Received: 19 June 2017 | Revised: 28 September 2018 | Accepted: 2 October 2018
DOI: 10.1002/nla.2221

RESEARCH ARTICLE

WILEY

A Riemannian variant of the Fletcher–Reeves conjugate gradient method for stochastic inverse eigenvalue problems with partial eigendata

Teng-Teng Yao[1] | Zheng-Jian Bai[2] | Zhi Zhao[3]

[1]Department of Mathematics, School of Sciences, Zhejiang University of Science and Technology, Hangzhou, China

[2]School of Mathematical Sciences and Fujian Provincial Key Laboratory on Mathematical Modeling & High Performance Scientific Computing, Xiamen University, Xiamen, China

[3]Department of Mathematics, School of Sciences, Hangzhou Dianzi University, Hangzhou, China

Correspondence
Zheng-Jian Bai, School of Mathematical Sciences and Fujian Provincial Key Laboratory on Mathematical Modeling & High Performance Scientific Computing, Xiamen 361005, China.
Email: zjbai@xmu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 11701514, 11671337, and 11601112; Natural Science Foundation of Fujian Province of China, Grant/Award Number: 2016J01035; Fundamental Research Funds for the Central Universities, Grant/Award Number: 20720180008

Summary

In this paper, we focus on the stochastic inverse eigenvalue problem with partial eigendata of constructing a stochastic matrix from the prescribed partial eigendata. A Riemannian variant of the Fletcher–Reeves conjugate gradient method is proposed for solving a general unconstrained minimization problem on a Riemannian manifold, and the corresponding global convergence is established under some assumptions. Then, we reformulate the inverse problem as a nonlinear least squares problem over a matrix oblique manifold, and the application of the proposed geometric method to the nonlinear least squares problem is investigated. The proposed geometric method is also applied to the case of prescribed entries and the case of column stochastic matrix. Finally, some numerical tests are reported to illustrate that the proposed geometric method is effective for solving the inverse problem.

KEYWORDS

Fletcher–Reeves conjugate gradient method, inverse eigenvalue problem, Riemannian manifold, stochastic matrix

1 | INTRODUCTION

An inverse eigenvalue problem (IEP) aims to reconstruct a structured matrix from prescribed spectral data, which arises in many applications such as structural dynamics, control design, vibration, circuit theory, geophysics, system identification, molecular spectroscopy, seismic tomography, remote sensing, and principal component analysis. On many applications, mathematical properties, and various numerical methods of general IEPs, one may refer to other works[1–7] and the references therein.

Abbreviations: IEP, inverse eigenvalue problem; StIEP, stochastic inverse eigenvalue problem; StIEP-ED, stochastic inverse eigenvalue problem with partial eigendata; StIEP-ED-PE, stochastic inverse eigenvalue problem with partial eigendata and prescribed entries.

An n-by-n matrix A is called nonnegative if $A_{ij} \geq 0$ for all $i,j = 1, \ldots, n$, where A_{ij} denote the (i,j)th entry of a square matrix A. An n-by-n matrix A is called (row) stochastic if it is nonnegative and $\sum_{j=1}^{n} A_{ij} = 1$ for all $i = 1, \ldots, n$. In this paper, we consider the following stochastic IEP with partial eigendata (StIEP-ED) and the stochastic IEP with partial eigendata and prescribed entries (StIEP-ED-PE).

StIEP-ED. *Given a self-conjugate set of complex pairs* $\{(\lambda_k, \mathbf{x}_k) \in \mathbb{C} \times \mathbb{C}^n\}_{k=1}^{p}$ $(p \ll n)$, *find an n-by-n stochastic matrix C such that*

$$C\mathbf{x}_k = \lambda_k \mathbf{x}_k, \quad k = 1, \ldots, p.$$

StIEP-ED-PE. *Given a self-conjugate set of complex pairs* $\{(\lambda_k, \mathbf{x}_k) \in \mathbb{C} \times \mathbb{C}^n\}_{k=1}^{p}$ $(p \ll n)$, *an n-by-n stochastic matrix* C_a, *and an index subset* $\mathcal{L} \subset \mathcal{N} := \{(i,j) \mid i,j = 1, \ldots, n\}$, *find an n-by-n stochastic matrix C such that*

$$C\mathbf{x}_k = \lambda_k \mathbf{x}_k, \quad k = 1, \ldots, p \quad \text{and} \quad C_{ij} = (C_a)_{ij}, \quad \forall (i,j) \in \mathcal{L}.$$

Nonnegative matrices and stochastic matrices arise in many applications including probability and statistics, queues, web information retrieval, dynamical geometry, and economics.[8–13] The stochastic IEP (StIEP) with complete eigenvalues was considered in many literature. On its solvability, see for instance other works.[1–3,7,11,14–17] There exist some numerical methods for solving the StIEP, including constructive methods,[18–20] an isospectral flow method,[21] an alternating projection-like algorithm,[22] and a recursive algorithm.[23]

Recently, there exist some Riemannian optimization methods for eigenproblems and IEPs.[24–29] Sparked by a variant of the Fletcher–Reeves (FR) conjugate gradient method for solving an unconstrained nonlinear optimization problem over $\mathbb{R}^n$,[30] we propose a Riemannian variant of the FR conjugate gradient method for solving the unconstrained minimization problem on a Riemannian manifold. The global convergence of the proposed method is established under some assumptions. The StIEP-ED is rewritten as a nonlinear least squares problem over an oblique manifold and the application of the proposed method to the StIEP-ED is discussed. We also apply the proposed method to the StIEP-ED-PE and the case of the column stochastic matrix. Finally, some numerical experiments are reported to show that the proposed geometric method is effective for solving the StIEP-ED, the StIEP-ED-PE, and the case of the column stochastic matrix.

Throughout the paper, we use the following notations. Let I_n be the identity matrix of order n. Let $\mathbb{R}^{m\times n}$ be the set of all m-by-n real matrices. The symbols A^T and $A^\dagger$ denote the transpose and the Moore–Penrose pseudoinverse of a matrix A. Let $A \odot B$ mean the Hadamard product of two matrices $A, B \in \mathbb{R}^{n\times n}$. For a matrix $A \in \mathbb{R}^{n\times n}$, diag($A$), tr($A$), and vec($A$) mean the diagonal matrix with the same diagonal entries as A, the sum of the diagonal entries of A, and an n^2-vector obtained by stacking the columns of A on top of one another accordingly. For a vector $\mathbf{x} \in \mathbb{R}^n$, Diag($\mathbf{x}$) is a diagonal matrix with $\mathbf{x}$ on its diagonal. The symbol "$\simeq$" means the identification of two sets.

The remaining part of this paper is organized as follows. In Section 2, we review a variant of the FR conjugate gradient method for solving the unconstrained nonlinear optimization problem over $\mathbb{R}^n$. In Section 3, we present a Riemannian variant of the FR conjugate gradient method for solving the unconstrained minimization problem on a Riemannian manifold. In Section 4, the global convergence of the proposed geometric method is established. In Section 5, the proposed geometric method is applied to the StIEP-ED, the StIEP-ED-PE, and the case of column stochastic matrix. Finally, some numerical tests are reported in Section 6 and some concluding remarks are given in Section 7.

2 | PRELIMINARIES

In this section, we recall a variant of the FR conjugate gradient method for solving general unconstrained nonlinear optimization problem.[30(p35)]

We consider the following unconstraint optimization problem:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth nonlinear function. The conjugate gradient method is given by

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

where α_k is the step length, which is computed by some line search and $\mathbf{d}_k$ is the search direction defined by

$$\mathbf{d}_k = \begin{cases} -\mathbf{g}_k, & \text{if } k = 0, \\ -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}, & \text{if } k \geq 1. \end{cases} \tag{2}$$

Here, β_k is a scalar and $\mathbf{g}_k = \mathbf{g}(\mathbf{z}_k)$, where $\mathbf{g}(\mathbf{z})$ means the gradient of f at $\mathbf{z} \in \mathbb{R}^n$.

There exist some classical formulas for computing β_k, for example, the FR and the Polak–Ribière–Polyak (PRP)[31–33]:

$$\beta_k^{\mathrm{FR}} = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}} \tag{3}$$

and

$$\beta_k^{\mathrm{PRP}} = \frac{\mathbf{g}_k^T(\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}.$$

For more conjugate gradient methods, one may refer to the works of Dai et al.,[34] Hager et al.,[35] and Sun et al.[36]

If the parameter β_k takes the variant formula of FR,

$$\beta_k^{\mathrm{VFR}}(\mu_1, \mu_2, \mu_3) = \frac{\mu_1 \mathbf{g}_k^T \mathbf{g}_k}{\mu_2 |\mathbf{g}_k^T \mathbf{d}_{k-1}| + \mu_3 \mathbf{g}_{k-1}^T \mathbf{g}_{k-1}},$$

where $\mu_1 > 0$, $\mu_2 \geq \mu_1 + \epsilon_1$, $\mu_3 > 0$, ϵ_1 is an arbitrary positive number, and $|a|$ means the absolute value of a real number a, then it is easy to check that, for all $k \geq 0$, we have

$$\mathbf{g}_k^T \mathbf{d}_k \leq -\left(1 - \frac{\mu_1}{\mu_1 + \epsilon_1}\right) \mathbf{g}_k^T \mathbf{g}_k.$$

This shows that $\mathbf{d}_k$ is a descent direction for the objective function f and it is independent of the line search used. We summarize the variant of the FR conjugate gradient algorithm using the standard Armijo line search for solving the optimization problem (1).

Algorithm 1 A variant of FR conjugate gradient method

Step 0. Choose an initial point $\mathbf{z}_0 \in \mathbb{R}^n$, $\rho \in (0, 1)$, $\delta \in (0, \frac{1}{2})$. Let $k := 0$.
Step 1. Compute $\mathbf{d}_k$ by (2) with $\beta_k = \beta_k^{\mathrm{VFR}}(\mu_{1k}, \mu_{2k}, \mu_{3k})$.
Step 2. Determine $\alpha_k = \max\{\rho^j, j = 0, 1, 2, \dots\}$ satisfying

$$f(\mathbf{z}_k + \alpha_k \mathbf{d}_k) - f(\mathbf{z}_k) \leq \delta \alpha_k \mathbf{g}_k^T \mathbf{d}_k. \tag{4}$$

Step 3. Set $\mathbf{z}_{k+1} := \mathbf{z}_k + \alpha_k \mathbf{d}_k$.
Step 4. Replace k by $k + 1$ and go to **Step 1**.

This algorithm was proved to be globally convergent in the work of Wei et al.[30(p36)] if the following conditions hold:

$$\sup\{\mu_{1k} \mid k \geq 1\} < +\infty \quad \text{and} \quad \sup\left\{ \frac{\mu_{1k}^2}{\mu_{3k}^2} \,\middle|\, k \geq 1 \right\} < 1.$$

3 | A RIEMANNIAN VARIANT OF THE FR CONJUGATE GRADIENT METHOD

Let $(\mathcal{M}, g)$ be a Riemannian manifold and $h : \mathcal{M} \to \mathbb{R}$ be a differentiable function on $\mathcal{M}$, where g is a Riemannian metric on $\mathcal{M}$. We consider a Riemannian variant of Algorithm 1 for solving the following unconstrained minimization problem:

$$\begin{array}{ll} \min & h(Z) \\ \text{subject to (s.t.)} & Z \in \mathcal{M}. \end{array} \tag{5}$$

To simplify notations, we use $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ to denote the Riemannian metric and its induced norm on $\mathcal{M}$. Let $T_Z\mathcal{M}$ be the tangent space of $\mathcal{M}$ at a point $Z \in \mathcal{M}$. Let $R : T\mathcal{M} \to \mathcal{M}$ be a retraction on $\mathcal{M}$, where $T\mathcal{M} := \cup_{Z \in \mathcal{M}} T_Z\mathcal{M}$ (definition 4.1.1 in the work of Absil et al.[37]). Given a point $z \in \mathcal{M}$, $R_z : T_Z\mathcal{M} \to \mathcal{M}$ denotes the restriction of R to $T_Z\mathcal{M}$. Let $\mathcal{T}$ be a vector transport on $\mathcal{M}$, which is associated with the retraction R (definition 8.1.1 in the work of Absil et al.[37]). Retraction and vector transport are generalizations of exponential mapping and parallel translation and they are computationally efficient for constructing optimization algorithms on Riemannian manifolds.[37] Based on these concepts, a Riemannian variant of Algorithm 1 for solving problem (5) can be described as follows.

Algorithm 2 A Riemannian Variant of FR conjugate gradient method

Step 0. Choose an initial point $Z_0 \in \mathcal{M}$, $\mu_1 > 0$, $\mu_2 \geq \mu_1$, $\mu_3 > 0$, $\alpha \geq 1$, $\rho, \delta_1 \in (0,1)$, $\delta_2 > 0$. Let $k := 0$.

Step 1. Compute

$$\Delta Z_k := \begin{cases} -\text{grad } h(Z_k), & \text{if } k = 0, \\ -\text{grad } h(Z_k) + \beta_k^{\text{RVFR}} \Delta Y_k, & \text{if } k \geq 1, \end{cases} \tag{6}$$

where

$$\Delta Y_k := \mathcal{T}_{\alpha_{k-1}(\Delta Z_{k-1})} \Delta Z_{k-1}$$

and

$$\beta_k^{\text{RVFR}} := \frac{\mu_1 \|\text{grad } h(Z_k)\|^2}{\mu_2 |\langle \text{grad } h(Z_k), \Delta Y_k \rangle| + \mu_3 \|\text{grad } h(Z_{k-1})\|^2}. \tag{7}$$

Step 2. Determine $\alpha_k = \max\{\alpha \rho^j, j = 0, 1, 2, \ldots\}$ such that

$$h\big(R_{Z_k}(\alpha_k(\Delta Z_k))\big) - h(Z_k) \leq \delta_1 \alpha_k \langle \text{grad } h(Z_k), \Delta Z_k \rangle - \delta_2 \alpha_k^2 \|\Delta Z_k\|^2. \tag{8}$$

Step 3. Set $Z_{k+1} := R_{Z_k}(\alpha_k \Delta Z_k)$.

Step 4. Replace k by $k+1$ and go to **Step 1**.

Remark 1. For purpose of convergence analysis, in Algorithm 2, we replace the standard Armijo line search (4) in Algorithm 1 by a Riemannian version of the form (8) as in the work of Yao et al.[27] In addition, in each iteration, we use the fixed parameters μ_1, μ_2, and μ_3.

Remark 2. It follows from (6) and (7) that, for all $k \geq 1$,

$$\begin{aligned}
\langle \Delta Z_k, \text{grad } h(Z_k) \rangle &= \big\langle -\text{grad } h(Z_k) + \beta_k^{\text{RVFR}} \Delta Y_k, \text{grad } h(Z_k) \big\rangle \\
&= - \|\text{grad } h(Z_k)\|^2 \\
&\quad + \frac{\mu_1 \|\text{grad } h(Z_k)\|^2}{\mu_2 |\langle \text{grad } h(Z_k), \Delta Y_k \rangle| + \mu_3 \|\text{grad } h(Z_{k-1})\|^2} \langle \Delta Y_k, \text{grad } h(Z_k) \rangle \\
&\leq - \|\text{grad } h(Z_k)\|^2 + \frac{\mu_1 \|\text{grad } h(Z_k)\|^2}{\mu_2 |\langle \text{grad } h(Z_k), \Delta Y_k \rangle|} |\langle \Delta Y_k, \text{grad } h(Z_k) \rangle| \\
&= - \left(1 - \frac{\mu_1}{\mu_2}\right) \|\text{grad } h(Z_k)\|^2 \\
&= - \varsigma \|\text{grad } h(Z_k)\|^2 < 0,
\end{aligned} \tag{9}$$

where $\varsigma := 1 - \mu_1/\mu_2 \in (0,1)$. Notice that $\langle \Delta Z_0, \text{grad}\, h(Z_0) \rangle = -\|\text{grad}\, h(Z_0)\|^2$. This shows that the search direction ΔZ_k is a descent direction of h for all $k \geq 0$.

On the other hand, we have, for all $k \geq 1$,

$$\begin{aligned}
\langle \Delta Z_k, \text{grad } h(Z_k) \rangle &\geq - \|\text{grad } h(Z_k)\|^2 \\
&\quad - \frac{\mu_1 \|\text{grad } h(Z_k)\|^2}{\mu_2 |\langle \text{grad } h(Z_k), \Delta Y_k \rangle| + \mu_3 \|\text{grad } h(Z_{k-1})\|^2} |\langle \Delta Y_k, \text{grad } h(Z_k) \rangle| \\
&\geq - \|\text{grad } h(Z_k)\|^2 - \frac{\mu_1 \|\text{grad } h(Z_k)\|^2}{\mu_2 |\langle \text{grad } h(Z_k), \Delta Y_k \rangle|} |\langle \Delta Y_k, \text{grad } h(Z_k) \rangle| \\
&= - \left(1 + \frac{\mu_1}{\mu_2}\right) \|\text{grad } h(Z_k)\|^2 \\
&= (\varsigma - 2) \|\text{grad } h(Z_k)\|^2.
\end{aligned} \tag{10}$$

4 | GLOBAL CONVERGENCE

In this section, we establish the global convergence of Algorithm 2. We note that the pullback function $\widehat{h} : T\mathcal{M} \to \mathbb{R}$ of $h : \mathcal{M} \to \mathbb{R}$ through the retraction R on $\mathcal{M}$ is given by[37(p55)]

$$\widehat{h}(\eta) = h(R(\eta)) \quad \text{for all } \eta \in T\mathcal{M}.$$

The restriction of $\widehat{h}$ to $T_Z\mathcal{M}$ is denoted by $\widehat{h}_Z : T_Z\mathcal{M} \to \mathbb{R}$, that is,

$$\widehat{h}_Z(\eta_Z) = h(R_Z(\eta_Z)) \quad \text{for all } \eta_Z \in T_Z\mathcal{M}.$$

Moreover, we have[37(p56)]

$$\text{grad } h(Z) = \text{grad } \widehat{h}_Z(0_Z),$$

where 0_Z is the origin of $T_Z\mathcal{M}$. To establish the global convergence of Algorithm 2, we need the following assumption.

Assumption 1.

1. The function $h : \mathcal{M} \to \mathbb{R}$ is differentiable and bounded below in the level set $\Omega := \{Z \in \mathcal{M} \mid h(Z) \leq h(Z_0)\}$.
2. The gradient grad $\widehat{h}_Z : T_Z\mathcal{M} \to T_Z\mathcal{M}$ is Lipschitz continuous at $0_Z \in T_Z\mathcal{M}$ uniformly in Z in the level set Ω, that is, there exist two constants $\kappa > 0$ and $\beta_L > 0$ such that
$$\left\| \text{grad } \widehat{h}_Z(\xi_Z) - \text{grad } \widehat{h}_Z(0_Z) \right\| \leq \beta_L \|\xi_Z\|, \tag{11}$$
for any $Z \in \Omega$ and $\xi_Z \in T_Z\mathcal{M}$ with $\|\xi_Z\| \leq \kappa$.
3. The vector tansport $\mathcal{T}$ satisfies the following inequality:
$$\|\mathcal{T}_{\eta_Z}\xi_Z\| \leq \|\xi_Z\|, \quad \forall \xi_Z, \eta_Z \in T_Z\mathcal{M},\ Z \in \mathcal{M}. \tag{12}$$

For the step size α_k determined by the line search (8) in **Step 2** of Algorithm 2, we have the following estimation.

Lemma 1. *Suppose Assumption 1 is satisfied. If $\varsigma := 1 - \mu_1/\mu_2 > 2\delta_1/(1 + \delta_1)$, then there exists a constant $\nu > 0$ such that, for all k sufficiently large,*

$$\alpha_k \geq \nu \frac{\|\text{grad } h(Z_k)\|^2}{\|\Delta Z_k\|^2}. \tag{13}$$

Proof. Using (8)–(9) and Assumption 1, it follows that the sequence $\{h(Z_k)\}$ is decreasing and bounded below and thus is convergent. Thus,

$$\sum_{k=0}^{\infty} \left(-\delta_1\alpha_k\langle \text{grad } h(Z_k), \Delta Z_k\rangle + \delta_2\alpha_k^2\|\Delta Z_k\|^2\right) < \infty.$$

From (9), we have

$$\sum_{k=0}^{\infty} \alpha_k^2\|\Delta Z_k\|^2 < \infty \tag{14}$$

and

$$\varsigma \sum_{k=0}^{\infty} \alpha_k\|\text{grad } h(Z_k)\|^2 \leq -\sum_{k=0}^{\infty} \alpha_k\langle \text{grad } h(Z_k), \Delta Z_k\rangle < \infty.$$

Thus,

$$\lim_{k\to\infty} \alpha_k\|\Delta Z_k\| = 0 \quad \text{and } \lim_{k\to\infty} \alpha_k\|\text{grad } h(Z_k)\|^2 = 0. \tag{15}$$

We now show that (13) holds for all k sufficiently large. By using (9), we find, for all $k \geq 1$,

$$|-\varsigma\|\text{grad } h(Z_k)\|^2| \leq |\langle \Delta Z_k, \text{grad } h(Z_k)\rangle| \leq \|\Delta Z_k\|\|\text{grad } h(Z_k)\|.$$

Thus,

$$\|\text{grad } h(Z_k)\| \leq \frac{1}{\varsigma}\|\Delta Z_k\|.$$

This shows that (13) holds for all k sufficiently large with $\nu = \varsigma^2$ if $\alpha_k \geq 1$ for all k sufficiently large.

Next, we assume that $\alpha_k < 1$ for all k sufficiently large. From **Step 3** of Algorithm 2 and (10), it follows that, for all k sufficiently large,

$$\begin{aligned} h\left(R_{Z_k}\left(\rho^{-1}\alpha_k\Delta Z_k\right)\right) - h(Z_k) &> \delta_1\rho^{-1}\alpha_k\langle \text{grad } h(Z_k), \Delta Z_k\rangle - \delta_2\rho^{-2}\alpha_k^2\|\Delta Z_k\|^2 \\ &\geq \delta_1\rho^{-1}(\varsigma-2)\alpha_k\|\text{grad } h(Z_k)\|^2 - \delta_2\rho^{-2}\alpha_k^2\|\Delta Z_k\|^2. \end{aligned} \tag{16}$$

From (15), we have, for all k sufficiently large,

$$\rho^{-1}\alpha_k\|\Delta Z_k\| \leq \kappa. \tag{17}$$

By using the mean-value theorem, (9), (11), and (17), there exists a $t_k \in (0,1)$ such that, for all k sufficiently large,

$$\begin{aligned} h\left(R_{Z_k}\left(\rho^{-1}\alpha_k\Delta Z_k\right)\right) - h(Z_k) =& h\left(R_{Z_k}\left(\rho^{-1}\alpha_k\Delta Z_k\right)\right) - h(R_{Z_k}(0_{Z_k})) \\ =& \widehat{h}_{Z_k}\left(\rho^{-1}\alpha_k\Delta Z_k\right) - \widehat{h}_{Z_k}(0_{Z_k}) \\ =& \rho^{-1}\alpha_k\langle \text{grad } \widehat{h}_{Z_k}\left(t_k\rho^{-1}\alpha_k\Delta Z_k\right), \Delta Z_k\rangle \\ =& \rho^{-1}\alpha_k\langle \text{grad } \widehat{h}_{Z_k}(0_{Z_k}), \Delta Z_k\rangle \\ &+ \rho^{-1}\alpha_k\langle \Delta Z_k, \text{grad } \widehat{h}_{Z_k}\left(t_k\rho^{-1}\alpha_k\Delta Z_k\right) - \text{grad } \widehat{h}_{Z_k}(0_{Z_k})\rangle \\ \leq& \rho^{-1}\alpha_k\langle \text{grad } \widehat{h}_{Z_k}(0_{Z_k}), \Delta Z_k\rangle + \beta_L t_k\rho^{-2}\alpha_k^2\|\Delta Z_k\|^2 \\ \leq& -\varsigma\rho^{-1}\alpha_k\|\text{grad } h(Z_k)\|^2 + \beta_L\rho^{-2}\alpha_k^2\|\Delta Z_k\|^2. \end{aligned} \tag{18}$$

Combining (16) and (18) yields for all k sufficiently large,

$$\alpha_k > \frac{(\varsigma + \delta_1(\varsigma-2))\rho\|\text{grad } h(Z_k)\|^2}{(\delta_2+\beta_L)\|\Delta Z_k\|^2}.$$

Therefore, (13) holds for all k sufficiently large with $\nu = \min\{\varsigma^2, (\varsigma+\delta_1(\varsigma-2))\rho/(\delta_2+\beta_L)\}$. □

By Lemma 1 and using (14), we have the following Riemannian analogy of the Zoutendijk condition (see the work of Sato et al.[38] for more details).

Lemma 2. *Suppose Assumption 1 is satisfied. Let* $\varsigma := 1 - \mu_1/\mu_2 > 2\delta_1/(1+\delta_1)$. *Then, we have*

$$\sum_{k=0}^{\infty}\frac{\|\text{grad } h(Z_k)\|^4}{\|\Delta Z_k\|^2} < \infty. \tag{19}$$

We now establish the global convergence of Algorithm 2.

Theorem 1. *Suppose Assumption 1 is satisfied and Algorithm 2 generates an infinite sequence* $\{Z_k\}$. *If* $\varsigma := 1 - \mu_1/\mu_2 > 2\delta_1/(1+\delta_1)$ *and* $\omega = \mu_1^2/\mu_3^2 < 1$, *then we have*

$$\liminf_{k\to\infty}\|\text{grad } h(Z_k)\| = 0.$$

Proof. For the sake of contradiction, we assume that there exists a constant $\epsilon > 0$ such that

$$\|\text{grad } h(Z_k)\| \geq \epsilon, \quad \forall k. \tag{20}$$

From (6) and (12), we have, for all $k \geq 1$,

$$
\begin{aligned}
\|\Delta Z_k\|^2 = \| - \text{grad } h(Z_k) + \beta_k^{\text{RVFR}} \Delta Y_k\|^2 =& \|\text{grad } h(Z_k)\|^2 + \left(\beta_k^{\text{RVFR}}\right)^2 \|\Delta Y_k\|^2 - 2\beta_k^{\text{RVFR}} \langle \text{grad } h(Z_k), \Delta Y_k \rangle \\
\leq& \|\text{grad}\, h(Z_k)\|^2 + \left(\frac{\mu_1 \|\text{grad } h(Z_k)\|^2}{\mu_3 \|\text{grad } h(Z_{k-1})\|^2} \right)^2 \|\Delta Y_k\|^2 \\
&+ 2 \frac{\mu_1 \|\text{grad } h(Z_k)\|^2}{\mu_2 |\langle \text{grad } h(Z_k), \Delta Y_k \rangle| + \mu_3 \|\text{grad } h(Z_{k-1})\|^2} |\langle \Delta Y_k, \text{grad } h(Z_k) \rangle| \\
\leq& \|\text{grad } h(Z_k)\|^2 + 2\frac{\mu_1}{\mu_2} \|\text{grad } h(Z_k)\|^2 + \frac{\mu_1^2}{\mu_3^2} \|\text{grad } h(Z_k)\|^4 \frac{\|\Delta Y_k\|^2}{\|\text{grad } h(Z_{k-1})\|^4} \\
\leq& \left(1 + 2\frac{\mu_1}{\mu_2}\right) \|\text{grad } h(Z_k)\|^2 + \omega \|\text{grad } h(Z_k)\|^4 \frac{\|\Delta Z_{k-1}\|^2}{\|\text{grad } h(Z_{k-1})\|^4}.
\end{aligned}
$$

This, together with (20), yields

$$
\begin{aligned}
\frac{\|\Delta Z_k\|^2}{\|\text{grad } h(Z_k)\|^4} &\leq \left(1 + 2\frac{\mu_1}{\mu_2}\right) \frac{1}{\|\text{grad } h(Z_k)\|^2} + \omega \frac{\|\Delta Z_{k-1}\|^2}{\|\text{grad } h(Z_{k-1})\|^4} \\
&\leq 3 \frac{1}{\|\text{grad } h(Z_k)\|^2} + \omega \frac{\|\Delta Z_{k-1}\|^2}{\|\text{grad } h(Z_{k-1})\|^4} \\
&\leq \frac{3}{\epsilon^2} + \omega \frac{3}{\epsilon^2} + \omega^2 \frac{\|\Delta Z_{k-2}\|^2}{\|\text{grad } h(Z_{k-2})\|^4} \\
&\leq \cdots \\
&\leq \frac{3}{\epsilon^2}(1 + \omega + \omega^2 + \cdots + \omega^k) \\
&< \frac{3}{\epsilon^2} \frac{1}{1 - \omega}, \quad \forall k \geq 1,
\end{aligned}
$$

which implies that

$$
\sum_{k=0}^{\infty} \frac{\|\text{grad } h(Z_k)\|^4}{\|(\Delta Z_k)\|^2} \geq \epsilon^2 \sum_{k=0}^{\infty} \frac{1 - \omega}{3} = \infty.
$$

This contradicts the Riemannian Zountendijk condition (19). The proof is complete. □

Finally, we have the following remark on Algorithm 2.

Remark 3. The generalization of the classical FR method (2) and (3) to Riemannian manifolds by using the concepts of retraction and vector transport was also introduced in the works of Absil et al.,[37] Sato et al.,[38] and Ring et al.[39] In the work of Absil et al.,[37(p182)] an algorithm was proposed and its convergence was not discussed. For line searches of the methods in the works of Sato et al.[38] and Ring et al.,[39] the step sizes are required to satisfy some Riemannian Wolfe conditions (see lemma 14 in the work of Ring et al.[39] and algorithm 3.2 in the work of Sato et al.[38]). Moreover, to achieve global convergence, the vector transports are required to satisfy the condition (12) in Assumption 1 and to be constructed through differentiated retraction, which may be computationally costly[37(p172)] (see proposition 15 in the work of Ring et al.[39] and algorithm 3.2 in the work of Sato et al.[38]). However, in Algorithm 2, we use a Riemannian Armijo line search and the used vector transport only needs to satisfy the condition (12) in Assumption 1.

5 | STOCHASTIC IEPS WITH PARTIAL EIGENDATA

In this section, we consider the application of Algorithm 2 to the StIEP-ED and the StIEP-ED-PE. We also use Algorithm 2 to the StIEP-ED and the StIEP-ED-PE where the required matrix $C \in \mathbb{R}^{n \times n}$ is assumed to be column stochastic, that is, an n-by-n nonnegative matrix, with each column summing to 1.

5.1 | StIEP-ED

To apply Algorithm 2 in solving the StIEP-ED, we turn the StIEP-ED into a nonlinear least squares problem on a matrix oblique manifold defined in (23) below. For simplicity, we assume that $\{\lambda_k\}_{k=1}^p$ are all distinct. We note that the prescribed complex pairs $\{(\lambda_k, \mathbf{x}_k) \in \mathbb{C} \times \mathbb{C}^n\}_{k=1}^p$ are closed under complex conjugate. Without loss of generality, we assume that

$$\begin{cases} \lambda_{2j-1} = a_j + \mathrm{i}b_j, \quad \lambda_{2j} = a_j - \mathrm{i}b_j, \quad j = 1, \dots, s, \\ \mathbf{x}_{2j-1} = \mathbf{x}_{jR} + \mathrm{i}\mathbf{x}_{jI}, \quad \mathbf{x}_{2j} = \mathbf{x}_{jR} - \mathrm{i}\mathbf{x}_{jI}, \quad j = 1, \dots, s, \\ \lambda_j \in \mathbb{R}, \quad \mathbf{x}_j \in \mathbb{R}^n, \quad j = 2s+1, \dots, p, \end{cases}$$

where $\mathrm{i} := \sqrt{-1}$, $a_j, b_j \in \mathbb{R}$ with $b_j \neq 0$ and $\mathbf{x}_{jR}, \mathbf{x}_{jI} \in \mathbb{R}^n$ with $\mathbf{x}_{jI} \neq \mathbf{0}$ for $j = 1, \dots, s$. Define the block diagonal matrix

$$\Lambda := \text{blkdiag}\left(\lambda_1^{[2]}, \dots, \lambda_s^{[2]}, \lambda_{2s+1}, \dots, \lambda_p\right),$$

where

$$\lambda_j^{[2]} = \begin{bmatrix} a_j & b_j \\ -b_j & a_j \end{bmatrix}, \quad j = 1, \dots, s.$$

Let

$$X = \left[\mathbf{x}_{1R} \ \mathbf{x}_{1I} \ \dots \ \mathbf{x}_{sR} \ \mathbf{x}_{sI} \ \mathbf{x}_{2s+1} \ \dots \ \mathbf{x}_p\right],$$

which is of full column rank. Hence, the StIEP-ED is to find an n-by-n stochastic matrix C such that

$$CX = X\Lambda.$$

Let the QR decomposition of X be given by

$$X = Q\begin{bmatrix} R_1 \\ 0 \end{bmatrix} \equiv \left[Q_1 \ Q_2\right]\begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

where $Q \in \mathbb{R}^{n\times n}$ is orthogonal with $Q_1 \in \mathbb{R}^{n\times p}$ and $R_1 \in \mathbb{R}^{p\times p}$ is upper triangular and nonsingular. Thus, the StIEP-ED is equivalent to finding an n-by-n stochastic matrix C such that

$$CQ_1 = Q_1R_1\Lambda R_1^{-1} \equiv Y. \tag{21}$$

Suppose that the StIEP-ED has at least one solution. Instead of solving the linear matrix Equation (21) for an n-by-n stochastic matrix C, one may find a global solution of the following least squares problem:

$$\begin{aligned} &\min && \tfrac{1}{2}\|CQ_1 - Y\|_F^2 \\ &\text{subject to (s.t.)} && C \geq 0, \\ & && C\mathbf{e} = \mathbf{e}, \end{aligned} \tag{22}$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm, $C \geq 0$ means that C is a nonnegative matrix, and $\mathbf{e} \in \mathbb{R}^n$ is an n-vector of all ones.

We observe that

$$\{C \in \mathbb{R}^{n\times n} \mid C \geq 0,\ C\mathbf{e} = \mathbf{e}\} = \{C = Z \odot Z \in \mathbb{R}^{n\times n} \mid \text{diag}(ZZ^T) = I_n\}.$$

Define the following oblique manifold[37]:

$$\mathcal{OB} := \{Z \in \mathbb{R}^{n\times n} \mid \text{diag}(ZZ^T) = I_n\}. \tag{23}$$

Then, problem (22) becomes the following nonlinear least squares problem:

$$\begin{aligned} &\min \phi(Z) := \frac{1}{2}\|\Phi(Z)\|_F^2 \\ &\text{s.t.} \quad Z \in \mathcal{OB}, \end{aligned} \tag{24}$$

where the mapping $\Phi : \mathcal{OB} \to \mathbb{R}^{n\times p}$ is defined by

$$\Phi(Z) = (Z \odot Z)Q_1 - Y, \quad Z \in \mathcal{OB}. \tag{25}$$

We see that if $\overline{Z} \in \mathcal{OB}$ is a solution of $\Phi(Z) = 0$, then $\overline{C} = \overline{Z} \odot \overline{Z}$ is a solution of the StIEP-ED. To apply Algorithm 2 in solving problem (24), we need some basic properties of the oblique manifold $\mathcal{OB}$, the differential of Φ defined in (25), and the Riemannian gradient of ϕ defined in problem (24).

We note that

$$\dim(\mathcal{OB}) = n(n-1) > \dim(\mathbb{R}^{n\times p}) \quad \text{for } p < n-1.$$

Thus, the nonlinear equation $\Phi(Z) = 0$ is underdetermined for $p < n-1$.

The tangent space of $\mathcal{OB}$ at point $Z \in \mathcal{OB}$ is given by[37(p42),40]

$$T_Z\mathcal{OB} = \{K \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(ZK^T) = 0\}.$$

A retraction R on $\mathcal{OB}$ takes the form of[37(p58),40]

$$R_Z(\xi_Z) = \left(\mathrm{diag}\left((Z+\xi_Z)(Z+\xi_Z)^T\right)\right)^{-\frac{1}{2}}(Z+\xi_Z) \quad \text{for } Z \in \mathcal{OB},\ \xi_Z \in T_Z\mathcal{OB}. \tag{26}$$

We now define the Riemannian metric on $\mathcal{OB}$. It is obvious that $\mathcal{OB}$ is an embedded submanifold of the Euclidean space $\mathbb{R}^{n\times n}$. A Riemannian metric of $\mathcal{OB}$ is inherited from $\mathbb{R}^{n\times n}$, which is given by

$$g_Z(\xi_Z, \eta_Z) := \mathrm{tr}\left(\xi_Z^T \eta_Z\right) \tag{27}$$

for all $Z \in \mathcal{OB}$ and $\xi_Z, \eta_Z \in T_Z\mathcal{OB}$. The orthogonal projection of any given point $\xi \in \mathbb{R}^{n\times n}$ onto $T_Z\mathcal{OB}$ is given by[37(p48),40]

$$\Pi_Z\xi = \xi - \mathrm{diag}(Z\xi^T)Z.$$

To transport a tangent vector $\xi_Z \in T_Z\mathcal{OB}$ from a point $Z \in \mathcal{OB}$ to another point $R_Z(\eta) \in \mathcal{OB}$, one may use parallel translation (see for instance the work of Absil et al.[37(p104)]), which is in general computationally costly.[37(pp103,169)] To improve efficiency, we use the vector transport on Riemannian submanifolds.[37(p174)] The vector transport on $\mathcal{OB}$ takes the form of[37(p174)]

$$\mathcal{T}_{\eta_Z}\xi_Z = \Pi_{R_Z(\eta_Z)}\xi_Z = \xi_Z - \mathrm{diag}\left(R_Z(\eta_Z)\xi_Z^T\right) R_Z(\eta_Z) \quad \text{for } \xi_Z, \eta_Z \in T_Z\mathcal{OB}. \tag{28}$$

Next, we establish the differential of Φ and the Riemannian gradient of ϕ. By simple calculation, the differential $\mathrm{D}\Phi(Z) : T_Z\mathcal{OB} \to T_{\Phi(Z)}\mathbb{R}^{n\times p} \simeq \mathbb{R}^{n\times p}$ of Φ at a point $Z \in \mathcal{OB}$ is determined by

$$\mathrm{D}\Phi(Z)[\Delta Z] = \Pi_Z(2(Z \odot \Delta Z)Q_1), \quad \forall \Delta Z \in T_Z\mathcal{OB}.$$

With respect to the Riemannian metric g on $\mathcal{OB}$, the adjoint of $D\Phi(Z)$ is given by

$$(\mathrm{D}\Phi(Z))^*[\Delta W] = \Pi_Z\left(2Z \odot \left(\Delta W Q_1^T\right)\right), \quad \forall \Delta W \in T_{\Phi(Z)}\mathbb{R}^{n\times p}.$$

The Riemannian gradient of ϕ at a point $Z \in \mathcal{OB}$ is given by[37(p185)]

$$\mathrm{grad}\ \phi(Z) = (\mathrm{D}\Phi(Z))^*[\Phi(Z)]. \tag{29}$$

For the level set of ϕ defined in (24), we have the following result. The proof is similar to lemma 3.1 of the work of Zhao et al.[29] Therefore, we omit it here.

Lemma 3. *For any given point $Z_0 \in \mathcal{OB}$, the level set*

$$\Omega := \{Z \in \mathcal{OB} \mid \phi(Z) \le \phi(Z_0)\}$$

is compact.

We now verify that Assumption 1 holds for ϕ defined in (24), the retraction (26), and the vector transport (28). We first note that the function ϕ is bounded below by zero. Because $\mathcal{OB}$ is an embedded Riemannian submanifold of $\mathbb{R}^{n\times n}$, for any $Z \in \mathcal{OB}$, $T_Z\mathcal{OB}$ can be regarded as a subspace of $\mathbb{R}^{n\times n}$. The Riemannian gradient grad ϕ can be seen as a continuous nonlinear mapping between $\mathcal{OB}$ and $\mathbb{R}^{n\times p}$. We also note that both ϕ and R are smooth and, thus, $\hat{\phi}$ is smooth. Because the level set Ω is compact, the Lipschitz condition (11) holds for $\hat{\phi}$. In addition, because the vector transport in (28) is defined via orthogonal projection, the condition (12) in Assumption 1 is naturally satisfied. Therefore, all the conditions in Assumption 1 are satisfied. Therefore, one may apply Algorithm 2 in solving problem (24), and its global convergence is guaranteed.

Remark 4. In **Step 3** of Algorithm 2 for solving problem (24), the initial step length of α_k is set to be α. As in the work of Zhang et al.,[41] one may take a reasonable initial step length guess as follows:

$$s_k = \left| \frac{\langle \text{grad}\, \phi(Z_k), \Delta Z_k \rangle}{\|\mathrm{D}\Phi(Z_k)[\Delta Z_k]\|^2} \right|, \tag{30}$$

where Φ is defined by (25). From the numerical tests in Section 6, we observe that Algorithm 2 with the initial step length guess (30) is very effective for solving problem (24).

We have the following result on the condition under which an accumulation point of the sequence $\{Z_k\}$ generated by Algorithm 2 for problem (24) is a solution to $\Phi(Z) = 0$.

Theorem 2. *Let $\overline{Z}$ be an accumulation point of the sequence $\{Z_k\}$ generated by Algorithm 2 for problem (24) such that grad $\phi(\overline{Z}) = 0_{\overline{Z}}$. Define $\widehat{T}, \widetilde{T} \in \mathbb{R}^{n^2 \times n^2}$ by*

$$\mathrm{vec}(A^T) = \widehat{T}\mathrm{vec}(A) \quad \textit{and} \quad \mathrm{vec}(\mathrm{diag}(A)) = \widetilde{T}\mathrm{vec}(A)$$

for all $A \in \mathbb{R}^{n\times n}$. Suppose that $\varsigma := 1 - \mu_1/\mu_2 > 2\delta_1/(1+\delta_1)$ and $\omega = \mu_1^2/\mu_3^2 < 1$. Then, $\overline{Z}$ is a solution to $\Phi(Z) = 0$ if the matrix

$$\left(I_{n^2} - \left(\overline{Z}^T \otimes I_n\right) \widetilde{T}\widehat{T} \left(\overline{Z} \otimes I_n\right)\right) \mathrm{Diag}(\mathrm{vec}(\overline{Z}))(Q_1 \otimes I_n)$$

is of full rank. In this case, $\overline{C} := \overline{Z} \odot \overline{Z}$ is a solution to the StIEP-ED.

Proof. By hypothesis, $\overline{Z}$ is a stationary point of ϕ, that is, grad $\phi(\overline{Z}) = 0_{\overline{Z}}$. It follows from (29) that

$$(\mathrm{D}\Phi(\overline{Z}))^*[\Phi(\overline{Z})] = 2\overline{Z} \odot \left(\Phi(\overline{Z})Q_1^T\right) - 2\mathrm{diag}\left(\overline{Z}\left(\overline{Z} \odot \left(\Phi(\overline{Z})Q_1^T\right)\right)^T\right)\overline{Z} = 0_{\overline{Z}}. \tag{31}$$

Thus, $\overline{Z}$ is a solution to $\Phi(Z) = 0$ if

$$\ker((\mathrm{D}\Phi(\overline{Z}))^*) = \{0\}, \tag{32}$$

where "ker" means the kernel of a linear map. By vectoring the matrix Equation (31), we obtain

$$\left(I_{n^2} - \left(\overline{Z}^T \otimes I_n\right) \widetilde{T}\widehat{T} \left(\overline{Z} \otimes I_n\right)\right) \mathrm{Diag}(\mathrm{vec}(\overline{Z}))(Q_1 \otimes I_n)\mathrm{vec}(\Phi(\overline{Z})) = \mathbf{0}_{\mathbf{n}^2},$$

where $\mathbf{0}_{\mathbf{n}^2}$ is a zero vector of order n^2. Therefore, (32) holds if and only if

$$\left(I_{n^2} - \left(\overline{Z}^T \otimes I_n\right) \widetilde{T}\widehat{T} \left(\overline{Z} \otimes I_n\right)\right) \mathrm{Diag}(\mathrm{vec}(\overline{Z}))(Q_1 \otimes I_n)$$

is of full rank. This completes the proof. □

5.2 | StIEP-ED-PE

In this subsection, we consider the application of Algorithm 2 to the StIEP-ED-PE. Let Λ, X, Q_1, and R_1 be defined as in Section 5.1. Then, the StIEP-ED-PE is to find an n-by-n stochastic matrix C such that

$$CQ_1 = Q_1R_1\Lambda R_1^{-1} \equiv Y \quad \text{and} \quad C_{ij} = (C_a)_{ij}, \quad \forall (i,j) \in \mathcal{L}.$$

Suppose that the StIEP-ED-PE has at least a solution. Then, the StIEP-ED-PE aims to find a global solution of the following least squares problem:

$$\begin{aligned} \min \ & \frac{1}{2}\|CQ_1 - Y\|_F^2 \\ \text{s.t.} \ & C \geq 0, \\ & C\mathbf{e} = \mathbf{e}, \\ & C_{ij} = (C_a)_{ij}, \quad \forall (i,j) \in \mathcal{L}. \end{aligned} \tag{33}$$

Define the matrix $G \in \mathbb{R}^{n\times n}$ by $G_{ij} = 1$ if $(i,j) \in \mathcal{L}$; 0, otherwise. Let

$$\widehat{C}_a := G \odot C_a \quad \text{and} \quad \widehat{I}_n := I_n - \mathrm{Diag}(\widehat{C}_a\mathbf{e}).$$

Here, we assume that the given index subset $\mathcal{L}$ is such that $\sum_{j=1}^{n}(\widehat{C}_a)_{ij} < 1$ for $i = 1, \dots, n$. In this case, $\widehat{I}_n$ is nonsingular. Then, the StIEP-ED-PE aims to find a global solution of the following nonlinear least squares problem:

$$\begin{aligned} \min \quad & \psi(Z) := \frac{1}{2}\|\Psi(Z)\|_F^2 \\ \text{s.t.} \quad & Z \in \widehat{\mathcal{OB}}, \end{aligned} \tag{34}$$

where $\widehat{\mathcal{OB}} := \{Z \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(ZZ^T) = \widehat{I}_n,\ G \odot Z = 0\}$ and Ψ is mapping from $\widehat{\mathcal{OB}}$ to $\mathbb{R}^{n\times p}$, which is defined by

$$\Psi(Z) := (\widehat{C}_a + Z \odot Z)Q_1 - Y, \quad Z \in \widehat{\mathcal{OB}}.$$

We note that $\dim(\widehat{\mathcal{OB}}) = n(n-1) - |\mathcal{L}|$, where $|\mathcal{L}|$ means the cardinality of $\mathcal{L}$. Hence, the nonlinear matrix equation $\Psi(Z) = 0$ is underdetermined if n is large and the number of prescribed entries is small.

We point out that if $\overline{Z} \in \widehat{\mathcal{OB}}$ is a global solution to problem (34), that is, $\Psi(\overline{Z}) = 0$, then $\overline{C} = \widehat{C}_a + \overline{Z} \odot \overline{Z}$ is a solution to the StIEP-ED-PE.

We note that $\widehat{\mathcal{OB}}$ is a smooth manifold. The tangent space of $\widehat{\mathcal{OB}}$ at a point $Z \in \widehat{\mathcal{OB}}$ is given by

$$T_Z\widehat{\mathcal{OB}} = \{K \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(ZK^T) = 0,\ G \odot K = 0\}.$$

The retraction on $\widehat{\mathcal{OB}}$ at a point $Z \in \widehat{\mathcal{OB}}$ can be defined by

$$\widehat{R}_Z(\xi_Z) = \widehat{I}_n^{\frac{1}{2}}\left(\mathrm{diag}\left((Z+\xi_Z)(Z+\xi_Z)^T\right)\right)^{-\frac{1}{2}}(Z+\xi_Z), \quad \forall \xi_Z \in T_Z\widehat{\mathcal{OB}}.$$

Let $\widehat{\mathcal{OB}}$ be equipped with the Riemannian metric defined as in (27). Then, the orthogonal projection of a point $\xi \in \mathbb{R}^{n\times n}$ onto $T_Z\widehat{\mathcal{OB}}$ is given by

$$\widehat{\Pi}_Z\xi = (E-G)\odot\xi - \widehat{I}_n^{-1}\mathrm{diag}(Z((E-G)\odot\xi)^T)Z,$$

where E is an n-by-n matrix of ones. In addition, we adopt the vector transport on $\widehat{\mathcal{OB}}$,

$$\widehat{\mathcal{T}}_{\eta_Z}(\xi_Z) := \widehat{\Pi}_{\widehat{R}_Z(\eta_Z)}\xi_Z,$$

for any $Z \in \widehat{\mathcal{OB}}$ and $\xi_Z, \eta_Z \in T_Z\widehat{\mathcal{OB}}$.

The differential $\mathrm{D}\Psi(Z) : T_Z\widehat{\mathcal{OB}} \to T_{\Psi(Z)}\mathbb{R}^{n\times n} \simeq \mathbb{R}^{n\times n}$ of Ψ at a point $Z \in \widehat{\mathcal{OB}}$ is determined by

$$\mathrm{D}\Psi(Z)[\Delta Z] = \widehat{\Pi}_Z(2(Z \odot \Delta Z)Q_1), \quad \forall \Delta Z \in T_Z\widehat{\mathcal{OB}}$$

and the Riemannian gradient of ψ at a point $Z \in \widehat{\mathcal{OB}}$ is given by

$$\mathrm{grad}\,\psi = (\mathrm{D}\Psi(Z))^*[\Psi(Z)].$$

As in Section 5.1, one may apply Algorithm 2 to problem (34). The corresponding global convergence can be established similarly.

5.3 | The case of the column stochastic matrix

We first consider the StIEP-ED where the required matrix $C \in \mathbb{R}^{n\times n}$ is assumed to be column stochastic. Let Λ, X, Q_1, and R_1 be defined as in Section 5.1. Then, the StIEP-ED with column stochastic matrix $C \in \mathbb{R}^{n\times n}$ aims to find a global solution to the following minimization problem:

$$\begin{aligned} \min \ & \phi^c(Z) := \frac{1}{2}\|\Phi^c(Z)\|_F^2 \\ \text{s.t.} \ & Z \in \mathcal{OB}_c, \end{aligned} \tag{35}$$

where $\mathcal{OB}_c := \{Z \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(Z^TZ) = I_n\}$ and the mapping $\Phi^c : \mathcal{OB}_c \to \mathbb{R}^{n\times p}$ is defined by

$$\Phi^c(Z) = (Z \odot Z)Q_1 - Y, \quad Z \in \mathcal{OB}_c.$$

It follows that the tangent space of $\mathcal{OB}_c$ at point $Z \in \mathcal{OB}_c$ is given by

$$T_Z\mathcal{OB}_c = \{K \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(Z^TK) = 0\}.$$

A retraction R^c on $\mathcal{OB}_c$ takes the form of

$$R_Z^c(\xi_Z) = (Z+\xi_Z)\left(\mathrm{diag}\left((Z+\xi_Z)^T(Z+\xi_Z)\right)\right)^{-\frac{1}{2}} \quad \text{for } Z \in \mathcal{OB}_c,\ \xi_Z \in T_Z\mathcal{OB}_c.$$

Let $\mathcal{OB}_c$ be equipped with the Riemannian metric defined as in (27). Then, the orthogonal projection of a point $\xi \in \mathbb{R}^{n\times n}$ onto $T_Z\mathcal{OB}_c$ is given by

$$\Pi_Z^c \xi = \xi - Z\mathrm{diag}(Z^T\xi).$$

The Riemannian gradient of ϕ^c at a point $Z \in \mathcal{OB}_c$ is given by

$$\mathrm{grad}\ \phi^c(Z) = (\mathrm{D}\Phi^c(Z))^*[\Phi^c(Z)],$$

where the differential $\mathrm{D}\Phi^c(Z) : T_Z\mathcal{OB}_c \to T_{\Phi^c(Z)}\mathbb{R}^{n\times p} \simeq \mathbb{R}^{n\times p}$ of Φ^c at a point $Z \in \mathcal{OB}_c$ is determined by

$$\mathrm{D}\Phi^c(Z)[\Delta Z] = \Pi_Z^c(2(Z \odot \Delta Z)Q_1), \quad \forall \Delta Z \in T_Z\mathcal{OB}_c.$$

Next, we consider the StIEP-ED-PE with a column stochastic matrix. Let Λ, X, Q_1, and R_1 be defined as in Section 5.1. Let G and $\widehat{C}_a$ be defined as in Section 5.2. Suppose that the given index subset $\mathcal{L}$ is such that $\sum_{i=1}^n (\widehat{C}_a)_{ij} < 1$ for $j = 1, \ldots, n$. Then, the StIEP-ED-PE with a column stochastic matrix is to solve the following minimization problem:

$$\begin{aligned} \min \quad & \psi^c(Z) := \frac{1}{2}\|\Psi^c(Z)\|_F^2 \\ \text{s.t.} \quad & Z \in \widehat{\mathcal{OB}}_c, \end{aligned} \tag{36}$$

where $\widehat{I}_n^c := I_n - \mathrm{Diag}(\widehat{C}_a^T\mathbf{e})$. $\widehat{\mathcal{OB}}_c := \{Z \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(Z^TZ) = \widehat{I}_n,\ G \odot Z = 0\}$ and Ψ^c is mapping from $\widehat{\mathcal{OB}}_c$ to $\mathbb{R}^{n\times p}$, which is defined by

$$\Psi^c(Z) := (\widehat{C}_a + Z \odot Z)Q_1 - Y, \quad Z \in \widehat{\mathcal{OB}}_c.$$

It is easy to know that the tangent space of $\widehat{\mathcal{OB}}_c$ at point $Z \in \widehat{\mathcal{OB}}_c$ is given by

$$T_Z\widehat{\mathcal{OB}}_c = \{K \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(Z^TK) = 0,\ G \odot K = 0\}.$$

A retraction $\widehat{R}^c$ on $\widehat{\mathcal{OB}}_c$ takes the form of

$$\widehat{R}_Z^c(\xi_Z) = (Z + \xi_Z)\left(\mathrm{diag}\left((Z+\xi_Z)^T(Z+\xi_Z)\right)\right)^{-\frac{1}{2}}\left(\widehat{I}_n^c\right)^{\frac{1}{2}} \quad \text{for } Z \in \widehat{\mathcal{OB}}_c,\ \xi_Z \in T_Z\widehat{\mathcal{OB}}_c.$$

Let $\widehat{\mathcal{OB}}_c$ be equipped with the Riemannian metric defined as in (27). Then, the orthogonal projection of a point $\xi \in \mathbb{R}^{n\times n}$ onto $T_Z\widehat{\mathcal{OB}}_c$ is given by

$$\widehat{\Pi}_Z^c\xi = (E - G) \odot \xi - Z\mathrm{diag}\left(Z^T((E-G)\odot\xi)\right)\left(\widehat{I}_n^c\right)^{-1}.$$

The Riemannian gradient of ψ^c at a point $Z \in \widehat{\mathcal{OB}}_c$ is given by

$$\mathrm{grad}\ \psi^c(Z) = (\mathrm{D}\Psi^c(Z))^*[\Psi^c(Z)],$$

where the differential $\mathrm{D}\Psi^c(Z) : T_Z\widehat{\mathcal{OB}}_c \to T_{\Psi^c(Z)}\mathbb{R}^{n\times p} \simeq \mathbb{R}^{n\times p}$ of Ψ^c at a point $Z \in \widehat{\mathcal{OB}}_c$ is determined by

$$\mathrm{D}\Psi^c(Z)[\Delta Z] = \widehat{\Pi}_Z^c(2(Z \odot \Delta Z)Q_1), \quad \forall \Delta Z \in T_Z\widehat{\mathcal{OB}}_c.$$

As in Section 5.1, we can use Algorithm 2 to problem (35) and problem (36). We can also show the global convergence by a similar way.

6 | NUMERICAL TESTS

In this section, we report some numerical results to illustrate the effectiveness of Algorithm 2 for solving problem (24) and problem (34). Our numerical experiments are implemented in MATLAB R2016a running on a workstation with an Intel Xeon CPU E5-2687 W at 3.10 GHz and 32 GB of RAM.

We consider the following two examples.

Example 1. We consider the StIEP-ED with varying n. Let $\widetilde{C}$ be a random $n \times n$ nonnegative matrix with each entry generated from the uniform distribution on the interval [0, 1]. Let $\widehat{C}$ be a random stochastic matrix given by

$$\widehat{C} := (\mathrm{diag}(\widetilde{C}\widetilde{C}^T))^{-\frac{1}{2}}\widetilde{C}.$$

We choose the first p eigenvalues of $\widehat{C}$ with largest absolute values and associated eigenvectors as prescribed eigendata.

Example 2. We consider the StIEP-ED-PE with varying n. Let $\widehat{C}$ be a random $n \times n$ stochastic matrix generated as in Example 1. We choose the first p eigenvalues of $\widehat{C}$ with largest absolute values and associated eigenvectors as prescribed eigendata. In addition, we choose the index subset $\mathcal{L} := \{(i,j) \mid 3/(5n) < \widehat{C}_{ij} < 4/(5n),\ i,j = 1, \dots, n\}$. The prescribed nonnegative matrix $C_a \in \mathbb{R}^{n\times n}$ is defined by $(C_a)_{ij} = \widehat{C}_{ij}$, if $(i,j) \in \mathcal{L}$; 0, otherwise.

In our numerical tests, we compare the numerical performances of Algorithm 2, the Riemannian FR conjugate gradient method (RFR)[27] and the Riemannian PRP-based nonlinear conjugate gradient method (RPRP)[29] for solving problem (24) and problem (34).

The starting points of Algorithm 2, RFR, and RPRP are generated randomly by the built-in functions `rand`. For problem (24),

$$\widehat{Z} \odot \widehat{Z} = \texttt{rand}(n,n),\ Z_0 = (\mathrm{diag}(\widehat{Z}\widehat{Z}^T))^{-\frac{1}{2}}\widehat{Z} \in \mathcal{OB}.$$

For problem (34),

$$\widehat{Z} \odot \widehat{Z} = \texttt{rand}(n,n),\ Z_0 = \widehat{I}_n^{\frac{1}{2}}(\mathrm{diag}(((E-G)\odot \widehat{Z})((E-G)\odot \widehat{Z})^T))^{-\frac{1}{2}}((E-G)\odot \widehat{Z}) \in \widehat{\mathcal{OB}}.$$

The stopping criteria of Algorithm 2 for solving problems (24) and (34) are respectively set to be

$$\|\Phi(Z_k)\|_F \le 10^{-12} \quad \text{and} \quad \|\Psi(Z_k)\|_F \le 10^{-12}. \tag{37}$$

We also set other parameters used in Algorithm 2 as $\alpha = 1.4$, $\rho = 0.5$, $\delta_1 = 10^{-3}$, $\delta_2 = 10^{-8}$, $\mu_1 = 1.0$, $\mu_2 = \mu_3 = 1.2$.

The numerical results for Examples 1 and 2 are listed in Tables 1–4, where "**CT.**," "**IT.**," "**NF.**," "**Res.**," and "**grad.**" mean the total computing time, the number of iterations, the number of function evaluations, the residual $\|\Phi(Z_k)\|_F$ or $\|\Psi(Z_k)\|_F$, and the residual $\|\mathrm{grad}\phi(Z_k)\|$ or $\|\mathrm{grad}\psi(Z_k)\|$ at the final iterate of the corresponding algorithms accordingly.

We see from Tables 1 and 2 that, when no initial guess of the step length α_k as (30) is used, Algorithm 2 works much better than RRPR and RFR in terms of both the number of iterations and the total computing time for most cases. We see from Tables 3 and 4 that, when an initial guess of the step length α_k as (30) is used, Algorithm 2, RRPR, and RFR are all efficient in terms of both the number of iterations and the total computing time. We also observe that the initial step length guess as (30) can drastically reduce the number of iterations and thus improves the efficiency.

To further illustrate the effectiveness of our proposed algorithm, we compare Algorithm 2 with the interior-point algorithm in the work of Tütüncü et al.[42] for solving the StIEP-ED and the StIEP-ED-PE. To apply the package SDPT3 in

TABLE 1 Numerical results for Example 1 without the initial step length guess

Alg.	n	p	CT.	IT.	NF.	Res.	grad.
RPRP	100	21	0.9470 s	978	979	9.9×10^{-13}	1.3×10^{-13}
	200	31	4.5440 s	1,627	1,628	9.9×10^{-13}	1.0×10^{-13}
	400	41	45.98 s	2,732	2,733	9.9×10^{-13}	7.9×10^{-14}
	600	61	2 min 49 s	4,090	4,091	9.9×10^{-13}	6.5×10^{-14}
	800	81	7 min 23 s	5,443	5,444	9.9×10^{-13}	5.6×10^{-14}
	1,000	100	17 min 14 s	6,735	6,736	9.9×10^{-13}	5.1×10^{-14}
RFR	100	21	0.6600 s	695	696	9.7×10^{-13}	1.6×10^{-13}
	200	31	3.2960 s	1,306	1,307	9.9×10^{-13}	1.2×10^{-13}
	400	41	34.69 s	2,385	2,386	9.9×10^{-13}	8.7×10^{-14}
	600	61	2 min 10 s	3,577	3,578	9.9×10^{-13}	7.1×10^{-14}
	800	81	5 min 40 s	4,784	4,785	9.9×10^{-13}	6.2×10^{-14}
	1,000	100	12 min 18 s	5,901	5,902	9.9×10^{-13}	5.6×10^{-14}
Alg. 2	100	21	0.6190 s	653	654	9.7×10^{-13}	1.3×10^{-13}
	200	31	2.7480 s	1,085	1,086	9.9×10^{-13}	1.0×10^{-13}
	400	41	26.45 s	1,822	1,823	9.9×10^{-13}	7.9×10^{-14}
	600	61	1 min 39 s	2,727	2,728	9.9×10^{-13}	6.5×10^{-14}
	800	81	4 min 19 s	3,629	3,630	9.9×10^{-13}	5.6×10^{-14}
	1,000	100	9 min 20 s	4,490	4,491	9.9×10^{-13}	5.1×10^{-14}

Note. RPRP = Riemannian Polak–Ribière–Polyak; RFR = Riemannian Fletcher–Reeves.

TABLE 2 Numerical results for Example 2 without the initial step length guess

Alg.	n	p	CT.	IT.	NF.	Res.	grad.
RPRP	100	21	1.5290 s	1,463	1,464	9.9×10^{-13}	1.1×10^{-13}
	200	30	6.5520 s	2,082	2,083	9.9×10^{-13}	8.7×10^{-14}
	400	41	1 min 11 s	3,382	3,383	9.9×10^{-13}	6.9×10^{-14}
	600	60	4 min 33 s	4,907	4,908	9.9×10^{-13}	5.9×10^{-14}
	800	80	12 min 20 s	6,476	6,477	9.9×10^{-13}	5.1×10^{-14}
	1,000	100	29 min 00 s	8,085	8,086	9.9×10^{-13}	4.6×10^{-14}
RFR	100	21	0.8710 s	8,86	8,87	9.9×10^{-13}	1.5×10^{-13}
	200	30	4.0030 s	1,525	1,526	9.9×10^{-13}	1.1×10^{-13}
	400	41	46.52 s	2,800	2,801	9.9×10^{-13}	8.1×10^{-14}
	600	60	2 min 52 s	4,065	4,066	9.9×10^{-13}	6.7×10^{-14}
	800	80	7 min 45 s	5,369	5,370	9.9×10^{-13}	5.8×10^{-14}
	1,000	100	17 min 36 s	6,724	6,725	9.9×10^{-13}	5.2×10^{-14}
Alg. 2	100	21	0.9570 s	976	977	9.8×10^{-13}	1.0×10^{-13}
	200	30	3.6560 s	1,388	1,389	9.9×10^{-13}	8.7×10^{-14}
	400	41	37.41 s	2,255	2,256	9.9×10^{-13}	6.9×10^{-14}
	600	60	2 min 19 s	3,272	3,273	9.9×10^{-13}	5.8×10^{-14}
	800	80	6 min 18 s	4,318	4,319	9.9×10^{-13}	5.1×10^{-14}
	1,000	100	14 min 01 s	5,390	5,391	9.9×10^{-13}	4.6×10^{-14}

Note. RPRP = Riemannian Polak–Ribière–Polyak; RFR = Riemannian Fletcher–Reeves.

the work of Tütüncü et al.[42] to our problems, one may rewrite problems (22) and (33) as

$$\begin{aligned} &\min \frac{1}{2}z \\ &\text{s.t. } C\mathbf{e} = \mathbf{e}, \\ &\quad \sqrt{z} \geq \|CQ_1 - Y\|_F, \\ &\quad C \geq 0 \end{aligned}$$

and

$$\begin{aligned} &\min \frac{1}{2}z \\ &\text{s.t. } C\mathbf{e} = \mathbf{e}, \\ &\quad C_{ij} = (C_a)_{ij}, \quad \forall (i,j) \in \mathcal{L}, \\ &\quad \sqrt{z} \geq \|CQ_1 - Y\|_F, \\ &\quad C \geq 0. \end{aligned}$$

For simplicity, for Algorithm 2, the stopping criterion is set to be as in (37) and the initial step length guess is used. For the interior-point algorithm in the work of Tütüncü et al.,[42] we use the default tolerance. The numerical results for Examples 1 and 2 are displayed in Tables 5 and 6. Here, “LS failed” means that the step length is too short to proceed during the computation.

We observe from Tables 5 and 6 that Algorithm 2 works much better than the interior-point algorithm in the work of Tütüncü et al.[42]

Finally, we consider two examples in Criminology and Random Walk.

Example 3. A study of male criminals in Philadelphia[43] showed that the offenses were classified into five types: nonindex, index injury, index theft, index damage, and index combination. In particular, the transition matrix (a column stochastic matrix) is given by

$$C = \begin{array}{r} \\ \text{Nonindex} \\ \text{Injury} \\ \text{Theft} \\ \text{Damage} \\ \text{Combination} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Nonindex} & \text{Injury} & \text{Theft} & \text{Damage} & \text{Combination} \end{array} \\ \begin{bmatrix} 0.6450 & ? & ? & ? & ? \\ ? & 0.1380 & ? & ? & ? \\ ? & ? & 0.2710 & ? & ? \\ ? & ? & ? & 0.0640 & ? \\ ? & ? & ? & ? & 0.1790 \end{bmatrix} \end{array}. \tag{38}$$

TABLE 3 Numerical results for Example 1 with the initial step length guess

Alg.	n	p	CT.	IT.	NF.	Res.	grad.
RPRP	100	21	0.0350 s	27	28	5.3×10^{-13}	1.0×10^{-13}
	200	31	0.0680 s	23	24	6.1×10^{-13}	8.3×10^{-14}
	400	41	0.3470 s	19	20	3.8×10^{-13}	3.7×10^{-14}
	600	61	0.8730 s	19	20	4.7×10^{-13}	3.7×10^{-14}
	800	81	1.6820 s	19	20	5.2×10^{-13}	3.6×10^{-14}
	1,000	100	3.1110 s	19	20	2.5×10^{-13}	1.5×10^{-14}
	2,000	140	11.766 s	16	17	6.5×10^{-13}	2.8×10^{-14}
	4,000	181	1 min 11 s	14	15	6.0×10^{-13}	1.9×10^{-14}
	6,000	221	3 min 10 s	13	14	9.9×10^{-13}	2.5×10^{-14}
	8,000	260	7 min 16 s	13	14	4.1×10^{-13}	9.0×10^{-15}
	10,000	301	13 min 21 s	13	14	2.3×10^{-13}	4.5×10^{-15}
RFR	100	21	0.0340 s	27	28	5.3×10^{-13}	1.0×10^{-13}
	200	31	0.0630 s	23	24	6.1×10^{-13}	8.4×10^{-14}
	400	41	0.3010 s	19	20	3.8×10^{-13}	3.7×10^{-14}
	600	61	0.7450 s	19	20	4.7×10^{-13}	3.7×10^{-14}
	800	81	1.4700 s	19	20	5.2×10^{-13}	3.6×10^{-14}
	1,000	100	2.4990 s	19	20	2.5×10^{-13}	1.5×10^{-14}
	2,000	140	9.8000 s	16	17	6.5×10^{-13}	2.8×10^{-14}
	4,000	181	54.608 s	14	15	6.0×10^{-13}	1.9×10^{-14}
	6,000	221	2 min 34 s	13	14	9.9×10^{-13}	2.5×10^{-14}
	8,000	260	5 min 35 s	13	14	4.1×10^{-13}	9.0×10^{-15}
	10,000	301	10 min 11 s	13	14	2.3×10^{-13}	4.5×10^{-15}
Alg. 2	100	21	0.0360 s	29	30	6.4×10^{-13}	1.2×10^{-13}
	200	31	0.0640 s	24	25	9.5×10^{-13}	1.3×10^{-13}
	400	41	0.3040 s	19	20	9.4×10^{-13}	9.2×10^{-14}
	600	60	0.7890 s	21	22	3.4×10^{-13}	2.7×10^{-14}
	800	81	1.5130 s	20	21	3.5×10^{-13}	2.4×10^{-14}
	1,000	100	2.4910 s	19	20	6.7×10^{-13}	4.2×10^{-14}
	2,000	140	10.345 s	17	18	2.3×10^{-13}	1.0×10^{-14}
	4,000	181	55.536 s	15	16	1.5×10^{-13}	4.7×10^{-15}
	6,000	221	2 min 45 s	14	15	2.1×10^{-13}	5.4×10^{-15}
	8,000	260	5 min 32 s	13	14	6.8×10^{-13}	1.5×10^{-14}
	10,000	301	9 min 59 s	13	14	3.9×10^{-13}	7.7×10^{-15}

Note. RPRP = Riemannian Polak–Ribière–Polyak; RFR = Riemannian Fletcher–Reeves.

Suppose that the long-term probabilities for each type of crime are as follows: 0.6068 for nonindex, 0.0969 for injury, 0.1740 for theft, 0.0325 for damage, and 0.0898 for combination. Determine the entries in the transition matrix C denoted by "?."

Example 4. Consider a one-dimensional random walk model.[29,44,45] Suppose that, from state i, a particle moves one unit to the right (state $i+1$) with probability p_i, moves to the left (state $i-1$) with probability q_i, or stays in state i with probability $r_i = 1 - p_i - q_i$. The particle starts at the origin and is absorbed in state $n-1$. The transition probabilities are described in the following transition matrix (a column stochastic matrix)

$$C = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \\ n-1 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & n-1 \end{array} \\ \begin{bmatrix} r_0 & q_1 & 0 & 0 & \cdots & 0 \\ p_0 & r_1 & q_2 & 0 & \cdots & 0 \\ 0 & p_1 & r_2 & q_3 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{n-3} & r_{n-2} & q_{n-1} \\ 0 & 0 & \cdots & 0 & p_{n-2} & r_{n-1} \end{bmatrix} \end{array}.$$

Suppose that $n = 10$ and the associated Markov chain has a stationary distribution vector

$$\boldsymbol{\pi} = \begin{bmatrix} 0.0264 & 0.0308 & 0.0835 & 0.1740 & 0.2230 & 0.1680 & 0.1161 & 0.0730 & 0.0490 & 0.0562 \end{bmatrix}^T.$$

TABLE 4 Numerical results for Example 2 with the initial step length guess

Alg.	n	p	CT.	IT.	NF.	Res.	grad.
RPRP	100	21	0.0440 s	33	34	6.2×10^{-13}	1.1×10^{-13}
	200	30	0.0980 s	26	27	8.0×10^{-13}	1.0×10^{-13}
	400	41	0.5080 s	22	23	7.6×10^{-13}	7.0×10^{-14}
	600	60	1.2730 s	21	22	8.7×10^{-13}	6.5×10^{-14}
	800	80	2.7270 s	21	22	4.0×10^{-13}	2.6×10^{-14}
	1,000	100	4.7340 s	21	22	4.0×10^{-13}	2.3×10^{-14}
	2,000	140	19.002 s	18	19	4.9×10^{-13}	2.0×10^{-14}
	4,000	180	1 min 49 s	16	17	1.8×10^{-13}	5.4×10^{-15}
	6,000	221	5 min 03 s	15	16	1.9×10^{-13}	4.8×10^{-15}
	8,000	260	10 min 58 s	14	15	5.0×10^{-13}	1.1×10^{-14}
	10,000	300	20 min 46 s	14	15	2.6×10^{-13}	5.0×10^{-15}
RFR	100	21	0.0420 s	33	34	6.1×10^{-13}	1.1×10^{-13}
	200	30	0.0840 s	26	27	8.0×10^{-13}	1.0×10^{-13}
	400	41	0.4010 s	22	23	7.6×10^{-13}	7.0×10^{-14}
	600	60	0.9690 s	21	22	8.7×10^{-13}	6.5×10^{-14}
	800	80	2.0350 s	21	22	4.0×10^{-13}	2.6×10^{-14}
	1,000	100	3.5090 s	21	22	4.0×10^{-13}	2.3×10^{-14}
	2,000	140	13.050 s	18	19	4.9×10^{-13}	2.0×10^{-14}
	4,000	180	1 min 17 s	16	17	1.8×10^{-13}	5.4×10^{-15}
	6,000	221	3 min 39 s	15	16	1.9×10^{-13}	4.8×10^{-15}
	8,000	260	7 min 39 s	14	15	5.0×10^{-13}	1.1×10^{-14}
	10,000	300	14 min 03 s	14	15	2.6×10^{-13}	5.0×10^{-15}
Alg. 2	100	21	0.0450 s	37	38	5.7×10^{-13}	9.2×10^{-14}
	200	30	0.0910 s	28	29	5.8×10^{-13}	7.3×10^{-14}
	400	41	0.4150 s	23	24	7.3×10^{-13}	6.6×10^{-14}
	600	60	0.9900 s	22	23	8.1×10^{-13}	6.1×10^{-14}
	800	80	2.1170 s	22	23	3.1×10^{-13}	2.0×10^{-14}
	1,000	100	3.7230 s	22	23	3.2×10^{-13}	1.8×10^{-14}
	2,000	140	14.476 s	19	20	2.5×10^{-13}	1.0×10^{-14}
	4,000	180	1 min 19 s	16	17	3.5×10^{-13}	1.0×10^{-14}
	6,000	221	3 min 36 s	15	16	3.6×10^{-13}	8.8×10^{-15}
	8,000	260	7 min 39 s	14	15	8.9×10^{-13}	1.9×10^{-14}
	10,000	300	14 min 01 s	14	15	4.7×10^{-13}	9.0×10^{-15}

Note. RPRP = Riemannian Polak–Ribière–Polyak; RFR = Riemannian Fletcher–Reeves.

TABLE 5 Numerical results for Example 1

		Alg. 2			IPM		
n	p	CT.	IT.	Res.	CT.	IT.	Res.
50	11	0.0310 s	30	4.1×10^{-13}	1.7760 s	22	6.6×10^{-9}
80	16	0.0370 s	31	5.0×10^{-13}	3.9490 s	24	1.3×10^{-9}
100	20	0.0340 s	27	7.0×10^{-13}	7.8180 s	26	5.9×10^{-10}
150	25	0.0530 s	25	6.7×10^{-13}	27.31 s	25	8.1×10^{-11}
200	31	0.0840 s	25	4.8×10^{-13}	1 min 7 s	25	3.2×10^{-9}
300	36	0.1280 s	21	9.0×10^{-13}	4 min 40 s	27	7.9×10^{-9}
400	41	0.3480 s	20	3.2×10^{-13}	23 min 52 s	LS failed	1.3×10^{-6}

Note. IPM = interior-point method.

Determine the parameters $\{p_i\}_{i=0}^{n-1}$ and $\{q_i\}_{i=1}^{n-1}$.

We use Algorithm 2 to Examples 3 and 4 corresponding to problem (36), where the starting point and stopping criterion are set to be

$$\hat{Z} \odot \hat{Z} = \mathtt{rand}(n,n),\ Z_0 = ((E-G) \odot \hat{Z})(\mathrm{diag}(((E-G) \odot \hat{Z})^T((E-G) \odot \hat{Z})))^{-\frac{1}{2}} \hat{I}_n^{\frac{1}{2}} \in \widehat{\mathcal{OB}_c}$$

and

$$\|\Psi^c(Z_k)\|_F \leq 10^{-12}.$$

TABLE 6 Numerical results for Example 2

		Alg. 2			IPM		
n	p	CT.	IT.	Res.	CT.	IT.	Res.
50	10	0.0530 s	36	8.0×10^{-13}	1.7850 s	22	7.9×10^{-9}
80	15	0.0470 s	35	6.1×10^{-13}	3.9480 s	24	2.3×10^{-9}
100	21	0.0460 s	37	8.8×10^{-13}	7.8780 s	25	2.2×10^{-9}
150	25	0.0620 s	32	5.1×10^{-13}	27.73 s	25	7.0×10^{-10}
200	30	0.1060 s	28	6.3×10^{-13}	1 min 10 s	26	5.7×10^{-9}
300	36	0.1800 s	24	6.1×10^{-13}	5 min 5 s	28	2.3×10^{-9}
400	40	0.4510 s	23	3.2×10^{-13}	24 min 12 s	LS failed	4.5×10^{-7}

Note. IPM = interior-point method.

The other parameters used in Algorithm 2 are set as above.

By using Algorithm 2 to Example 3 where $\lambda_1 = 1$ and

$$\mathbf{x}_1 = \begin{bmatrix} 0.6068 & 0.0969 & 0.1740 & 0.0325 & 0.0898 \end{bmatrix}^T,$$

the computed transition matrix is given by

$$\overline{C} = \begin{bmatrix} 0.6450 & 0.6323 & 0.6426 & 0.2517 & 0.3803 \\ 0.1157 & 0.1380 & 0.0306 & 0.2326 & 0.0051 \\ 0.1217 & 0.1853 & 0.2710 & 0.2809 & 0.2888 \\ 0.0210 & 0.0037 & 0.0237 & 0.0640 & 0.1468 \\ 0.0966 & 0.0407 & 0.0321 & 0.1707 & 0.1790 \end{bmatrix}.$$

By applying Algorithm 2 to Example 4, where $\lambda_1 = 1$ and $\mathbf{x}_1 = \boldsymbol{\pi}$, we get the following transition matrix:

$$\overline{C} = \begin{bmatrix} 0.7455 & 0.2182 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2545 & 0.4576 & 0.1196 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3243 & 0.4706 & 0.1966 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.4098 & 0.4204 & 0.2988 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3830 & 0.5575 & 0.1907 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1436 & 0.5943 & 0.3112 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2151 & 0.4981 & 0.3033 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.1907 & 0.4925 & 0.3042 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.2042 & 0.2370 & 0.4001 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4588 & 0.5999 \end{bmatrix}.$$

7 | CONCLUDING REMARKS

In this paper, we focus on the stochastic IEP with partial eigendata. We propose a Riemannian variant of the FR conjugate gradient algorithm for the solving unconstrained minimization problem on a Riemannian manifold and establish the global convergence of the proposed algorithm under some mild conditions. Then, we reformulate the inverse problem as a nonlinear least squares problem over a matrix oblique manifold, and the application of the proposed geometric algorithm to this problem is investigated. We also apply our geometric algorithm to the case of prescribed entries and the case of the column stochastic matrix. Numerical experiments show that our proposed geometric algorithm is very effective for solving the stochastic IEP with partial eigendata. We point out that the considered inverse problem often has many solutions. An interesting question is how to combine our geometric algorithm in finding a best approximation to a given stochastic matrix, which needs further investigation.

ACKNOWLEDGEMENTS

We are very grateful to the editor and the anonymous referees for their valuable comments on the paper, which have considerably improved this paper.

ORCID

Zheng-Jian Bai http://orcid.org/0000-0002-5134-3500

REFERENCES

1. Chu MT. Inverse eigenvalue problems. SIAM Rev. 1998;40(1):1–39.
2. Chu MT, Golub GH. Structured inverse eigenvalue problems. Acta Numer. 2002;11(1):1–71.
3. Chu MT, Golub GH. Inverse eigenvalue problems: Theory, algorithms, and applications. Oxford, UK: Oxford University Press; 2005.
4. Datta BN. Numerical methods for linear control systems: Design and analysis. New York, NY: Elsevier Academic Press; 2004.
5. Friswell MI, Mottershead JE. Finite element model updating in structural dynamics. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1995.
6. Gladwell GML. Inverse problems in vibration. 2nd ed. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2004.
7. Xu SF. An introduction to inverse algebraic eigenvalue problems. Beijing, China: Peking University Press and Vieweg Publishing; 1998.
8. Asmussen S. Applied probability and queues. 2nd ed. New York, NY: Springer; 2003.
9. Cohen JE, Kemperman JHB, Zbăganu G. Comparisons of stochastic matrices with applications in information theory, statistics, economics and population sciences. Boston, MA: Birkhäuser; 1998.
10. Ding J, Zhou AH. Nonnegative matrices, positive operators, and applications. Singapore: World Scientific Press; 2009.
11. Minc H. Nonnegative matrices. New York, NY: Wiley; 1988.
12. Petoukhov S, He M. Symmetrical analysis techniques for genetic systems and bioinformatics: Advanced patterns and applications. Hershey, PA: IGI Global; 2010.
13. Seneta E. Non-negative matrices and Markov chains. New York, NY: Springer-Verlag; 1981.
14. Karpelevič FI. On the characteristic roots of matrices with nonnegative elements. Izv Akad Nauk SSSR Ser Mat. 1951;15:361–383. in Russian.
15. Perfect H. Methods of constructing certain stochastic matrices. Duke Math J. 1953;20(3):395–404.
16. Perfect H. Methods of constructing certain stochastic matrices. II. Duke Math J. 1955;22(2):305–311.
17. Yang SJ, Xu CQ. Row stochastic inverse eigenvalue problem. J Inequalities Appl. 2011;24(1):137–141.
18. Mourad B, Abbas H, Mourad A, Ghaddar A, Kaddoura I. An algorithm for constructing doubly stochastic matrices for the inverse eigenvalue problem. Linear Algebra Appl. 2013;439(5):1382–1400.
19. Perfect H, Mirsky L. Spectral properties of doubly-stochastic matrices. Monatsh Math. 1965;69(1):35–37.
20. Soules GW. Constructing symmetric nonnegative matrices. Linear Multilinear Algebra. 1983;13(3):241–251.
21. Chu MT, Guo Q. A numerical method for the inverse stochastic spectrum problem. SIAM J Matrix Anal Appl. 1998;19(4):1027–1039.
22. Orsi R. Numerical methods for solving inverse eigenvalue problems for nonnegative matrices. SIAM J Matrix Anal Appl. 2006;28(1):190–212.
23. Lin MM. An algorithm for constructing nonnegative matrices with prescribed real eigenvalues. Appl Math Comput. 2015;256:582–590.
24. Absil PA, Baker CG, Gallivan KA. A truncated-CG style method for symmetric generalized eigenvalue problems. J Comput Appl Math. 2006;189(1–2):274–285.
25. Absil P-A, Baker CG, Gallivan KA. An implicit Riemannian trust-region method for the symmetric generalized eigenproblem. Int Conf Comput Sci. 2006;3991(42):210–217.
26. Smith ST. Optimization techniques on Riemannian manifolds. In: Hamiltonian and gradient flows, algorithms, and control. Vol. 3. Providence, RI: American Mathematical Society; 1994:113–136.
27. Yao T-T, Bai Z-J, Zhao Z, Ching W-K. A Riemannian Fletcher–Reeves conjugate gradient method for doubly stochastic inverse eigenvalue problems. SIAM J Matrix Anal Appl. 2016;37(1):215–234.
28. Zhao Z, Bai Z-J, Jin X-Q. A Riemannian Newton algorithm for nonlinear eigenvalue problems. SIAM J Matrix Anal Appl. 2015;36(2):752–774.
29. Zhao Z, Jin X-Q, Bai Z-J. A geometric nonlinear conjugate gradient method for stochastic inverse eigenvalue problems. SIAM J Numer Anal. 2016;54(4):2015–2035.
30. Wei Z, Lu S. Nonlinear optimization algorithms. Beijing, China: Science Press; 2016. in Chinese.
31. Fletcher R, Reeves C. Function minimization by conjugate gradients. Comput J. 1964;7(2):149–154.
32. Polak E, Ribière G. Note sur la convergence des méthodes de directions conjuguées. Rev Fr Imform Rech Oper. 1969;16(3):35–43.
33. Polyak BT. The conjugate gradient method in extreme problems. USSR Comput Math Math Phys. 1969;9(4):94–112.
34. Dai Y-H, Yuan Y. Nonlinear conjugate gradient methods. Shanghai, China: Shanghai Scientific & Technical Publishers; 2000. in Chinese.
35. Hager WW, Zhang HC. A survey of nonlinear conjugate gradient methods. Pacific J Optim. 2006;2(1):35–58.
36. Sun WY, Yuan YX. Optimization theory and methods: Nolinear programming. New York, NY: Springer; 2006. Springer optimization and its applications.
37. Absil P-A, Mahony R, Sepulchre R. Optimization algorithms on matrix manifolds. Princeton, NJ: Princeton University Press; 2008.
38. Sato H, Iwai T. A new, globally convergent Riemannian conjugate gradient method. Optimization. 2015;64(4):1011–1031.
39. Ring W, Wirth B. Optimization methods on Riemannian manifolds and their application to shape space. SIAM J Optim. 2012;22(2):596–627.
40. Absil P-A, Gallivan KA. Joint diagonalization on the oblique manifold for independent component analysis. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP); 2006 May 14–19; Toulouse, France. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2006. p. V-945–V-948.

41. Zhang L, Zhou WJ, Li DH. Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo-type line search. Numer Math. 2006;104(4):561–572.
42. Tütüncü RH, Toh KC, Todd MJ. Solving semidefinite-quadratic-linear programs using SDPT3. Math Program. 2003;95(2):189–217.
43. Stander J, Farrington DP, Hill G, Altham PME. Markov chain analysis and specialization in criminal careers. Br J Criminol. 1989;29(4):317–335.
44. Bapat RB, Raghavan TES. Nonnegative matrices and applications. Cambridge, UK: Cambridge University Press; 1997.
45. Van Kampen NG. Stochastic processes in physics and chemistry. Oxford, UK: Elsevier; 2007.

How to cite this article: Yao T-T, Bai Z-J, Zhao Z. A Riemannian variant of the Fletcher–Reeves conjugate gradient method for stochastic inverse eigenvalue problems with partial eigendata. *Numer Linear Algebra Appl.* 2018;e2221. https://doi.org/10.1002/nla.2221