我们建立了以下定理。

定理 3.3 设序列 $\{\mathbf{z}_k\}$ 由 MGM 生成。那么，该方法要么在驻点处终止，要么收敛，具体而言：

$$\liminf_{k\to\infty} \|g(\mathbf{z}_k)\| = 0 \tag{3.18}$$

证明。由 (3.3) 可得

$$\begin{aligned}\sum_{i=1}^{N} \left(\gamma_k\varphi_{k_i} - g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}\right)\beta_{k_i} &> (\gamma_k\varphi_{k_1} - \gamma_k\varphi_{k_1})\beta_{k_1} + \sum_{i=2}^{N} \left(\gamma_k\varphi_{k_i} - g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}\right)\beta_{k_i} \\ &\geq \sum_{i=2}^{N}(\gamma_k\varphi_{k_i} - \gamma_k\varphi_{k_i})\beta_{k_i} = 0.\end{aligned}$$

由于定理 3.1 的所有假设均已满足，搜索方向是下降方向，这意味着

$$g(\mathbf{z}_k)^\top \mathbf{d}_k < 0 \quad \text{for all } k.$$

因此，由 (3.1) 和 (3.2) 可得

$$\begin{aligned}\left|g(\mathbf{z}_k)^\top \mathbf{d}_k\right| &= -g(\mathbf{z}_k)^\top \mathbf{d}_k \\ &= \gamma_k\|g(\mathbf{z}_k)\|^2 - \frac{1}{N}\sum_{i=1}^{N} \beta_{k_i} g(\mathbf{z}_k)^\top \mathbf{d}_{k-i} \\ &= \frac{1}{N}\sum_{i=1}^{N} \left(\gamma_k\|g(\mathbf{z}_k)\|^2 - \beta_{k_i} g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} \left(\gamma_k\varphi_{k_i} - g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}\right)\beta_{k_i} > 0.\end{aligned} \tag{3.19}$$

将 (3.2) 改写为

$$\mathbf{d}_k + \gamma_k g(\mathbf{z}_k) = \frac{1}{N}\sum_{i=1}^{N} \beta_{k_i}\mathbf{d}_{k-i}.$$

两边平方并化简后

$$\|\mathbf{d}_k\|^2 = \left\|\frac{1}{N}\sum_{i=1}^{N} \beta_{k_i}\mathbf{d}_{k-i}\right\|^2 - 2\gamma_k g(\mathbf{z}_k)^\top \mathbf{d}_k - \gamma_k^2\|g(\mathbf{z}_k)\|^2.$$

两边除以 $\left(g(\mathbf{z}_k)^\top \mathbf{d}_k\right)^2$，并应用（3.19），可得

$$\begin{aligned}
\frac{\|\mathbf{d}_k\|^2}{\left(g(\mathbf{z}_k)^\top \mathbf{d}_k\right)^2} &= \frac{\left\|\frac{1}{N}\sum_{i=1}^N \beta_{k_i}\mathbf{d}_{k-i}\right\|^2}{\left(g(\mathbf{z}_k)^\top \mathbf{d}_k\right)^2} - 2\gamma_k \frac{g(\mathbf{z}_k)^\top \mathbf{d}_k}{\left(g(\mathbf{z}_k)^\top \mathbf{d}_k\right)^2} - \gamma_k^2 \frac{\|g(\mathbf{z}_k)\|^2}{(g(\mathbf{z}_k)^\top \mathbf{d}_k)^2} \\
&= \frac{\left\|\frac{1}{N}\sum_{i=1}^N \beta_{k_i}\mathbf{d}_{k-i}\right\|^2}{(g(\mathbf{z}_k)^\top \mathbf{d}_k)^2} - \frac{2\gamma_k}{g(\mathbf{z}_k)^\top \mathbf{d}_k} - \gamma_k^2 \frac{\|g(\mathbf{z}_k)\|^2}{(g(\mathbf{z}_k)^\top \mathbf{d}_k)^2} \\
&= \frac{\left\|\frac{1}{N}\sum_{i=1}^N \beta_{k_i}\mathbf{d}_{k-i}\right\|^2}{(g(\mathbf{z}_k)^\top \mathbf{d}_k)^2} - \left(\frac{1}{\|g(\mathbf{z}_k)\|} + \gamma_k \frac{\|g(\mathbf{z}_k)\|}{g(\mathbf{z}_k)^\top \mathbf{d}_k}\right)^2 + \frac{1}{\|g(\mathbf{z}_k)\|^2} \qquad (3.20) \\
&\le \frac{\left\|\frac{1}{N}\sum_{i=1}^N \beta_{k_i}\mathbf{d}_{k-i}\right\|^2}{(g(\mathbf{z}_k)^\top \mathbf{d}_k)^2} + \frac{1}{\|g(\mathbf{z}_k)\|^2} \\
&= \left(\frac{\frac{1}{N}\sum_{i=1}^N \beta_{k_i}\|\mathbf{d}_{k-i}\|}{\frac{1}{N}\sum_{i=1}^N \beta_{k_i}(\gamma_k\varphi_{k_i} - g(\mathbf{z}_k)^\top \mathbf{d}_{k-i})}\right)^2 + \frac{1}{\|g(\mathbf{z}_k)\|^2}.
\end{aligned}$$

注意到 $\gamma_k\varphi_{k_i} \ge g(\mathbf{z}_k)^\top \mathbf{d}_{k-i} + \|g(\mathbf{z}_k)\|\|\mathbf{d}_{k-i}\|$，并将其乘以 $\beta_{k_i} \ge 0$，可得

$$\beta_{k_i}\left(\gamma_k\varphi_{k_i} - g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}\right) \ge \beta_{k_i}\|g(\mathbf{z}_k)\|\|\mathbf{d}_{k-i}\|.$$

对 i 求和上述不等式，可得

$$\sum_{i=1}^N \beta_{k_i}\left(\gamma_k\varphi_{k_i} - g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}\right) \ge \|g(\mathbf{z}_k)\| \sum_{i=1}^N \beta_{k_i}\|\mathbf{d}_{k-i}\|.$$

由 (3.19) 可知 $\sum_{i=1}^N \beta_{k_i}(\gamma_k\varphi_{k_i} - g(\mathbf{z}_k)^\top \mathbf{d}_{k-i}) > 0$，因此可得

$$\frac{\sum_{i=1}^N \beta_{k_i}\|\mathbf{d}_{k-i}\|}{\sum_{i=1}^N \beta_{k_i}(\gamma_k\varphi_{k_i} - g(\mathbf{z}_k)^\top \mathbf{d}_{k-i})} \le \frac{1}{\|g(\mathbf{z}_k)\|}. \qquad (3.21)$$

由 (3.20) 和 (3.21)，可得

$$\frac{\|\mathbf{d}_k\|^2}{\left(g_1(\mathbf{x}_k, \mathbf{y}_k)^\top \mathbf{d}_k\right)^2} \le \frac{1}{\|g(\mathbf{z}_k)\|^2} + \frac{1}{\|g(\mathbf{z}_k)\|^2} = \frac{2}{\|g(\mathbf{z}_k)\|^2} \quad \text{for all } k. \qquad (3.22)$$

若 (3.18) 不成立，则存在常数 $c_1 > 0$，使得

$$\|g(\mathbf{z}_k)\| \ge c_1 \quad \text{for all } k. \qquad (3.23)$$

因此，由式 (3.22) 和 (3.23) 可得

$$\frac{(g(\mathbf{z}_k)^\top \mathbf{d}_k)^2}{\|\mathbf{d}_k\|^2} \ge \frac{c_1^2}{2}.$$

于是，我们得到

$$\sum_{k=0}^{\infty} \frac{\left(g(\mathbf{z}_k)^\top \mathbf{d}_k\right)^2}{\|\mathbf{d}_k\|^2} = \infty.$$

这与引理 3.4 矛盾，从而确认 (3.18) 成立。 □

该定理表明，对于满足条件 (3.6) 的任意选择的 $\gamma_k, \varphi_{k_i} (i = 1, \ldots, N)$，我们的方法都能实现全局收敛。

4 数值结果

本节通过一些计算实验，评估所提出方法在求解问题 (2.1) 或 (2.5) 时的有效性和稳定性。我们将 MGM 方法与 [47] 中的 SS-HOPM 以及 [46], 中的 SIPM 进行比较，据报道，这些方法在无约束优化中效率较高。所有数值实验均在 MATLAB 9.0 中，使用配备 AMD Ryzen 7 4800H CPU 2.90GHz 和 16 GB 随机存取存储器 (RAM) 的个人计算机完成。

如定理 3.2 所述，对于满足条件 (3.6) 的任意参数 γ_k, φ_{k_i}，该方法的全局收敛性均有保证。在实验中，我们首先选取 γ_k，然后为 $i = 1, \ldots, N$ 确定相应的 φ_{k_i} 值，使得 (3.6) 成立。对于所有方法，我们均设置 $\bar{\gamma} = 10^{-15}$。具体而言，我们采用了以下两种 γ_k 选择。

(i) 对所有 $\gamma_k = 1 \ \ k$;

(ii) $\gamma_0 = 1$，且对于 $k \geq 1$

$$\gamma_k = \begin{cases} 1, & \text{if } \dfrac{\mathbf{w}_{k-1}^\top \mathbf{s}_{k-1}}{\mathbf{w}_{k-1}^\top \mathbf{w}_{k-1}} < \bar{\gamma}, \\ \dfrac{\mathbf{w}_{k-1}^\top \mathbf{s}_{k-1}}{\mathbf{w}_{k-1}^\top \mathbf{w}_{k-1}}, & \text{otherwise.} \end{cases}$$

为阐明自适应规则，我们定义

$$\mathbf{s}_{k-1} = \mathbf{z}_k - \mathbf{z}_{k-1}, \mathbf{t}_{k-1} = g(\mathbf{z}_k) - g(\mathbf{z}_{k-1}).$$

修正向量定义为

$$\mathbf{w}_{k-1} = \mathbf{t}_{k-1} + \frac{\theta_{k-1}}{\mathbf{s}_{k-1}^\top \boldsymbol{\xi}_{k-1}},$$

其中， $\boldsymbol{\xi}_{k-1}$ 是满足 $\mathrm{s}_{k-1}^{\top}\boldsymbol{\xi}_{k-1} \neq 0$ 的任意向量，以确保除法有定义。依赖历史的修正系数为

$$\theta_{k-1} = 6(f(\mathbf{z}_k) - f(\mathbf{z}_{k-1})) + 3(g(\mathbf{z}_k) - g(\mathbf{z}_{k-1}))^{\top}\mathbf{s}_{k-1}.$$

（ii）中的 γ_k 选择，借鉴了 Zhang 等人 [51, 52]提出的修正割线条件定尺寸技术。给定 γ_k，我们对 $i = 1$、.....、 N定义 φ_{k_i} 如下

$$\varphi_{k_i} = \frac{\|g(\mathbf{z}_k)\|\|\mathbf{d}_{k-i}\| + g(\mathbf{z}_k)^{\top}\mathbf{d}_{k-i} + m + n}{\gamma_k}.$$

此处 φ_{k_i} 的选择满足条件（3.6）。我们将分别采用（i）和（ii）中的 γ_k 的 MGM 方法称为 MGM-1 和 MGM-2。

在 MGM 算法的实现中，我们将参数 $\epsilon = 10^{-6}$、 $\rho = 0.1$、 σ =设为 0.5。允许的最大外层迭代次数设为 2000。

4.1 对 γ_k 和 N的分析

为了评估 γ_k 对算法性能的影响，我们将迭代次数和 CPU时间作为关键指标。我们采用 Dolan 和 Mor$\acute{e}$ 提出的性能剖面 [13] 来分析 MGM 的性能。设 Y 和 W 分别为方法集和测试问题集， n_y、 n_w 分别为方法数和测试问题数。对于每个 $y \in Y$ 和 $w \in W$，性能剖面 ϕ： $\mathbb{R} \to [0, 1]$ 定义如下： $b_{w,y} > 0$ 表示方法 y求解问题 w 所需的迭代次数（或 CPU时间）。

此外，性能剖面由下式给出

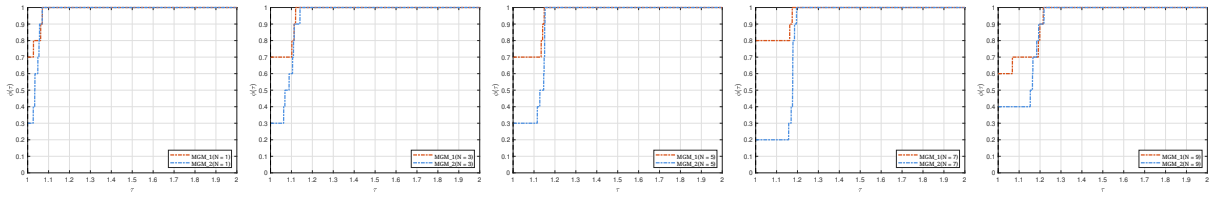
$$\phi_y(\tau) = \frac{1}{n_w} size\left\{w \in W : r_{w,y} \leq \tau\right\},$$

其中 $\tau > 0$，size$\{\cdot\}$ 表示集合中元素的个数，且 $r_{w,\ y}$ 是如下定义的性能比:

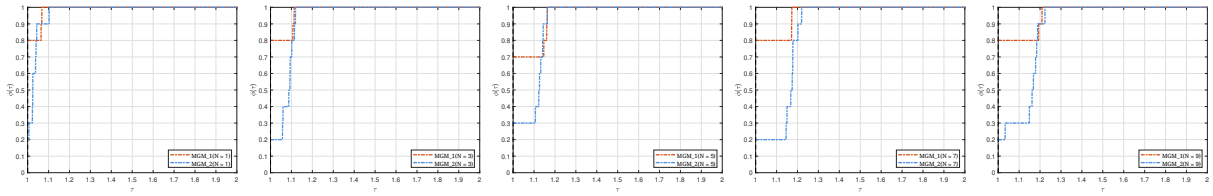
$$r_{w,y} = \frac{b_{w,y}}{\min\left\{b_{w,y} : y \in Y\right\}}.$$

因此，性能剖面通过绘制累积分布函数 ϕ_y 进行可视化。值得注意的是， ϕ_y(1) 表示该求解器优于所有其他求解器的概率。图像中 ϕ_y 的右侧显示了与求解器相关的稳健性。

为讨论 γ_k 的选择对算法性能的影响，我们在 N取值相同的情况下，基于迭代次数和 CPU时间对MGM-1与MGM-2进行比较。



(a) 迭代次数



(b) CPU时间

图 1：MGM-1 和 MGM-2 在相同 N 值下的性能曲线。

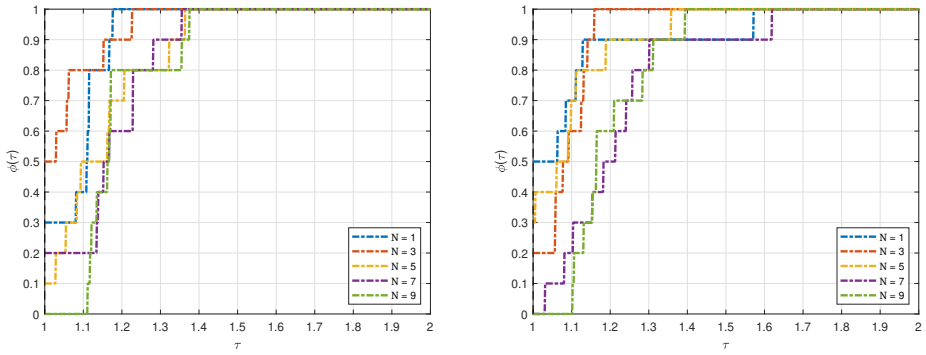
我们分别取 $N = 1$、3、5、7、9进行测试。图 1给出了基于迭代次数和CPU时间的性能曲线。结果表明，在 N取值相同的情况下，我们方法的性能会受到参数 γ_k 的影响。

为了研究 MGM-1 和 MGM-2 中 N 值的选择如何影响数值性能，我们使用 $N = 1$、3、5、7、9 进行了测试。图 2 和图 3 分别展示了 MGM-1 和 MGM-2 在迭代次数和CPU时间方面的性能剖面。结果表明，MGM-1 和 MGM-2 在不同范围内的性能会因所选 N 值而有所不同。因此，MGM-1 和 MGM-2 的性能取决于参数 N 的选择。尽管理论上很难确定最优选择，但我们的实验表明，MGM-1 采用 $N = 3$ 、MGM-2 采用 $N = 1$ 时，性能相对稳健。

4.2 所有对比算法的分析

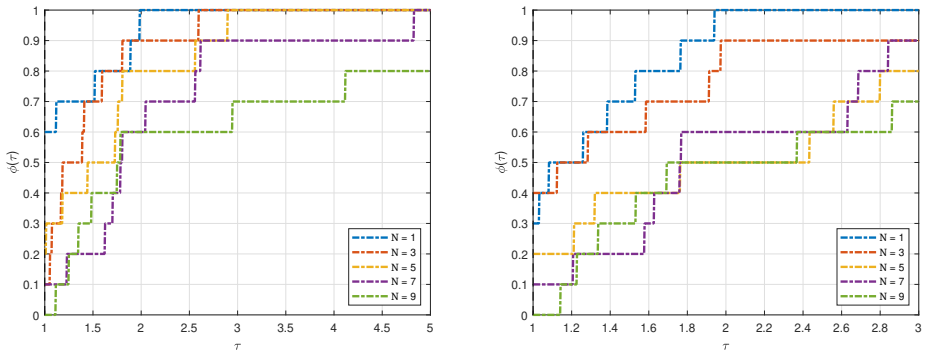
在本小节中，我们给出 SIPM、SS-HOPM、使用 $N = 3$ 的 MGM-1 以及使用 $N = 1$ 的 MGM-2 的结果，以验证我们方法的性能。

在表1中，“iter”表示迭代次数，“time”表示算法的平均CPU时间（以秒为单位），“λ^*”表示算法输出的最大M特征值。



(a) 迭代次数　　(b) CPU时间

图 2：不同 N 值下 MGM-1 的性能曲线。



(a) 迭代次数　　(b) CPU时间

图 3：MGM-2 在不同 N 值下的性能曲线。

示例4.1 设 $\mathcal{A} = (a_{ijkl}) \in \mathbb{H}^{2\times2\times2\times2}$ 为一个层次对称张量，其元素为

$$a_{1111} = 2, \quad a_{1211} = 3, \quad a_{2111} = 6, \quad a_{1121} = 6, \quad a_{1112} = 3, \quad a_{1212} = 2,$$
$$a_{2212} = 10, \quad a_{1222} = 10, \quad a_{2222} = 5,$$

以及其他 $a_{ijkl} = 0$。

例4.2 考虑层次对称张量 $\mathcal{A} = (a_{ijk\ell}) \in \mathbb{H}^{3\times3\times3\times3}$，其满足

$$
\begin{aligned}
&a_{2222} = a_{1111} = 196.6, \quad a_{3311} = a_{2233} = 83.2, \quad a_{2323} = a_{2322} = a_{3131} = a_{1331} = 54.7,\\
&a_{2233} = a_{2323} = -a_{1213} = -a_{2131} = -31.7, \quad a_{3333} = 110, \quad a_{2121} = a_{2121} = 64.4,\\
&a_{2321} = a_{1232} = -a_{1331} = -a_{1331} = -25.3, \quad a_{3112} = a_{1321} = 44.8,\\
&a_{2132} = a_{1223} = -35.84, \quad a_{1122} = 132.2,
\end{aligned}
$$

以及其他 $a_{ijk\ell} = 0$。

例4.3 考虑一个各项在 $(-1, 1)$ 上均匀生成的张量：

$$
\mathcal{A}(:,:,1,1) = \begin{pmatrix} -0.9727 & 0.3169 & -0.3437 \\ -0.6332 & -0.7866 & 0.4257 \\ -0.3350 & -0.9896 & -0.4323 \end{pmatrix},
$$

$$
\mathcal{A}(:,:,2,1) = \begin{pmatrix} -0.6332 & -0.7866 & 0.4257 \\ 0.7387 & 0.6873 & -0.3248 \\ -0.7986 & -0.5988 & -0.9485 \end{pmatrix},
$$

$$
\mathcal{A}(:,:,3,1) = \begin{pmatrix} -0.3350 & -0.9896 & -0.4323 \\ -0.7986 & -0.5988 & -0.9485 \\ 0.5853 & 0.5921 & 0.6301 \end{pmatrix},
$$

$$
\mathcal{A}(:,:,1,2) = \begin{pmatrix} 0.3169 & 0.6158 & -0.0184 \\ -0.7866 & 0.0160 & 0.0085 \\ -0.9896 & -0.6663 & 0.2559 \end{pmatrix},
$$

$$
\mathcal{A}(:,:,2,2) = \begin{pmatrix} -0.7866 & 0.0160 & 0.0085 \\ 0.6873 & 0.5160 & -0.0216 \\ -0.5988 & 0.0411 & 0.9857 \end{pmatrix},
$$

$$
\mathcal{A}(:,:,3,2) = \begin{pmatrix} -0.9896 & -0.6663 & 0.2559 \\ -0.5988 & 0.0411 & 0.9857 \\ 0.5921 & -0.2907 & -0.3881 \end{pmatrix},
$$

$$
\mathcal{A}(:,:,1,3) = \begin{pmatrix} -0.3437 & -0.0184 & 0.5649 \\ 0.4257 & 0.0085 & -0.1439 \\ -0.4323 & 0.2559 & 0.6162 \end{pmatrix},
$$

$$\mathcal{A}(:,:,2,3) = \begin{pmatrix} 0.4257 & 0.0085 & -0.1439 \\ -0.3248 & -0.0216 & -0.0037 \\ -0.9485 & 0.9857 & -0.7734 \end{pmatrix},$$

$$\mathcal{A}(:,:,3,3) = \begin{pmatrix} -0.4323 & 0.2559 & 0.6162 \\ -0.9485 & 0.9857 & -0.7734 \\ 0.6301 & -0.3881 & -0.8526 \end{pmatrix}.$$

表1：例4.1–4.3的数值结果。

示例	λ^*	MGM-1 迭代次数/时间	MGM-2 迭代次数/时间	SIPM 迭代次数/时间	SS-HOPM 迭代次数/时间
4.1	13.8616	5/0.0312	10/0.0899	11/0.0595	15/0.1286
4.2	318	11/0.0236	25/0.0521	34/0.0677	40/0.1459
4.3	2.3227	10/0.1233	29/0.2221	30/0.2767	35/0.3292

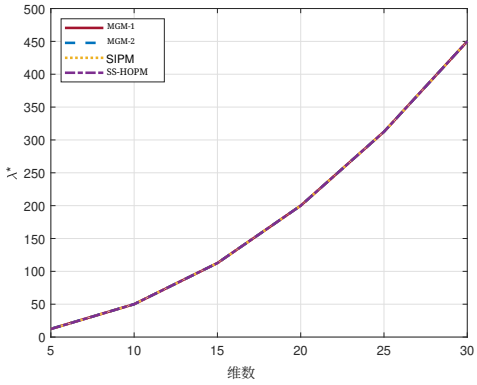
表1展示了 MGM-1 在计算例4.1–4.3的 M-特征值时的优越性能。与其他方法相比，MGM-1 的收敛速度显著更快，所需迭代次数更少，计算成本也更低。

例4.4 设 $\mathcal{A} \in \mathbb{R}^{n\times n\times n\times n}$ 是一个层次对称张量，其条目在 (0,1) 上均匀生成，即 $\mathcal{A} \sim U$ (0, 1)。

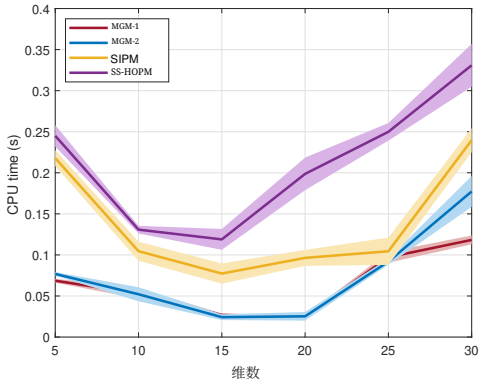
由于元素是随机生成的，我们为每种情况生成 10 组数据，以考察这四种方法的平均性能。具体而言，我们考察问题维数的六种情况，即 $n = \{5, 10, 15, 20, 25, 30\}$。所有方法均从相同的随机向量开始，这些向量取自标准正态分布。

图4展示了平均 M-特征值和 CPU时间，并以实线表示。相应颜色的阴影区域表示 10 次随机实验中数值的分布，阴影区域的宽度表示标准差。阴影区域越宽，说明该方法的性能越不稳定。图4中的两幅图充分表明 MGM-1 的表现相当稳定。

例 4.5 设 $\mathcal{A} \in \mathbb{R}^{m\times n\times m\times n}$ 为一个元素从 (-5,5) 中随机生成的层次对称张量。在此实验中，我们为 MGM-1 设置 $N = 3$，为 MGM-2 设置 $N = 1$。



(a) M-特征值



(b) CPU时间

图4：MGM-1、MGM-2、SIPM 和 SS-HOPM 的整体性能，包括 10 次随机测试的平均值和标准差。

结果见表 2，结果表明，在所有测试案例中，MGM-1 在效率和计算成本方面均优于其竞争方法。

表 2：例 4.5 的数值结果

		MGM-1	MGM-2	SIPM	SS-HOPM
(m,n)	λ^*	迭代次数/时间	迭代次数/时间	迭代次数/时间	迭代次数/时间
(12,18)	14.9986	13/0.2352	15/0.3576	21/0.5373	42/0.8522
(30,18)	18.9550	24/1.5623	33/1.2134	53/1.9543	84/2.3742
(30,30)	23.2705	43/1.7412	40/1.5431	72/2.5241	93/3.6652
(50,60)	30.0465	51/3.2582	58/3.4671	68/4.2521	153/6.3823

5 结论

本文通过将 M-特征值问题转化为无约束优化框架，重点研究了四阶层次对称张量的极值 M-特征值计算问题。基于一种新的重构方法，我们提出了一种记忆梯度法，用于求解该问题的近似解。此外，全局收敛

性得到了证明，且若干数值算例验证了该算法的性能。

参考文献

[1] Auchmuty G, 《对称特征值问题的全局快速收敛算法》, 《SIAM矩阵分析与应用杂志》, （1991年）, 12(4)：690-706。[2] Andrei N, 《一种用于无约束优化的缩放BFGS预条件共轭梯度算法》, 《应用数学快报》, （2007年）, 20(6)：645-650。[3] Andrei N, 《一种用于无约束优化的缩放无记忆BFGS预条件共轭梯度算法》, 《优化方法与软件》, （2007年）, 22(4)：561-571。[4] Backus G, 《各向异性弹性张量的几何图像》, 《地球物理学评论》, （1970年）, 8：633-671。[5] Bloy L, Verma R, 《在医学图像计算与计算机辅助干预中，根据扩散取向分布函数计算潜在纤维方向》, 施普林格：柏林/海德堡，德国, （2008年）, 第1-8页。[6] Chang K, Qi L, Zhou G, 《实矩形张量的奇异值》, 《数学分析及其应用杂志》, （2010年）, 370(1)：284-294。[7]Che H, Chen H, Wang Y, 《四阶部分对称张量的M-特征值估计》, 《工业与管理优化杂志》, （2020年）, 16(1)：309-324。[8] Chirita S, Danescu A, Ciarletta, 《各向异性线弹性材料的强椭圆性》, 《弹性学杂志》, （2007年）, 87：1-27。[9] Chen H, He H, Wang Y 等, 《一种高效的四次多项式优化交替最小化方法》, 《全球优化杂志》, （2022年）, 82：83-103。[10] Dahl G, Leinaas J, Myrheim J, 《量子物理中的张量积矩阵近似问题》, 《线性代数及其应用》, （2007年）, 420(2-3)：711-725。

[11] De L, De M, Vandewalle J, 《高阶张量的最佳秩-1及秩-($R_1, R_2, \ldots, R_N$)近似》, 《SIAM矩阵分析与应用杂志》, (2000年), 21(4): 1324-1342。[12] Ding W, Qi L, Wei Y, 《M张量与非奇异M张量》, 《线性代数及其应用》, (2013年), 439(10): 3264-3278。[13]Dolan E, Moré J, 《利用性能曲线对优化软件进行基准测试》, 《数学规划》, (2002年), 91(2): 201-213。[14] Einstein A, Podolsky B, Rosen N, 《量子力学对物理实在的描述是否可以认为是完备的?》, 《物理评论》, (1935年), 47(10): 777-780。[15] Fjær E, Holt R M, Rathore J S, 《地震各向异性》, 第3章: 弹性张量的表示与近似, 勘探地球物理学家协会(1996年)[16] Gao L, Cao Z, Wang G, 《带脉冲信号的离散时间非线性马尔可夫跳跃时滞系统的几乎必然稳定性》, 《非线性分析: 混杂系统》, (2019年), 34: 248-263。[17] Gao L, Luo F, Yan Z, 《脉冲切换系统的有限时间环形区域稳定性: 模态依赖参数方法》, 《国际控制杂志》, (2019年), 92(6): 1381-1392。[18] Gao L, Zhang M, Yao X, 《具有多重跳跃的脉冲切换随机非线性系统的随机输入到状态稳定性》, 《国际系统科学杂志》, (2019年), 50(9): 1860-1871。[19] Gurtin M, 《弹性线性理论》, 载于《物理学手册》, 施普林格, 柏林, (1972年)。[20] Han D, Dai H, Qi L, 《各向异性弹性材料强椭圆性的条件》, 《弹性学杂志》, (2009年), 97: 1-13。[21] Han L, 《一种求偶数阶对称张量实特征值的无约束优化方法》, 《数值代数、控制与优化》, 3 (2013年), 583-599。

[22]Hao C, Cui C, Dai Y, 《超对称张量极端 Z-特征值的序贯子空间投影法》, 《数值线性代数及其应用》, (2015年), 22(2): 283-298。[23]Hao C, Cui C, Dai Y, 《计算对称张量极端 Z-特征值的可行信赖域法》, (2015年)。[24] He J, Liu Y, Xu G, 《四阶部分对称张量的新 M-特征值包含集及其应用》, 《马来西亚数学科学学会通报》, (2021年), 44(6): 3929-3947。[25] He J, Li C, Wei Y, 《M-特征值区间及强椭圆性的可检验充分条件》, 《应用数学快报》, (2020年), 102: 106137。[26] Helbig K, 《弹性张量的表示与近似》, 作为第六届地震各向异性国际研讨会的海报 PI-1, 特隆赫姆, (1994年), 七月: 3-8。[27] Huang Z, Qi L, 《成对对称张量和弹性张量的正定性》, 《计算与应用数学杂志》, (2018年), 338: 22-43。[28] Hu S, Li G, Qi L, 等, 《通过平方和规划求本质非负对称张量的最大特征值》, 《优化理论与应用杂志》, (2013年), 158(3): 717-738。[29] He J, Xu G, Liu Y, 《关于弹性 M-张量最小 M-特征值的一些不等式》, 《工业与管理优化杂志》, (2017年), 13(5): 1-11。[30] Kofidis E, Regalia P, 《高阶超对称张量的最佳秩一逼近》, 《SIAM矩阵分析与应用杂志》, (2002年), 23(3): 863-884。[31] Li C, Li Y, 《双 B-张量与拟双 B-张量》, 《线性代数及其应用》, (2015年), 466: 343-356。[32] Li C, Li Y, Kong X, 《张量的新特征值包含集》, 《数值线性代数及其应用》, (2014年), 21(1): 39-50。[33] Li C, Wang F, Zhao J, 《实超对称张量正定性的判据》, 《计算与应用数学杂志》, (2014年), 255: 1-14。

[34] Li G, Qi L, Yu G, 《对称张量的 Z-特征值及其在谱超图理论中的应用》, 《数值线性代数及其应用》, (2013年), 20, 1001-1029。[35] Li S, Li C, Li Y, 《四阶部分对称张量的 M-特征值包含区间》, 《计算与应用数学杂志》, (2019年), 356：391-401。[36] Miele A, Cantrell J, 《函数最小化的记忆梯度法研究》, 《优化理论与应用杂志》, (1969年), 3(6)：459-470。[37] Narushima Y, Yabe H, 《无约束优化中记忆梯度法的全局收敛性》, 《计算优化与应用》, (2006年), 35(3)：325-346。[38] Ni Q, Qi L, 《求非负齐次多项式映射最大特征值的二次收敛算法》, 《全球优化杂志》, (2015年), 61(4)：627-641。[39] Nocedal J, Wright S, 《数值优化》, 纽约, NY：Springer New York, (1999年)。[40] Padovani C, 《横观各向同性弹性张量的强椭圆性》, 《力学》, (2002年), 37(6)：515-525。[41] Peressini A, Sullivan F E, Uhl J J, 《非线性规划数学》, 纽约：Springer-Verlag, (1988年)。[42] Qi L, Dai H, Han D, 《强椭圆性和M-特征值条件》, 《中国数学前沿》, (2009年), 4(2)：349-364。[43] Qi L, Luo Z, 《张量分析：谱理论与特殊张量》, 工业与应用数学学会, (2017年)。[44] Qi L, Yu G, Wu E, 《高阶半正定扩散张量成像》, 《SIAM成像科学杂志》, (2010年), 3(3)：416-433。[45]Schrödinger E, 《量子力学中的当前情形》, 《自然科学》, (1935年), 23(50)：844-849。

[46] Wang C, Chen H, Wang Y, 《四阶部分对称张量极端特征值的交替移位逆幂法》, 《应用数学快报》, （2023年）, 141：108601。[47] Wang Y, Qi L, Zhang X, 《计算四阶部分对称张量最大M特征值的一种实用方法》, 《应用数值线性代数》, （2009年）, 16(7)：589-601。[48] Wang G, Sun L, Liu L, 《四阶部分对称张量正定性的基于M特征值的充分条件》, 《复杂性》2020, （2020年）, 3：1-8。[49] Walton J, Wilber J, 《一类各向异性材料强椭圆性的充分条件》, 《国际非线性力学杂志》, （2003年）, 38(4)：441-455。[50] Yang Q, Yang Y, 《非负张量佩龙-弗罗贝尼乌斯定理的进一步结果 II》, 《SIAM矩阵分析与应用杂志》, （2011年）, 32(4)：1236-1250。[51] Zhang J, Xu C, 《采用修正拟牛顿方程的拟牛顿法的性质与数值性能》, 《计算与应用数学杂志》, （2001年）, 137(2)：269-278。[52] Zhang J, Deng N, Chen L, 《无约束优化的新拟牛顿方程及相关方法》, 《优化理论与应用杂志》, （1999年）, 102(1)：147-167。[53] Zoutendijk G, 《非线性规划》, 《整数与非线性规划中的计算方法》, J. Abadie（编）, 北荷兰出版社, 阿姆斯特丹, （1970年）：37-86。