

求解对称非负逆特征值问题的一阶黎曼算法

吕文平, 王湘美*

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2022年9月24日; 录用日期: 2022年10月17日; 发布日期: 2022年10月26日

摘要

基于对称非负逆特征值问题的黎曼优化模型, 提出了求解该问题的黎曼梯度下降算法和黎曼共轭梯度算法, 并分析了这两种算法的收敛性。通过数值实验, 比较了两种算法的收敛效率。数值结果表明, 求解高阶的非负逆特征值问题时, 黎曼共轭梯度算法的收敛效率要高于黎曼梯度下降算法。

关键词

对称非负逆特征值问题, 黎曼梯度下降算法, 黎曼共轭梯度算法, 收敛性

Riemannian Algorithms of Order One for Solving Symmetric Nonnegative Inverse Eigenvalue Problems

Wenping Lv, Xiangmei Wang*

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Sep. 24th, 2022; accepted: Oct. 17th, 2022; published: Oct. 26th, 2022

Abstract

The Riemannian gradient descent algorithm and the Riemannian conjugate gradient algorithm are proposed for solving the symmetric nonnegative inverse eigenvalue problems, respectively, together with the convergence analysis. At last, some numerical experiments are provided to illustrate the convergence of the algorithms, and they show that the Riemannian conjugate gradient algorithm is more effective than the Riemannian gradient descent algorithm.

Keywords

Symmetric Nonnegative Inverse Eigenvalue Problems, Riemannian Gradient Descent Algorithm,

*通讯作者。

Riemannian Conjugate Gradient Algorithm, Convergence

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

矩阵中所有元素都是非负实数的矩阵称为非负矩阵。非负矩阵理论一直是矩阵理论中重要的研究领域之一,并且在数学和自然科学的其它分支以及社会科学中都被广泛涉及,例如博弈论、Markov 链、概率论、数值分析、离散分布、群论、小振荡弹性系统和经济学等。经典的非负逆特征值问题是已知矩阵的全体特征值,求非负矩阵(简称 NIEP)。近来,白正简老师等人研究了已知矩阵部分特征值的非负逆特征值问题[1] [2] [3]。特别地,若所求非负矩阵是对称的,则称为对称非负逆特征值问题(简称 SNIEP);若所求非负矩阵是随机矩阵(行和为一的非负矩阵),则称为对称随机逆特征值问题(简称 StNIEP)。NIEP 在物理学、控制设计、结构动力学、博弈论、马尔可夫链、概率算法、数值分析、离散分布,分类数据,群论和经济学等领域有广泛的应用[4] [5] [6]。自从 1937 年数学家 Kolmogorov 提出矩阵非负逆特征值问题以来,一直备受关注。关于 NIEP 的研究主要是两个方面,一方面关于解的理论研究,包括解的存在性和唯一性,另一方面算法研究,即研究求解各类非负逆特征值问题的数值算法研究。关于第一个方面的研究,虽然取得了一些重要的研究成果,至今仍未彻底解决。1949 年数学家 Sulemanova 提出了矩阵理论第二大猜想:给定一个封闭的共轭 n 元复数组 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$,能否找出使得此 n 元复数组是某个非负矩阵谱的充分必要条件。文献[4] [7] [8] [9] [10]研究了建立当矩阵的阶数小于等于 5 阶时,有解的充分必要条件;当阶数高于 5 阶时,文献[9] [10]研究了有解的充分条件。其中 Perron-Frobenius 定理的运用,使得这个问题的解决取得了很大的进步;关于解的存在性比较完整的研究结果可以参考 M.T. Chu 和 G.H. Golub 合著的文献[11]。

本文关注的重点是 NIEP 的数值算法。关于已知部分特征值的 NIEP,白正简老师等人在[2]中提出了非光滑牛顿法,并证明了算法局部收敛,且算法收敛时收敛速度是二阶的。进一步,白正简老师等人在文献[1]中将这类问题转化为黎曼流形上的优化问题,提出了共轭梯度算法,并证明了算法的收敛性。最近,杨丹在[12]中,通过这类问题转化为凸可行性问题,提出了交替投影算法,并证明了算法线性收敛。关于经典的 NIEP,即已知矩阵的全体特征值,求非负矩阵,Oris 等人在文献[8]中将非负逆特征值问题转化为求解两个集合交集的可行性问题,从而提出了交替投影算法,并在一定条件下建立了算法的收敛性。受这一思想的启发,党婵娟在文献[13]中研究了 StNIEP。通过将该问题转化为求解可行性问题,提出了交替投影算法和 Average 交替投影算法,并借助黎曼流形上投影算子的性质研究了算法的收敛性和收敛速度。值得一提的是,Chu 等人在文献[7]将 SNIEP 转化为黎曼流形上的优化问题,通过建立相应变量的微分方程,提出了“梯度流”算法。该问题本质上是 SNIEP 转化为微分方程来求解,当文中的 ω 极限点满足一定条件时,算法收敛(详见[7])。最近,Chu 等人在[11]运用类似的技术研究了一般 NIEP 的梯度流算法,同时运用到求解 StNIEP。

自从 1972 年 Luenberger 提出黎曼流形上沿测地线的梯度下降算法[14],黎曼流形上的优化算法的研究受到了很大的关注。其他黎曼流形上优化算法的研究,包括更一般的梯度算法,信赖域算法,牛顿法,次梯度法等见[15] [16]-[21]及其参考文献。正如这些文献中指出,运用黎曼流形技术求解欧氏空间的优化

问题, 不论在理论上还是实际应用中都有重要的意义。其主要优点包括可将欧氏空间中的约束优化问题转化为无约束优化问题; 将病态问题转化为非病态问题; 将非凸优化问题转化为凸优化问题。本文在第三节考虑的优化问题就运用把欧氏空间中的约束优化问题转化为无约束优化问题这一优点。受到 *Chu* 等人在文献[7]研究的启发, 基于该文献提出的黎曼优化模型, 我们分别提出黎曼梯度算法和黎曼共轭梯度算法求解 *SNIEP*。通过对 *SNIEP* 的分析, 分别建立了两类算法的全局收敛性(算法的收敛性不依赖于初始点的选取)。正如我们上面指出, 已有求解 *SNIEP* 的算法, 均不具备全局收敛性。文献[7]中的梯度流算法当 ω 极限点满足一定条件时收敛, 而文献[8]中交替投影算法在初始点满足一定条件下收敛。最后通过数值例子, 我们验证了算法的收敛性。在数值实验中, 将 *Stiefel* 流形上的拉回映射和平移映射应用到算法中, 解决了乘积黎曼流形上的拉回映射和平移映射的计算问题。通过数值例子可以看出, 求解高阶的 *SNIEP* 时, 黎曼共轭梯度算法的收敛效率高于黎曼梯度算法。

本文结构如下。在下一节, 我们将介绍一些黎曼流形上的相关知识。第三节分别提出来求解 *SNIEP* 的黎曼梯度下降算法和黎曼共轭梯度算法。第四节, 分析得到了算法的全局收敛性。最后, 通过第五节的数值实验, 验证并比较了两种算法收敛性。

2. 预备知识

本节主要介绍黎曼流形上的一些基本的概念和结论, 它们可以在黎曼流形的书上找到, 例如[15] [22] [23]。

设 \mathcal{M} 是一个 Hausdorff 拓扑空间, 若 \mathcal{M} 的每一个点 x 都有一个开邻域 $U \subset \mathcal{M}$, 使得 U 和 n 维欧氏空间 \mathbb{R}^n 中的一个开子集是同胚的, 则称 \mathcal{M} 是一个 n 维拓扑流形。若 \mathcal{M} 上有 C^∞ 微分结构, 则称 \mathcal{M} 为 C^∞ 微分流形(或光滑流形)。以下我们总假设 \mathcal{M} 是光滑流形。设 $x \in \mathcal{M}$, 则在点 x 处的所有切向量组成的集合称为流形 \mathcal{M} 在点 x 处的切空间, 记为 $T_x\mathcal{M}$, 而 $T\mathcal{M} := \bigcup_{x \in \mathcal{M}} T_x\mathcal{M}$ 称为切丛。设 g 是 \mathcal{M} 上的一个光滑的二阶协变张量场。如果 g 是对称、正定的, 则称 g 为 \mathcal{M} 上的黎曼度量。若在 \mathcal{M} 上指定一个黎曼度量 g , 则称 (\mathcal{M}, g) 为一个黎曼流形, 简记为 \mathcal{M} 。本文也用符号 $\langle \cdot, \cdot \rangle$ 来表示在光滑流形上指定的黎曼度量 g 。在 \mathcal{M} 上存在唯一一个与度量 g 相容的黎曼联络 D 。设 $I \subset \mathbb{R}$ 是 \mathbb{R} 中一个区间, $\gamma: I \rightarrow \mathcal{M}$ 为 \mathcal{M} 上一条光滑

曲线, 如果满足 $\frac{D}{dt} \left(\frac{d\gamma}{dt} \right) = 0$, 则称其为 \mathcal{M} 上一条测地线。设 \mathcal{M} 是完备的黎曼流形, 则对任意 $x \in \mathcal{M}$,

任意 $\xi \in T_x\mathcal{M}$, 则存在唯一测地线 $\gamma = \gamma(t; x, \xi): \mathbb{R} \rightarrow \mathcal{M}$, 满足:

$$\gamma(0) = x, \quad \left. \frac{d\gamma}{dt} \right|_{t=0} = \xi。$$

设 $x \in \mathcal{M}$, 定义点 x 处的指数映射 $\text{Exp}_x: T_x\mathcal{M} \rightarrow \mathcal{M}$ 为:

$$\text{Exp}_x(\xi) = \gamma(1; x, \xi) \quad \forall \xi \in T_x\mathcal{M}。$$

我们知道欧氏空间 \mathbb{R}^n 作为特殊的黎曼流形, 其每一点处的切空间都是它自己。 \mathbb{R}^n 中点和点以及点和切向量之间可以直接进行线性运算。但是一般的黎曼流形上不存在这样的线性运算, 为进行类似的运算, 我们需要引入拉回映射、平移映射作等。以下拉回映射的概念是指数映射的推广, 它将切空间 $T_x\mathcal{M}$ 上的点一一映射到流形 \mathcal{M} 上, 这一概念最先由 Adler [24] 提出。

定义 2.1 称流形 \mathcal{M} 上的光滑映射 $\mathcal{R}: T\mathcal{M} \rightarrow \mathcal{M}$ 是拉回映射, 如果 \mathcal{R} 在 $T_x\mathcal{M}$ 上的映射 \mathcal{R}_x 满足:

- 1) $\mathcal{R}_x(0_x) = x$, 0_x 是 $T_x\mathcal{M}$ 的零元;
- 2) 若 $T_0T_x\mathcal{M} \cong T_x\mathcal{M}$, 则 \mathcal{R}_x 满足 $D\mathcal{R}_x(0_x) = \text{id}_{T_x\mathcal{M}}$, $\text{id}_{T_x\mathcal{M}}$ 是 $T_x\mathcal{M}$ 上的恒等映射。

可以验证, 指数映射是拉回映射。除指数映射以外, 还有其他拉回映射, 例如在球面 $\mathcal{M} = S^{n-1} = \{x \in \mathbb{R}^n : x^T x = 1\}$ 上, $x \in \mathcal{M}$, \mathcal{R} 在 $T_x\mathcal{M}$ 上的映射 \mathcal{R}_x 定义为:

$$\mathcal{R}_x(\xi) = \frac{x + \xi}{\|x + \xi\|} \quad \forall \xi \in T_x \mathcal{M},$$

则 \mathcal{R} 是球面上的拉回映射。

为在黎曼流形 \mathcal{M} 上不同点的切空间之间实现加(减)法运算, 定义平移映射[15]。为此, 记 $T\mathcal{M} \oplus T\mathcal{M} = \{(\eta, \xi) : \eta, \xi \in T_x \mathcal{M}, x \in \mathcal{M}\}$ 。

定义 2.2 称黎曼流形 \mathcal{M} 上的光滑映射 $\mathcal{T} : T\mathcal{M} \oplus T\mathcal{M} \rightarrow T\mathcal{M}$ 是 \mathcal{M} 上的**平移映射**, 如果 \mathcal{T} 满足:

- 1) 对于任意 $\xi, \eta \in T_x \mathcal{M}, x \in \mathcal{M}$, 存在一个拉回映射 \mathcal{R} , 使得 $\mathcal{T}_\eta(\xi) \in T_{\mathcal{R}_x(\eta)} \mathcal{M}$;
- 2) 对于任意 $\xi \in T_x \mathcal{M}, x \in \mathcal{M}$, 都有 $\mathcal{T}_{0_x}(\xi) = \xi$;
- 3) 对于任意 $\xi, \zeta \in T_x \mathcal{M}, x \in \mathcal{M}$, 都有 $\mathcal{T}_\eta(a\xi + b\zeta) = a\mathcal{T}_\eta(\xi) + b\mathcal{T}_\eta(\zeta)$ 。

特别地, 我们介绍 *Stiefel* 流形 $\mathcal{M} = \text{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}$ ($p \geq 2$) 及其上的拉回映射和平移映射等, 这些结论均可以在[15]中找到。设 $X \in \mathcal{M}$, 在点 X 处切空间为

$$T_x \mathcal{M} = \{Z \in \mathbb{R}^{n \times p} : X^T Z + Z^T X = 0\} = \{X\Omega + X_\perp K : \Omega^T = -\Omega, K \in \mathbb{R}^{(n-p) \times p}\},$$

黎曼度量为

$$g(Z_1, Z_2) = \langle Z_1, Z_2 \rangle = \text{trace}(Z_1^T Z_2) \quad \forall Z_1, Z_2 \in T_x \mathcal{M},$$

其中 $\text{trace}(A)$ 表示矩阵 A 的迹。设 $X \in \mathcal{M}, Z \in T_x \mathcal{M}$ 。在点 X 处指数映射为:

$$\text{Exp}_x(t) = [X, Z] \exp \left(t \begin{bmatrix} A(0) & -S(0) \\ I_p & A(0) \end{bmatrix} \right) \begin{bmatrix} I_p \\ 0 \end{bmatrix} \exp(-A(0)t),$$

其中 $A(0) := X^T Z, S(0) = Z^T Z$ 。基于极分解的拉回映射定义为:

$$\mathcal{R}_x(Z) = (X + Z)(I_p + Z^T Z)^{-\frac{1}{2}};$$

基于 QR 分解的拉回映射可定义为:

$$\mathcal{R}_x(Z) = \text{qf}(X + Z), \tag{2.1}$$

其中 $\text{qf}(X + Z)$ 是对矩阵 $X + Z$ 进行 QR 分解后所得到的正交矩阵 Q [15]。在流形上可以沿着测地线进行线搜索, 也可以直接使用相应的拉回映射, 但他们在不同的流形上的计算效率是不一样的。在球面上无论是沿测地线进行线搜索还是使用拉回映射, 计算效率相差不大[3] [25]。一般而言, 即使同一个流形上, 不同的拉回映射也可能会有不同的计算效率。设 $X \in \mathcal{M}, Z_1, Z_2 \in T_x \mathcal{M}$ 。在后面的数值实验中, 我们采用以下 *Stiefel* 流形上的平移映射[22]:

$$\mathcal{T}_{Z_2} Z_1 = (I - Y Y^T) Z_1 + Y \text{skew}(Y^T Z_1) \in T_Y \mathcal{M}, \tag{2.2}$$

其中 $Y = \mathcal{R}_x(Z_2), \text{skew}(Y^T Z_1) := \frac{1}{2}(Y^T Z_1 - Z_1^T Y)$ 。注意, 其中拉回映射 \mathcal{R} 可以是上面指出的拉回映射中的任意一个。

设 $f : \mathcal{M} \rightarrow \mathbb{R}$ 是黎曼流形 \mathcal{M} 上的光滑函数, 定义 f 的梯度为一个光滑切向量场 $\text{grad}f$ 如下:

$$g(\text{grad}f(x), \xi) = \langle \text{grad}f(x), \xi \rangle = \xi(f) \quad \forall x \in \mathcal{M}, \xi \in T_x \mathcal{M}.$$

设黎曼流形 \mathcal{M} 是黎曼流形 $\bar{\mathcal{M}}$ 的子流形, $x \in \mathcal{M}, \bar{f} : \bar{\mathcal{M}} \rightarrow \mathbb{R}$ 是 $\bar{\mathcal{M}}$ 的光滑函数。设 $f : \mathcal{M} \rightarrow \mathbb{R}$ 是 \bar{f} 在 \mathcal{M} 上的限制, 则 f 在 \mathcal{M} 上梯度函数等于 \bar{f} 在 $\bar{\mathcal{M}}$ 上的梯度在 $T_x \mathcal{M}$ 上的投影, 即有:

$$\text{grad}f(x) = P_x \text{grad}\bar{f}(x),$$

其中 P_x 是从切空间 $T_x \bar{\mathcal{M}}$ 到切空间 $T_x \mathcal{M}$ 的投影算子。本文中的范数“ $\|\cdot\|$ ”均表示为 *Frobenius* 矩阵范数。

3. 黎曼梯度下降算法及黎曼共轭梯度算法

考虑黎曼流形上的优化问题:

$$\min_{X \in \mathcal{M}} f(X),$$

其中 \mathcal{M} 是一个黎曼流形, $f: \mathcal{M} \rightarrow \mathbb{R}$ 是定义在 \mathcal{M} 上的光滑函数。正如我们在引言部分指出, 黎曼流形上的一阶优化算法的研究取得了丰富的研究结果。特别地, 文献[1] [15]分别研究了黎曼流形上带 *Armijo* 步长的梯度下降算法和共轭梯度算法, 具体见下面的算法 1 和算法 2。

算法 1. 采用 *Armijo* 步长的黎曼梯度下降算法

Step 0 任取初始点 $X_0 \in \mathcal{M}$, 给定参数 $\alpha, \rho, \delta \in (0, 1)$.

Step 1 计算 $\eta_k = -\text{grad}f(X_k)$.

Step 2 确定 $\alpha_k = \max\{\alpha\rho^j, j = 0, 1, 2, \dots\}$ 使得

$$f(\mathcal{R}_{x_k}(\alpha_k \eta_k)) - f(X_k) \leq -\delta \alpha_k \|\eta_k\|^2.$$

Step 3 令 $X_{k+1} := \mathcal{R}_{X_k}(\alpha_k \eta_k)$.

Step 4 $k = k + 1$ 并转 **Step 1**.

算法 2. 采用 *Armijo* 步长的黎曼共轭梯度算法

Step 0 任取初始点 $X_0 \in \mathcal{M}$, 给定参数 $\mu_2 \geq \mu_1 > 0, \mu_3 \geq 0, \alpha, \rho, \delta_1 \in (0, 1), \delta_2 > 0$.

Step 1 计算 $\eta_k := \begin{cases} -\text{grad}f(X_k) & \text{if } k = 0 \\ -\text{grad}f(X_k) + \beta_k \Delta \eta_k & \text{if } k \geq 1 \end{cases}$,

$$\text{其中 } \Delta \eta_k := \mathcal{T}_{\alpha_{k-1}(\eta_{k-1})} \eta_{k-1}, \quad \beta_k := \frac{\mu_1 \|\text{grad}f(X_k)\|^2}{\mu_2 \langle \text{grad}f(X_k), \Delta \eta_k \rangle + \mu_3 \|\text{grad}f(X_{k-1})\|^2}.$$

Step 2 确定 $\alpha_k = \max\{\alpha\rho^j, j = 0, 1, 2, \dots\}$ 使得

$$f(\mathcal{R}_{x_k}(\alpha_k \eta_k)) - f(X_k) \leq \delta_1 \alpha_k \langle \text{grad}f(X_k), \eta_k \rangle - \delta_2 \alpha_k^2 \|\eta_k\|^2.$$

Step 3 令 $X_{k+1} := \mathcal{R}_{X_k}(\alpha_k \eta_k)$.

Step 4 $k = k + 1$ 并转 **Step 1**.

关于这两个算法的收敛性见下面的命题 3.1 和 3.2, 它们分别来自文献[15] [18] [19] [25]和[1] [19]。

命题 3.1 假设 $f: \mathcal{M} \rightarrow \mathbb{R}$ 是一个光滑映射, $\{X_k\}$ 是在 f 作用下由算法 1 产生的无穷点列, 且满足:

- ① 对于任意给定点 $X_0 \in \mathcal{M}$, $\Omega = \{X \in \mathcal{M}: f(X) \leq f(X_0)\}$ 是紧集;
- ② 梯度函数 $\text{grad}f: T_x \mathcal{M} \rightarrow T_x \mathcal{M}$ 在 $0_x \in T_x \mathcal{M}$ 处 Lipschitz 连续, 即存在一个常数 $L > 0, X_1, X_2 \in \mathcal{M}$, 使得:

$$\|\text{grad}f(X_1) - \text{grad}f(X_2)\| \leq L \cdot \text{dist}(X_1, X_2),$$

其中 $\text{dist}(X_1, X_2)$ 是 X_1, X_2 两点的黎曼距离。则有

$$\liminf_{k \rightarrow \infty} \|\text{grad}f(X_k)\| = 0 \tag{3.1}$$

命题 3.2 假设 $f: \mathcal{M} \rightarrow \mathbb{R}$ 是一个光滑映射, $\{X_k\}$ 是在 f 作用下由算法 2 产生的无穷点列, 满足命题 3.1 中的条件①和②, 且算法 2 中的参数还满足条件

$$\textcircled{3} \quad 1 - \frac{\mu_1}{\mu_2} > \frac{2\delta_1}{1 + \delta_1}, \quad \text{且} \quad \frac{\mu_1}{\mu_3} < 1, \tag{3.2}$$

则有(3.1)成立。

4. 对称非负逆特征值问题

本文研究求解 *SNIEP* 的算法, 即求一个对称非负矩阵, 使得它的特征值是给定的 n 个常数 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 。我们将通过 *SNIEP* 转化为将黎曼流形上优化问题, 并提出黎曼梯度算法和黎曼共轭梯度算法进行求解。为此, 先引进一些记号。记 $\Lambda := \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 。记所有的 n 阶正交矩阵构成的 Stiefel 流形为 $\mathcal{O}(n)$ 集合, 即

$$\mathcal{O}(n) := \{Q \in \mathbb{R}^{n \times n} : Q^T Q = I_n\}。$$

设 $X \in \mathcal{O}(n)$, 则流形 $\mathcal{O}(n)$ 在点 X 处的切空间是

$$T_X \mathcal{O}(n) = \{Z \in \mathbb{R}^{n \times n} : X^T Z + Z^T X = 0\}。$$

黎曼度量为

$$g(Z_1, Z_2) = \langle Z_1, Z_2 \rangle = \text{trace}(Z_1^T Z_2) \quad \forall Z_1, Z_2 \in T_X \mathcal{O}(n)。$$

记 n 阶实对称矩阵组成的集合 $\mathcal{S}(n) := \{B \in \mathbb{R}^{n \times n} : B = B^T\}$, 显然 $\mathcal{S}(n)$ 是欧氏空间, 也是特殊的黎曼流形。设 $X \in \mathcal{S}(n)$, 则 $\mathcal{S}(n)$ 在 X 点处的切空间为

$$T_X \mathcal{S}(n) = \mathcal{S}(n)。$$

在点 X 处的黎曼度量为

$$g(W_1, W_2) = \langle W_1, W_2 \rangle = \text{trace}(W_1^T W_2) \quad \forall W_1, W_2 \in T_X \mathcal{S}(n)。$$

显然乘积空间 $\mathcal{O}(n) \times \mathcal{S}(n)$ 也构成一个黎曼流形, 在任意一点 $(Q, R) \in \mathcal{O}(n) \times \mathcal{S}(n)$ 处的切空间为

$$T_{(Q,R)} \mathcal{O}(n) \times \mathcal{S}(n) = T_Q \mathcal{O}(n) \times T_R \mathcal{S}(n) = (T_Q \mathcal{O}(n), T_R \mathcal{S}(n))。$$

设 $(Q, R) \in \mathcal{O}(n) \times \mathcal{S}(n)$ 。在点 (Q, R) 的黎曼度量为: 对任意 $(Z_1, W_1), (Z_2, W_2) \in T_{(Q,R)} \mathcal{O}(n) \times \mathcal{S}(n)$,

$$g_{(Q,R)}((Z_1, W_1), (Z_2, W_2)) = \langle (Z_1, W_1), (Z_2, W_2) \rangle = \text{trace}(Z_1^T Z_2) + \text{trace}(W_1^T W_2)。$$

对于黎曼流形 $\mathcal{O}(n)$, 本文采用了基于 QR 分解的拉回映射(2.1)和平移映射(2.2)。设 $(Q, R) \in \mathcal{O}(n) \times \mathcal{S}(n)$, $(Z_1, W_1), (Z_2, W_2) \in T_{(Q,R)} \mathcal{O}(n) \times \mathcal{S}(n)$, 则可以分别定义流形 $\mathcal{O}(n) \times \mathcal{S}(n)$ 上的拉回映射和平移映射如下:

$$\mathcal{R}_{(Q,R)}(Z_1, W_1) = (\text{qf}(Q + Z_1), R + W_1),$$

$$\mathcal{T}_{(Z_1, W_1)}(Z_2, W_2) = \left(\frac{1}{2}(Z_1 - YZ_1^T Y), W_2 \right),$$

其中 $Y := \mathcal{R}_Q(Z_2)$ 。

文献[1]将 *SNIEP* 转化成了下面的黎曼优化问题:

$$\begin{aligned} \min f(Q, R) &:= \frac{1}{2} \|Q^T \Lambda Q - R * R\|^2 \\ \text{s.t. } (Q, R) &\in \mathcal{O}(n) \times \mathcal{S}(n) \end{aligned} \tag{4.1}$$

其中 $\|\cdot\|$ 表示矩阵的 F 范数, $R * R$ 表示矩阵 R 和 R 的 Hadamard 乘积(即 $R * R = (R_{ij}^2)$, $R = (R_{ij})$)。当 *SNIEP* 有解时, *SNIEP* 等价于黎曼流形 $\mathcal{O}(n) \times \mathcal{S}(n)$ 上优化问题(4.1)。同时文献[7]给出了目标函数 f 在流形 $\mathcal{O}(n) \times \mathcal{S}(n)$ 上的梯度表达式如下: 对于任意的 $(Q, R) \in \mathcal{O}(n) \times \mathcal{S}(n)$, 都有

$$\text{grad}f(Q, R) = (Q(R * R)Q^T \Lambda Q - \Lambda Q(R * R), -2(Q^T \Lambda Q - R * R) * R). \tag{4.2}$$

Chu 等人在[7]提出了用“梯度流”的方法求解问题(4.1)。具体地, 将 $X(t) = Q(t)^T \Lambda Q(t)$ 、 $Y(t) = R(t) * R(t)$ 看作关于 t 的“流”, 通过建立 $Q(t)$ 和 $R(t)$ 满足的微分方程, 运用微分方程的求解方法求解其极限点。如果轨道 $(X(t), Y(t))$ 的 ω -极限集包含一个类型为 (\hat{X}, \hat{X}) 的点, 则 $(X(t), Y(t))$ 收敛到 (\hat{X}, \hat{X}) 。容易看出, “梯度流”的求解思路本质上是极速下降法的思路, 但不是所有梯度“流”都会收敛到问题(4.1)的解。基于黎曼流形上的优化模型(4.1), 我们提出采用线搜索的黎曼优化算法 1 和算法 2 进行求解。以下是这两个算法运用到求解问题(4.1)的收敛性分析, 为此我们先证明下面的引理。

引理 4.1 在模型(4.1)中, 对于任意给定点 $(Q_0, R_0) \in \mathcal{O}(n) \times \mathcal{S}(n)$, 以下水平集是紧集:

$$\Omega = \{(Q, R) \in \mathcal{O}(n) \times \mathcal{S}(n) : f(Q, R) \leq f(Q_0, R_0)\}.$$

证明: 设 $(Q, R) \in \Omega$, 则

$$\|Q\| = \sqrt{\text{trace}(Q^T Q)} = \sqrt{n}, \tag{4.3}$$

$$\|Q^T \Lambda Q - R * R\| \leq \sqrt{2f(Q_0, R_0)}. \tag{4.4}$$

由柯西不等式,

$$\|R\| = \sqrt{\sum_{i,j} r_{ij}^2} = \sqrt{\sum_{i,j} (1 \cdot r_{ij}^2)} \leq \sqrt{n^2 \cdot \sum_{i,j} r_{ij}^4} = n \cdot \|R * R\|. \tag{4.5}$$

又由矩阵 F 范数的酉不变性[26]可得

$$\|Q^T \Lambda Q\| = \|\Lambda\|.$$

所以由(4.4), 得

$$\|R * R\| \leq \|Q^T \Lambda Q - R * R\| + \|Q^T \Lambda Q\| \leq \sqrt{2f(Q_0, R_0)} + \|\Lambda\|,$$

所以由(4.5)

$$\|R\| \leq n \cdot (\sqrt{2f(Q_0, R_0)} + \|\Lambda\|),$$

于是由(4.3)及上式可得

$$\|(Q, R)\| = \|Q\| + \|R\| \leq n \cdot (\sqrt{2f(Q_0, R_0)} + \|\Lambda\|) + \sqrt{n},$$

故水平集 Ω 是有界的。又 f 在 $\mathcal{O}(n) \times \mathcal{S}(n)$ 上显然是连续的, 故水平集 Ω 是闭集, 从而 Ω 是紧集。

引理 4.2 设 $(Q_0, R_0) \in \mathcal{O}(n) \times \mathcal{S}(n)$, 则 $\text{grad}f$ 在水平集 Ω 上是 Lipschitz 连续的(Ω 如引理 4.1 中定义)。

证明: 由目标函数 f 的定义, 可知 f 是定义在 $\mathcal{O}(n) \times \mathcal{S}(n)$ 上的光滑函数。又由引理 4.1 知 Ω 是紧集,

f 的 Hessian 矩阵在 Ω 上有界。再由微分中值定理, $\text{grad}f$ 在 Ω 上是 Lipschitz 连续。

由命题 2.1, 2.2 和引理 4.1, 4.2, 可得以下定理。

定理 4.1 设 $\{X_k\}$ 是运用算法 1 或者算法 2 求解问题(4.1)产生的无穷点列, 且算法 2 中参数满足命题 2.2 中条件③, 则有

$$\liminf_{k \rightarrow \infty} \|\text{grad}f(X_k)\| = 0.$$

5. 数值实验

本节利用数值实验来验证算法 1 和算法 2 有效性, 并比较这两种算法之间的收敛性。停机准则: $\|\text{grad}f(Q, R)\|^2 \leq 10^{-12}$, 实验所用软件是 Matlab2018a。实验中初始矩阵 $Q_0 = I_n$ 是 n 阶单位矩阵, $R_0 = \text{ones}(n)$ 是所有元素都是 1 的 n 阶矩阵。实验中的参数取值情况如下: $\alpha = \rho = 0.5$, $\delta = \delta_1 = 0.1$, $\delta_2 = 10$ 。为了满足命题 3.2 的条件③ (即(3.2)式), 取 $\mu_1 = 1, \mu_2 = 2, \mu_3 = 3$ 。下图 1~4 所示的是不同阶数, 两种算法的收敛比较(横坐标是迭代次数, 纵坐标是函数值)。

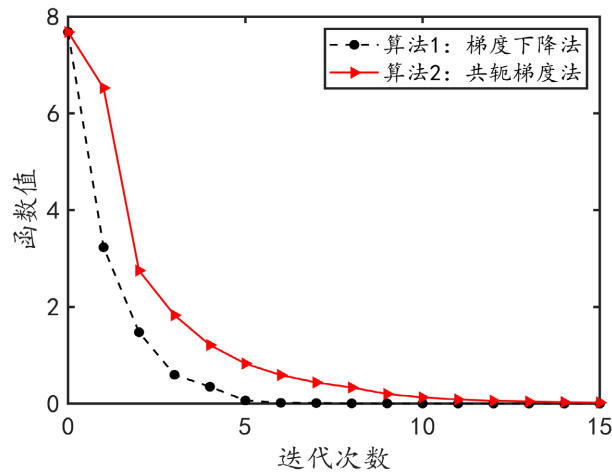


Figure 1. $n = 4$

图 1. $n = 4$

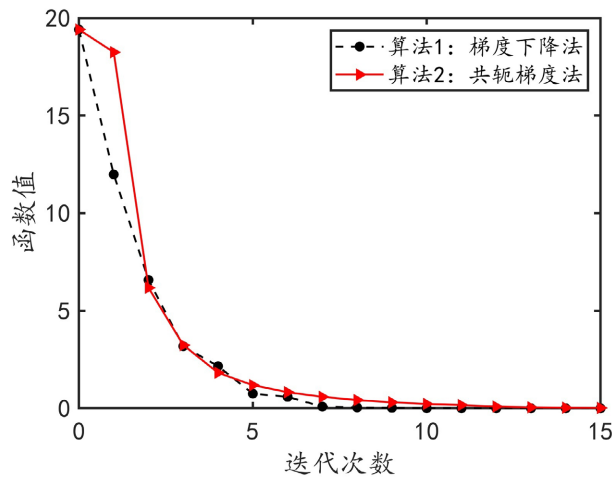


Figure 2. $n = 6$

图 2. $n = 6$

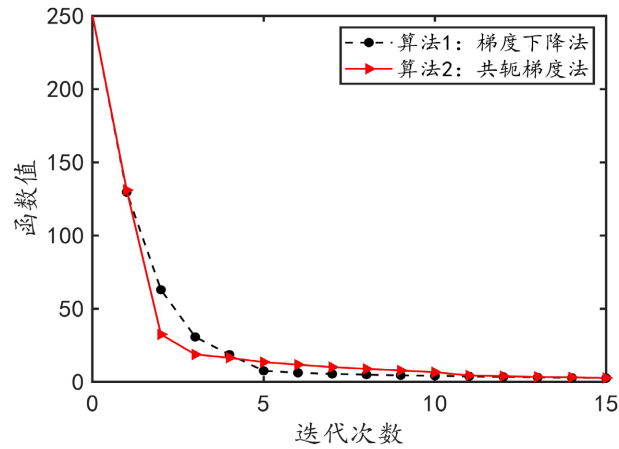


Figure 3. $n = 20$
图 3. $n = 20$

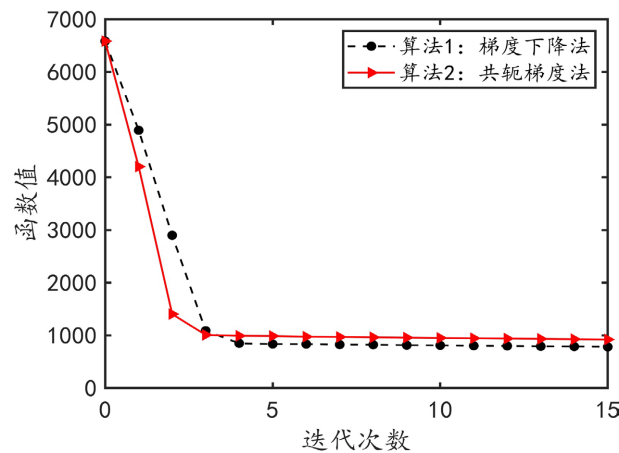


Figure 4. $n = 100$
图 4. $n = 100$

为了进一步比较黎曼梯度下降算法和黎曼共轭梯度算法的收敛效率, 在选取和以上实验相同参数的情况下, 将函数值的精确度 Res 取为 $1.0E-13$; 矩阵的阶数 n 依次取为 6, 10, 20, 50, 100, 迭代时间都是运行程序 10 次后取的平均值。其中 i 是迭代的次数, T 是迭代时间(单位是秒), Res 表示函数值 f 的精确度, n 是矩阵的阶数。见表 1。

Table 1. The experimental results of Algorithm 1 and Algorithm 2 at different orders

表 1. 算法 1 和算法 2 在不同阶数下的实验结果

阶数		算法 1 (黎曼梯度下降算法)			算法 2 (黎曼共轭梯度算法)		
n	Res	i	T	Res	i	T	
6	$8.71E-13$	72	0.0029	$3.36E-13$	103	0.007	
10	$3.28E-13$	70	0.004	$3.59E-13$	125	0.01	
20	$1.50E-13$	322	0.039	$2.08E-13$	272	0.04	
50	$5.12E-13$	6038	4.5	$2.43E-13$	1201	1.12	
100	$5.40E-13$	22226	109	$2.68E-13$	4616	23	

从以上各图可以看出, 在相同的迭代次数时, 黎曼梯度下降算法比的函数值下降量更多。但是从上图可以看出, 在阶数较高的情形下($n \geq 50$), 黎曼共轭梯度算法达到相同精度的运行时间比黎曼梯度下降算法更短。这表明, 在求解高阶 *SNIEP* 时, 黎曼共轭梯度算法比黎曼梯度下降算法更有效。

6. 结论

本文提出了用黎曼梯度下降算法(算法 1)和黎曼共轭梯度算法(算法 2)求解对称非负逆特征值问题, 并建立了算法的收敛性。在数值实验中比较了这两种算法的收敛效率, 实验的结果是: 在高阶的情况下, 黎曼共轭梯度算法比黎曼梯度下降算法收敛效率更高。接下来我们将研究用黎曼二阶算法求解对称非负逆特征值问题, 并比较各种黎曼算法的收敛效率。

基金项目

本文受国家自然科学基金(12161017)和贵州省省级科技计划项目(ZK[2022]110)资助项目资助。

参考文献

- [1] Yao, T.T., Bai, Z.J. and Zhao, Z. (2018) A Riemannian Variant of the Fletcher-Reeves Conjugate Gradient Method for Stochastic Inverse Eigenvalue Problems with Partial Eigendata. *Numerical Linear Algebra with Applications*, **26**, 1-19. <https://doi.org/10.1002/nla.2221>
- [2] Zhao, Z. and Bai, Z.J. (2018) A Riemannian Inexact Newton-CG Method for Constructing a Nonnegative Matrix with Prescribed Realizable Spectrum. *Numerische Mathematik*, **140**, 827-855. <https://doi.org/10.1007/s00211-018-0982-2>
- [3] Bai, Z.J., Stefano, S.C. and Zhao, Z. (2012) Nonnegative Inverse Eigenvalue Problems with Partial Eigendata. *Numerische Mathematik*, **120**, 387-431. <https://doi.org/10.1007/s00211-011-0415-y>
- [4] Barcilon, V. (1990) A Two-Dimensional Inverse Eigenvalue Problem. *Journal of Inverse and Ill-Posed Problems*, **6**, 11-20. <https://doi.org/10.1088/0266-5611/6/1/004>
- [5] Dawson, C.B. and Cha, P.D. (2019) A Sensitivity-Based Approach to Solving the Inverse Eigenvalue Problem for Linear Structures Carrying Lumped Attachments. *Materials Science and Engineering R—Reports*, **120**, 537-566. <https://doi.org/10.1002/nme.6147>
- [6] Joseph, K.T. (2012) Inverse Eigenvalue Problem in Structural Design. *AIAA Journal*, **30**, 2890-2896. <https://doi.org/10.2514/3.11634>
- [7] Chu, M.T. and Driessel, K.R. (1991) Constructing Symmetric Nonnegative Matrices with Prescribed Eigenvalues by Differential Equations. *SIAM Journal on Numerical Analysis*, **22**, 1372-1387. <https://doi.org/10.1137/0522088>
- [8] Orsi, R. (2006) Numerical Methods for Solving Inverse Eigenvalue Problems for Nonnegative Matrices. *SIAM Journal on Matrix Analysis and Applications*, **28**, 190-212. <https://doi.org/10.1137/050634529>
- [9] Bauschke, H.H. and Borwein, J.M. (1996) On Projection Algorithms for Solving Convex Feasibility Problems. *SIAM Review*, **38**, 367-426. <https://doi.org/10.1137/S0036144593251710>
- [10] Loewy, R. and London, D.A. (1978) Note on an Inverse Problem for Nonnegative Matrices. *Linear and Multilinear Algebra*, **6**, 83-90. <https://doi.org/10.1080/03081087808817226>
- [11] Chu, M.T. and Golub, G.H. (2005) *Inverse Eigenvalue Problems: Theory, Algorithms and Applications*. Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780198566649.001.0001>
- [12] 杨丹, 王湘美. 交替投影算法求解非负逆特征值问题[J]. 运筹与模糊学, 2021, 11(1): 9-14. <https://doi.org/10.12677/orf.2021.111002>
- [13] 党娟娟, 王湘美. 交替投影法求解对称随机逆特征值问题[J]. 应用数学进展, 2021, 10(6): 2206-2216. <https://doi.org/10.12677/aam.2021.106230>
- [14] Luenberger, D.G. (1972) The Gradient Projection Method along Geodesics. *Management Science*, **18**, 620-631. <https://doi.org/10.1287/mnsc.18.11.620>
- [15] Absil, P.A., Mahony, R. and Sepulchre, R. (2008) *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton. <https://doi.org/10.1515/9781400830244>
- [16] Gabay, D. (1982) Minimizing a Differentiable Function over a Differential Manifold. *Journal of Optimization Theory & Applications*, **37**, 177-219. <https://doi.org/10.1007/BF00934767>

-
- [17] Smith, S.T. (1993) *Geometric Optimization Methods for Adaptive Filtering*. Harvard University, Cambridge.
- [18] Yang, Y. (2007) Globally Convergent Optimization Algorithms on Riemannian Manifolds: Uniform Framework for Unconstrained and Constrained Optimization. *Journal of Optimization Theory & Applications*, **132**, 245-265. <https://doi.org/10.1007/s10957-006-9081-0>
- [19] Wang, J.H., Wang, X.M., Li, C. and Yao, J.C. (2021) Convergence Analysis of Gradient Algorithms on Riemannian Manifolds without Curvature Constraints and Application to Riemannian Mass. *SIAM Journal on Optimization: A Publication of the Society for Industrial and Applied Mathematics*, **31**, 172-199. <https://doi.org/10.1137/19M1289285>
- [20] Wang, J.H., Huang, S.C. and Li, C. (2009) Extended Newton's Algorithm for Mappings on Riemannian Manifolds with Values in a Cone. *Taiwanese Journal of Mathematics*, **13**, 633-656. <https://doi.org/10.11650/twjm/1500405392>
- [21] Wang, X.M. (2018) Subgradient Algorithms on Riemannian Manifolds of Lower Bounded Curvatures. *Optimization: A Journal of Mathematical*, **67**, 179-194. <https://doi.org/10.1080/02331934.2017.1387548>
- [22] 潘汉. 黎曼流形优化及其应用[M]. 北京: 科学出版社, 2020.
- [23] 陈维桓. 微分流形初步[M]. 北京: 高等教育出版社, 2019.
- [24] Adler, R., Dedieu, J.P., Martens, M. and Shub, M. (2002) Newton's Method on Riemannian Manifolds and a Geometric Model for the Human Spine. *IMA Journal of Numerical Analysis*, **22**, 359-390. <https://doi.org/10.1093/imanum/22.3.359>
- [25] Chu, M.T. and Guo, Q.Y. (1998) A Numerical Method for the Inverse Stochastic Spectrum Problem. *SIAM Journal on Matrix Analysis and Applications*, **19**, 1027-1039. <https://doi.org/10.1137/S0895479896292418>
- [26] 朱元国, 饶玲. 矩阵分析与计算[M]. 北京: 国防工业出版社, 2010.