SIAM J. NUMER. ANAL.
Vol. 54, No. 4, pp. 2015–2035

A GEOMETRIC NONLINEAR CONJUGATE GRADIENT METHOD FOR STOCHASTIC INVERSE EIGENVALUE PROBLEMS*

ZHI ZHAO†, XIAO-QING JIN‡, AND ZHENG-JIAN BAI§

Abstract. In this paper, we focus on the stochastic inverse eigenvalue problem of reconstructing a stochastic matrix from the prescribed spectrum. We directly reformulate the stochastic inverse eigenvalue problem as a constrained optimization problem over several matrix manifolds to minimize the distance between isospectral matrices and stochastic matrices. Then we propose a geometric Polak–Ribière–Polyak-based nonlinear conjugate gradient method for solving the constrained optimization problem. The global convergence of the proposed method is established. Our method can also be extended to the stochastic inverse eigenvalue problem with prescribed entries. An extra advantage is that our models yield new isospectral flow methods. Finally, we report some numerical tests to illustrate the efficiency of the proposed method for solving the stochastic inverse eigenvalue problem and the case of prescribed entries.

Key words. inverse eigenvalue problem, stochastic matrix, geometric nonlinear conjugate gradient method, oblique manifold, isospectral flow method

AMS subject classifications. 65F18, 65F15, 15A18, 65K05, 90C26, 90C48

DOI. 10.1137/140992576

1. Introduction. An n-by-n real matrix A is called a nonnegative matrix if all its entries are nonnegative, i.e., $A_{ij} \geq 0$ for all $i, j = 1, \ldots, n$, where A_{ij} means the (i, j)th entry of A. An n-by-n real matrix A is called a (row) stochastic matrix if it is nonnegative and the sum of the entries in each row equals one, i.e., $A_{ij} \geq 0$ and $\sum_{j=1}^{n} A_{ij} = 1$ for all $i, j = 1, \ldots, n$. Stochastic matrices arise in various applications such as probability theory, statistics, economics, computer science, physics, chemistry, and population genetics. See, for instance, [6, 13, 15, 20, 25, 32, 35] and the references therein.

In this paper, we consider the stochastic inverse eigenvalue problem (StIEP) and the stochastic inverse eigenvalue problem with prescribed entries (StIEP-PE) as follows:

StIEP. *Given a self-conjugate set of complex numbers* $\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*\}$, *find an* n*-by-*n *stochastic matrix* C *such that its eigenvalues are* $\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*$.

StIEP-PE. *Given a self-conjugate set of complex numbers* $\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*\}$, *an index subset* $\mathcal{L} \subset \mathcal{N} := \{(i, j) \mid i, j = 1, \ldots, n\}$, *and a certain* n*-by-*n *nonnegative matrix* $\widehat{C}_a$ *with* $(\widehat{C}_a)_{ij} = 0$ *for* $(i, j) \notin \mathcal{L}$ *(i.e., the set of nonnegative numbers* $\{(\widehat{C_a})_{ij} \mid (i, j) \in \mathcal{L}\}$ *is prescribed), find an* n*-by-*n *stochastic matrix* C *such that its*

*Received by the editors October 22, 2014; accepted for publication (in revised form) March 31, 2016; published electronically July 6, 2016.
http://www.siam.org/journals/sinum/54-4/99257.html
†Department of Mathematics, School of Sciences, Hangzhou Dianzi University, Hangzhou 310018, People's Republic of China (zzhao@hdu.edu.cn).
‡Department of Mathematics, University of Macau, Taipa, Macau, People's Republic of China (xqjin@umac.mo). The research of this author was supported by the research grant MYRG098(Y3-L3)-FST13-JXQ from University of Macau.
§Corresponding author. School of Mathematical Sciences, Xiamen University, Xiamen 361005, People's Republic of China (zjbai@xmu.edu.cn). The research of this author was partially supported by the National Natural Science Foundation of China (11271308), the Natural Science Foundation of Fujian Province of China (2016J01035), and the Fundamental Research Funds for the Central Universities (20720150001).

eigenvalues are $\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*$ *and*

$$C_{ij} = (\widehat{C}_a)_{ij} \quad \forall\, (i,j) \in \mathcal{L}.$$

As stated in [11], the StIEP aims to construct a Markov chain (i.e., a stochastic matrix) from the given spectral property. The StIEP is a special nonnegative inverse eigenvalue problem which aims to find a nonnegative matrix from the given spectrum. These two problems are related to each other by the following result [19, p. 142].

THEOREM 1.1. *If* A *is a nonnegative matrix with positive maximal eigenvalue* r *and a positive maximal eigenvector* $\mathbf{a}$*, then* $r^{-1}D^{-1}AD$ *is a stochastic matrix, where* $D := \mathrm{Diag}(\mathbf{a})$ *means a diagonal matrix with the vector* $\mathbf{a}$ *on its diagonal.*

Theorem 1.1 shows that one may first construct a nonnegative matrix with the prescribed spectrum, and then obtain a stochastic matrix with the desired spectrum by a diagonal similarity transformation. In many applications (see, for instance, [6, 32, 35]), the underlying structure of a desired stochastic matrix is often characterized by the prescribed entries. The corresponding inverse problem can be described as the StIEP-PE. However, the StIEP-PE may not be solved by first constructing a nonnegative solution followed by a diagonal similarity transformation.

There is a large literature on the existence condition of solutions to the StIEP in many special cases. The reader may refer to [8, 10, 11, 17, 19, 23, 24, 36, 37] for various necessary or sufficient conditions. However, there exist few numerical methods for solving the StIEP, especially for the large-scale StIEP (say $n \geq 1000$). In [34], Soules gave a constructive method where additional inequalities of the Perron eigenvector are required. In [12], based on a matrix similarity transformation, Chu and Guo presented an isospectral flow method for constructing a stochastic matrix with the desired spectrum. An inverse matrix involved in the isospectral flow method was further characterized by an analytic singular value decomposition (ASVD). Recently, in [21] Orsi proposed an alternating projection-like scheme for the construction of a stochastic matrix with the prescribed spectrum. The main computational cost is an eigenvalue-eigenvector decomposition or a Schur matrix decomposition. In [18], Lin presented a fast recursive algorithm for solving the StIEP, where the prescribed eigenvalues are assumed to be real and satisfy additional inequality conditions. We see that all the numerical methods in [12, 18, 21] are based on Theorem 1.1 and may not be extended to the StIEP-PE.

Recently, there appeared a few Riemannian optimization methods for eigenproblems. For instance, in [1], a truncated conjugate gradient method was presented for the symmetric generalized eigenvalue problem. In [5], a Riemannian trust-region method was proposed for the symmetric generalized eigenproblem. Newton's method and the conjugate gradient method proposed in [33] are applicable to the symmetric eigenvalue problem. In [40], a Riemannian Newton method was proposed for a kind of nonlinear eigenvalue problem.

In this paper, we directly focus on the StIEP. Based on the real Schur matrix decomposition, we reformulate the StIEP as a constrained optimization problem over several matrix manifolds, which aims to minimize the distance between isospectral matrices and stochastic matrices. We propose a geometric Polak–Ribière–Polyak-based nonlinear conjugate gradient method for solving the constrained optimization problem. This is motivated by Yao et al. [38] and Zhang, Zhou, and Li [39]. In [38], a Riemannian Fletcher–Reeves conjugate gradient approach was proposed for solving the doubly stochastic inverse eigenvalue problem. In [39], Zhang, Zhou, and Li gave a modified Polak–Ribière–Polyak conjugate gradient method for solving the unconstrained optimization problem, where the generated direction is always a descent direction for the objective function. We investigate some basic properties of these matrix manifolds (e.g., tangent spaces, Riemannian metrics, orthogonal projections, retractions, and vector transports) and derive an explicit expression of the Riemannian gradient of the cost function. The global convergence of the proposed method is established. We also extend the proposed method to the StIEP-PE. An additional advantage of our models is that they also yield new geometric isospectral flow methods for both the StIEP and the StIEP-PE. Finally, we report some numerical tests to show that the proposed geometric methods perform very effectively and are applicable to large-scale problems. Moreover, the new geometric isospectral flow methods work acceptably for small- and medium-scale problems.

The rest of this paper is organized as follows. In section 2 we propose a geometric nonlinear conjugate gradient approach for solving the StIEP. In section 3 we establish the global convergence of the proposed approach. In section 4 we extend the proposed approach to the StIEP-PE. Finally, some numerical tests are reported in section 5, and some concluding remarks are given in section 6.

2. A geometric nonlinear conjugate gradient method.

In this section, we propose a geometric nonlinear conjugate gradient approach for solving the StIEP. We first reformulate the StIEP as a nonlinear minimization problem defined on some matrix manifolds. We also discuss some basic properties of these matrix manifolds and derive the Riemannian gradient of the cost function. Then we present a geometric nonlinear conjugate gradient approach for solving the nonlinear minimization problem.

2.1. Reformulation.

As noted in [19, p. 141], we only assume that the set of prescribed complex numbers $\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*\}$ satisfies the following necessary condition: One element is 1 and $\max_{1\le i\le n} |\lambda_i^*| = 1$. Note that the set of complex numbers $\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*\}$ is closed under complex conjugation. Without loss of generality, we can assume

$$\lambda_{2j-1}^* = a_j + b_j\sqrt{-1},\ \lambda_{2j}^* = a_j - b_j\sqrt{-1},\quad j = 1,\ldots,s;\quad \lambda_j \in \mathbb{R},\quad j = 2s+1,\ldots,n,$$

where $a_j, b_j \in \mathbb{R}$ with $b_j \neq 0$ for $j = 1,\ldots,s$. Then we define a block diagonal matrix by

$$\Lambda := \mathrm{blkdiag}\big(\lambda_1^{[2]*}, \ldots, \lambda_s^{[2]*}, \lambda_{2s+1}^*, \ldots, \lambda_n^*\big),$$

with diagonal blocks $\lambda_1^{[2]*}, \ldots, \lambda_s^{[2]*}, \lambda_{2s+1}^*, \ldots, \lambda_n^*$, where

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

Denote by $\mathcal{O}(n)$ the set of all n-by-n orthogonal matrices, i.e.,

$$\mathcal{O}(n) := \big\{P \in \mathbb{R}^{n\times n} \mid P^T P = I_n\big\},$$

where P^T means the transpose of P and I_n is the identity matrix of order n. In addition, we define the set $\mathcal{V}$ by

$$\mathcal{V} := \big\{V \in \mathbb{R}^{n\times n} \mid V_{ij} = 0,\ (i,j) \in \mathcal{I}\big\},$$

where $\mathcal{I}$ is the index subset,

$$\mathcal{I} := \big\{(i,j) \mid i \ge j \text{ or } \Lambda_{ij} \neq 0,\ i,j = 1,\ldots,n\big\}.$$

Let $\mathcal{S}$ be the set of all n-by-n stochastic matrices, which can be defined as the following matrix manifold:

$$\mathcal{S} := \{S \odot S \mid \operatorname{diag}(SS^T) = I_n,\ S \in \mathbb{R}^{n\times n}\},$$

where $\odot$ denotes the Hadamard product [7] and $\operatorname{diag}(A)$ is a diagonal matrix with the same diagonal entries as a square matrix A. The set of isospectral matrices defined by

$$\mathcal{M}(\Lambda) := \left\{Z \in \mathbb{R}^{n\times n} \mid Z = P(\Lambda + V)P^T,\ P \in \mathcal{O}(n), V \in \mathcal{V}\right\} \tag{2.1}$$

is a smooth manifold. Then the StIEP has a solution if and only if $\mathcal{M}(\Lambda) \cap \mathcal{S} \neq \emptyset$.

Suppose that the StIEP has at least one solution. Then the StIEP is equivalent to solving the following nonlinear matrix equation:

$$H(S, P, V) = \mathbf{0}_{n\times n}$$

for $(S, P, V) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$, where $\mathbf{0}_{n\times n}$ is the zero matrix of order n and the set $\mathcal{OB}$ is the oblique manifold defined by [3, 4],

$$\mathcal{OB} := \left\{S \in \mathbb{R}^{n\times n} \mid \operatorname{diag}(SS^T) = I_n\right\}.$$

The mapping $H : \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V} \to \mathbb{R}^{n\times n}$ is defined by

$$H(S, P, V) = S \odot S - P(\Lambda + V)P^T, \quad (S, P, V) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}. \tag{2.2}$$

If we find a solution $(\overline{S}, \overline{P}, \overline{V}) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ to $H(S, P, V) = \mathbf{0}_{n\times n}$, then $\overline{C} = \overline{S} \odot \overline{S}$ is a solution to the StIEP. Alternatively, one may solve the StIEP by finding a global solution to the following nonlinear minimization problem:

$$\begin{array}{ll} \min & h(S, P, V) := \frac{1}{2}\|H(S, P, V)\|_F^2 = \frac{1}{2}\|S \odot S - P(\Lambda + V)P^T\|_F^2 \\ \text{s.t.} & S \in \mathcal{OB}, \quad P \in \mathcal{O}(n), \quad V \in \mathcal{V}, \end{array} \tag{2.3}$$

where "s.t." means "subject to" and $\|\cdot\|_F$ means the matrix Frobenius norm [16].

2.2. Basic properties. In the following, we establish some basic properties for the product manifold $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$. It is easy to check that the dimensions of $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{V}$ are given by [4, p. 27]

$$\dim \mathcal{OB} = n(n-1), \quad \dim \mathcal{O}(n) = \frac{1}{2}n(n-1), \quad \dim \mathcal{V} = |\mathcal{J}|,$$

where $|\mathcal{J}|$ is the cardinality of the index subset $\mathcal{J} = \mathcal{N} \setminus \mathcal{I}$, which is the relative complement of $\mathcal{I}$ in $\mathcal{N}$. Then we have

$$\dim(\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}) = n(n-1) + \frac{n(n-1)}{2} + |\mathcal{J}| > \dim \mathbb{R}^{n\times n}, \quad n \geq 3.$$

This shows that the nonlinear equation $H(S, P, V) = \mathbf{0}_{n\times n}$ is an under-determined system defined from $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ to $\mathbb{R}^{n\times n}$ for $n \geq 3$.

The tangent spaces of $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{V}$ at $S \in \mathcal{OB}$, $P \in \mathcal{O}(n)$, and $V \in \mathcal{V}$ are, respectively, given by (see [3] and [4, p. 42])

$$\left\{\begin{array}{lcl} T_S\mathcal{OB} & = & \left\{T \in \mathbb{R}^{n\times n} \mid \operatorname{diag}(ST^T) = \mathbf{0}_{n\times n}\right\}, \\ T_P\mathcal{O}(n) & = & \left\{P\Omega \mid \Omega^T = -\Omega,\ \Omega \in \mathbb{R}^{n\times n}\right\}, \\ T_V\mathcal{V} & = & \mathcal{V}. \end{array}\right.$$

Hence, the tangent space of $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ at $(S,P,V)\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ is given by

$$T_{(S,P,V)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}=T_S\mathcal{OB}\times T_P\mathcal{O}(n)\times T_V\mathcal{V}.$$

For the isospectral manifold $\mathcal{M}(\Lambda)$ defined in (2.1), it follows that

$$T_Z\mathcal{M}(\Lambda)=\big\{[Z,\Omega]+PUP^T\mid \Omega^T=-\Omega,\ U\in\mathcal{V},\ \Omega\in\mathbb{R}^{n\times n}\big\},$$

where $Z=P(\Lambda+V)P^T\in\mathcal{M}(\Lambda)$ and $[Z,\Omega]:=Z\Omega-\Omega Z$ means the Lie-bracket product.

Let the Euclidean space $\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}$ be equipped with the following natural inner product:

$$\langle (X_1,Y_1,Z_1),(X_2,Y_2,Z_2)\rangle:=\mathrm{tr}(X_1^TX_2)+\mathrm{tr}(Y_1^TY_2)+\mathrm{tr}(Z_1^TZ_2)$$

for all $(X_1,Y_1,Z_1),(X_2,Y_2,Z_2)\in\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}$ and its induced norm $\|\cdot\|$, where $\mathrm{tr}(A)$ denotes the sum of the diagonal entries of a square matrix A. Since $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ is an embedded submanifold of $\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}$, we can equip $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ with an induced Riemannian metric,

$$g_{(S,P,V)}\big((\xi_1,\zeta_1,\eta_1),(\xi_2,\zeta_2,\eta_2)\big):=\langle(\xi_1,\zeta_1,\eta_1),(\xi_2,\zeta_2,\eta_2)\rangle \tag{2.4}$$

for all $(S,P,V)\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ and $(\xi_1,\zeta_1,\eta_1),(\xi_2,\zeta_2,\eta_2)\in T_{(S,P,V)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$. Without causing any confusion, in what follows we use $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ to denote the Riemannian metric on $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ and its induced norm. Then, the orthogonal projections of any $\xi,\zeta,\eta\in\mathbb{R}^{n\times n}$ onto $T_S\mathcal{OB}$, $T_P\mathcal{O}(n)$, and $T_V\mathcal{V}$ are, respectively, given by (see [3] and [4, p. 48])

$$\Pi_S\xi=\xi-\mathrm{diag}(S\xi^T)S,\quad \Pi_P\zeta=P\mathrm{skew}(P^T\zeta),\quad \Pi_V\eta=W\odot\eta, \tag{2.5}$$

where $\mathrm{skew}(A):=\frac{1}{2}(A-A^T)$ for a matrix $A\in\mathbb{R}^{n\times n}$ and the matrix $W\in\mathbb{R}^{n\times n}$ is defined by

$$W_{ij}:=\begin{cases}0 & \text{if } (i,j)\in\mathcal{I},\\ 1 & \text{otherwise.}\end{cases}$$

Therefore, the orthogonal projection of a point $(\xi,\zeta,\eta)\in\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}$ onto $T_{(S,P,V)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ has the following form:

$$\Pi_{(S,P,V)}(\xi,\zeta,\eta)=\big(\xi-\mathrm{diag}(S\xi^T)S,P\mathrm{skew}(P^T\zeta),W\odot\eta\big).$$

In addition, the retractions on the manifolds $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{V}$ can be chosen as (see [3] and [4, p. 58]),

$$\begin{cases} R_S(\xi_S) &=& \big(\mathrm{diag}\left((S+\xi_S)(S+\xi_S)^T\right)\big)^{-1/2}(S+\xi_S)\quad \text{for } \xi_S\in T_S\mathcal{OB},\\ R_P(\zeta_P) &=& \mathrm{qf}(P+\zeta_P)\quad \text{for } \zeta_P\in T_P\mathcal{O}(n),\\ R_V(\eta_V) &=& V+\eta_V\quad \text{for } \eta_V\in T_V\mathcal{V},\end{cases}$$

where $\mathrm{qf}(A)$ denotes the Q factor of the QR decomposition of a nonsingular matrix $A\in\mathbb{R}^{n\times n}$ in the form of $A=Q\widetilde{R}$. Here, $Q\in\mathcal{O}(n)$ and $\widetilde{R}$ is an upper triangular matrix with strictly positive diagonal elements. Thus, a retraction R on $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ takes the form of

$$R_{(S,P,V)}(\xi_S,\zeta_P,\eta_V)=\big(R_S(\xi_S),R_P(\zeta_P),R_V(\eta_V)\big)$$

for all $(\xi_S, \zeta_P, \gamma_V) \in T_{(S,P,V)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$.

Since all the manifolds $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{V}$ are embedded Riemannian submanifolds of $\mathbb{R}^{n\times n}$, we can define vector transports on $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{V}$ by orthogonal projections [4, p. 174]:

(2.6)
$$\begin{cases} \mathcal{T}_{\eta_S}\xi_S & := & \Pi_{R_S(\eta_S)}\xi_S = \xi_S - \operatorname{diag}\left(R_S(\eta_S)\xi_S^T\right)R_S(\eta_S) \quad \text{for } \xi_S, \eta_S \in T_S\mathcal{OB}, \\ \mathcal{T}_{\eta_P}\xi_P & := & \Pi_{R_P(\eta_P)}\xi_P = R_P(\eta_P)\text{skew}\Big(\big(R_P(\eta_P)\big)^T\xi_P\Big) \quad \text{for } \xi_P, \eta_P \in T_P\mathcal{O}(n), \\ \mathcal{T}_{\eta_V}\xi_V & := & \xi_V \quad \text{for } \xi_V, \eta_V \in T_V\mathcal{V}. \end{cases}$$

Thus the vector transport on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ has the following form:

$$\mathcal{T}_{(\eta_S,\eta_P,\eta_V)}(\xi_S, \xi_P, \xi_V) = \left(\mathcal{T}_{\eta_S}\xi_S, \mathcal{T}_{\eta_P}\xi_P, \mathcal{T}_{\eta_V}\xi_V\right) \tag{2.7}$$

for any $(\xi_S, \xi_P, \xi_V), (\eta_S, \eta_P, \eta_V) \in T_{(S,P,V)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$.

We now establish explicit formulas for the differential of H defined in (2.2) and the Riemannian gradient of the cost function h defined in (2.3). To derive the Riemannian gradient of h, we define the mapping $\overline{H} : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and the cost function $\overline{h} : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}$ by

$$\overline{H}(S, P, V) := S \odot S - P(\Lambda + V)P^T \quad \text{and} \quad \overline{h}(S, P, V) := \frac{1}{2}\|\overline{H}(S, P, V)\|_F^2 \tag{2.8}$$

for all $(S, P, V) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. Then the mapping H and the function h can be seen as the restrictions of $\overline{H}$ and $\overline{h}$ onto $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$, respectively. That is, $H = \overline{H}|_{\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}}$ and $h = \overline{h}|_{\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}}$. The Euclidean gradient of $\overline{h}$ at a point $(S, P, V) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ is given by

$$\operatorname{grad} \overline{h}(S, P, V) = \left(\frac{\partial}{\partial S}\overline{h}(S, P, V), \frac{\partial}{\partial P}\overline{h}(S, P, V), \frac{\partial}{\partial V}\overline{h}(S, P, V)\right),$$

where

$$\begin{cases} \frac{\partial}{\partial S}\overline{h}(S, P, V) & = & 2S \odot \overline{H}(S, P, V), \\ \frac{\partial}{\partial P}\overline{h}(S, P, V) & = & -\overline{H}(S, P, V)P(\Lambda + V)^T - \overline{H}^T(S, P, V)P(\Lambda + V), \\ \frac{\partial}{\partial V}\overline{h}(S, P, V) & = & -P^T\overline{H}(S, P, V)P. \end{cases}$$

Then the Riemannian gradient of h at a point $(S, P, V) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ is given by [4, p. 48]

$$\begin{aligned} \operatorname{grad} h(S, P, V) &= \Pi_{(S,P,V)} \operatorname{grad} \overline{h}(S, P, V) \\ &= \left(\Pi_S \frac{\partial}{\partial S}\overline{h}(S, P, V), \Pi_P \frac{\partial}{\partial P}\overline{h}(S, P, V), \Pi_V \frac{\partial}{\partial V}\overline{h}(S, P, V)\right). \end{aligned} \tag{2.9}$$

It follows from (2.5) that

$$\begin{cases} \Pi_S \frac{\partial}{\partial S}\overline{h}(S, P, V) = 2S \odot H(S, P, V) - 2\operatorname{diag}\left(S(S \odot H(S, P, V))^T\right)S, \\ \Pi_P \frac{\partial}{\partial P}\overline{h}(S, P, V) = \dfrac{1}{2}\big([P(\Lambda + V)P^T, H^T(S, P, V)] + [P(\Lambda + V)^TP^T, H(S, P, V)]\big)P, \\ \Pi_V \frac{\partial}{\partial V}\overline{h}(S, P, V) = -W \odot \left(P^TH(S, P, V)P\right). \end{cases}$$

By using a similar idea from [9] and [12], we can define the steepest descent flow for the cost function h over $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ as follows:

$$\frac{d(S,P,V)}{dt} = -\operatorname{grad} h(S,P,V). \tag{2.10}$$

This, together with an initial value $(S(0),P(0),V(0))\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$, leads to an initial value problem for problem (2.3). Thus we can apply existing ODE solvers to the above differential equation and get a solution for problem (2.3). From the numerical tests in section 5, one can see that this extra geometric isospectral flow method is acceptable for small- and medium-scale problems.

On the other hand, the differential $\mathrm{D}H(S,P,V): T_{(S,P,V)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}\to T_{H(S,P,V)}\mathbb{R}^{n\times n}\simeq\mathbb{R}^{n\times n}$ of H at a point $(S,P,V)\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ is determined by

$$\mathrm{D}H(S,P,V)[(\Delta S,\Delta P,\Delta V)] = 2S\odot\Delta S + [P(\Lambda+V)P^T,\Delta PP^T] - P\Delta VP^T$$

for all $(\Delta S,\Delta P,\Delta V)\in T_{(S,P,V)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$. Then we have [4, p. 185]

$$\operatorname{grad} h(S,P,V) = (\mathrm{D}H(S,P,V))^*[H(S,P,V)], \tag{2.11}$$

where $(\mathrm{D}H(S,P,V))^*$ denotes the adjoint of the operator $\mathrm{D}H(S,P,V)$.

Let $T\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ denote the tangent bundle of $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$. As in [4, p. 55], the pullback mappings $\widehat{H}: T\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}\to\mathbb{R}^{n\times n}$ and $\widehat{h}: T\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}\to\mathbb{R}$ of H and h through the retraction R on $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$ can be defined by

$$\widehat{H}(\xi) = H(R(\xi)) \quad\text{and}\quad \widehat{h}(\xi) = h(R(\xi)),\quad \xi\in T\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}, \tag{2.12}$$

where $\widehat{H}_{(S,P,V)}$ and $\widehat{h}_{(S,P,V)}$ are the restrictions of $\widehat{H}$ and $\widehat{h}$ to $T_{(S,P,V)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$, respectively. By using the local rigidity of R [4, p. 55], one has

$$\mathrm{D}H(X) = \mathrm{D}\widehat{H}_X(\mathbf{0}_X),\quad X\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}, \tag{2.13}$$

where $\mathbf{0}_X$ is the origin of $T_X\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$. For the Riemannian gradient of h and the Euclidean gradient of its pullback function $\widehat{h}$, we have [4, p. 56]

$$\operatorname{grad} h(S,P,V) = \operatorname{grad}\widehat{h}_{(S,P,V)}\big(\mathbf{0}_{(S,P,V)}\big). \tag{2.14}$$

2.3. A geometric nonlinear conjugate gradient method. In this section, we propose a geometric modified Polak–Ribière–Polyak conjugate gradient method for solving the StIEP. This is motivated by the modified Polak–Ribière–Polyak conjugate gradient method in [39], where the generated direction is always a descent direction for the objective function and the modified method with Armijo-type line search is globally convergent. For the original Polak–Ribière–Polyak conjugate gradient method, the reader may refer to [26, 27]. A geometric Polak–Ribière–Polyak-based nonlinear conjugate gradient algorithm for solving problem (2.3) can be stated as follows.

ALGORITHM 2.1 (a geometric nonlinear conjugate gradient method).
Step 0. *Given* $X^0 := (S^0,P^0,V^0)\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{V}$, $\alpha\geq 1$, $\rho,\delta\in(0,1)$. $k:=0$.
Step 1. *If* $h(X^k)=0$, *then stop. Otherwise, go to Step* 2. *Here* $X^k := (S^k,P^k,V^k)$.
Step 2. *Set*

$$\Delta X_k = -\operatorname{grad} h(X^k) + \beta_k\mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}}\Delta X_{k-1} - \theta_k Y^{k-1}, \tag{2.15}$$

where $\Delta X_k := (\Delta S_k, \Delta P_k, \Delta V_k)$ *and*

$$Y^{k-1} := \operatorname{grad} h(X^k) - \mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}} \operatorname{grad} h(X^{k-1}), \tag{2.16}$$

$$\beta_k = \frac{\langle \operatorname{grad} h(X^k), Y^{k-1}\rangle}{\|\operatorname{grad} h(X^{k-1})\|^2}, \tag{2.17}$$

$$\theta_k = \frac{\langle \operatorname{grad} h(X^k), \mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}} \Delta X_{k-1}\rangle}{\|\operatorname{grad} h(X^{k-1})\|^2}. \tag{2.18}$$

Step 3. *Determine* $\alpha_k = \max\{\alpha\rho^j, j = 0, 1, 2, \ldots\}$ *such that*

$$h\big(R_{X^k}(\alpha_k \Delta X_k)\big) \le h(X^k) - \delta\alpha_k^2 \|\Delta X_k\|^2. \tag{2.19}$$

Set

$$X^{k+1} := R_{X^k}(\alpha_k \Delta X_k). \tag{2.20}$$

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

We have several remarks for the above algorithm as follows:

- In Algorithm 2.1, we set
$$\Delta X_0 = -\operatorname{grad} h(X^0).$$
- From (2.15), (2.16), (2.17), and (2.18), we get for all $k \ge 1$,
$$\begin{aligned}\langle \Delta X_k, \operatorname{grad} h(X^k)\rangle &= -\langle \operatorname{grad} h(X^k), \operatorname{grad} h(X^k)\rangle - \theta_k \langle Y^{k-1}, \operatorname{grad} h(X^k)\rangle \\ &\quad + \beta_k \langle \mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}} \Delta X_{k-1}, \operatorname{grad} h(X^k)\rangle \\ &= -\|\operatorname{grad} h(X^k)\|^2. \end{aligned} \tag{2.21}$$
This shows that the search direction ΔX_k is a descent direction of h.
- According to (2.6), the vector transport in (2.15) is given by
$$\begin{aligned}&\mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}} \Delta X_{k-1} \\ &\quad = \Big(\Delta S_{k-1} - \operatorname{diag}\big(S^k(\Delta S_{k-1})^T\big)S^k, P^k \operatorname{skew}\big((P^k)^T \Delta P_{k-1}\big), \Delta V_{k-1}\Big).\end{aligned}$$
- Since the function h is bounded from below, it follows from (2.19) and (2.20) that the sequence $\{h(X^k)\}$ is monotone decreasing and thus is convergent. Hence, we have $\sum_{k=0}^{\infty} \alpha_k^2 \|\Delta X_k\|^2 < \infty$. Then it follows that
$$\lim_{k\to\infty} \alpha_k \|\Delta X_k\| = 0. \tag{2.22}$$
- In Step 3 of Algorithm 2.1, the initial step length is set to be α. As mentioned in [39], the line-search process may not be very efficient. In the following, we propose a good initial guess to the initial step length. For the nonlinear mapping $\widehat{H}_{X^k}$ defined between the tangent space $T_{X^k}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ and $T_{H(X^k)}\mathbb{R}^{n\times n}$, we consider the first-order approximation of $\widehat{H}_{X^k}$ at the origin $\mathbf{0}_{X^k}$ along the search direction ΔX_k:
$$\widehat{H}_{X^k}\big(t\Delta X_k\big) \approx \widehat{H}_{X^k}(\mathbf{0}_{X^k}) + t\mathrm{D}\widehat{H}_{X^k}(\mathbf{0}_{X^k})[\Delta X_k] \tag{2.23}$$

for all t close to 0. Based on (2.12) and (2.13), this is reduced to

$$H\big(R_{X^k}(t\Delta X_k)\big) \approx H(X^k) + t\mathrm{D}H\big(X^k\big)[\Delta X_k]$$

for all t close to 0. We have by (2.12) and (2.23),

$$\begin{aligned} \widehat{h}_{X^k}\big(t\Delta X_k\big) &= \frac{1}{2}\|\widehat{H}_{X^k}\big(t\Delta X_k\big)\|^2 \\ &\approx \frac{1}{2}\|\widehat{H}_{X^k}(\mathbf{0}_{X^k})\|^2 + t\big\langle\big(\mathrm{D}\widehat{H}_{X^k}(\mathbf{0}_{X^k})\big)^*\big[\widehat{H}_{X^k}(\mathbf{0}_{X^k})\big], \Delta X_k\big\rangle \\ &\quad + \frac{1}{2}t^2\|\mathrm{D}\widehat{H}_{X^k}(\mathbf{0}_{X^k})[\Delta X_k]\|^2 \end{aligned}$$

for all t close to 0. Based on (2.12) and (2.13), this becomes

(2.24)
$$h\big(R_{X^k}(t\Delta X_k)\big) \approx h(X^k) + t\langle(\mathrm{D}H(X^k))^*[H(X^k)], \Delta X_k\rangle + \frac{1}{2}t^2\|\mathrm{D}H(X^k)[\Delta X_k]\|^2$$

for all t close to 0. By (2.11), we have the following local approximation of h at X^k along the direction ΔX_k:

$$m_k(t) := h(X^k) + t\langle \mathrm{grad}\, h(X^k), \Delta X_k\rangle + \frac{1}{2}t^2\|\mathrm{D}H(X^k)[\Delta X_k]\|^2$$

for all t close to 0. Let

$$m_k^{'}(t) = \langle \mathrm{grad}\, h(X^k), \Delta X_k\rangle + t\|\mathrm{D}H(X^k)[\Delta X_k]\|^2 = 0.$$

We have

$$t = -\frac{\langle \mathrm{grad}\, h(X^k), \Delta X_k\rangle}{\|\mathrm{D}H(X^k)[\Delta X_k]\|^2}.$$

Hence, a reasonable initial step length can be set to

$$t_k = \frac{\big|\langle \mathrm{grad}\, h(X^k), \Delta X_k\rangle\big|}{\|\mathrm{D}H(X^k)[\Delta X_k]\|^2}. \tag{2.25}$$

Motivated by [39], the line-search step, i.e., Step 3 of Algorithm 2.1, can be modified such that if

$$\|\mathrm{D}H(X^k)[\Delta X_k]\| > 0 \quad \text{and} \quad h\big(R_{X^k}(t_k\Delta X_k)\big) \le h(X^k) - \delta t_k^2\|\Delta X_k\|^2,$$

then we set $\alpha_k = t_k$; otherwise, the step length α_k can be selected by Step 3 of Algorithm 2.1. The numerical tests in section 5 show that the initial step length (2.25) is very effective.

Finally, we point out that there also exist other nonlinear conjugate gradient methods, which were generalized to Riemannian manifolds (see, for instance, [14, 29, 30, 31]). The global convergence of these methods is guaranteed under the strong or weak Wolfe conditions and specially constructed vector transports defined by parallel translation or scaled differentiated retraction. In general, parallel translation or scaled differentiated retraction is computationally expensive. For the StIEP, we note that $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ is an embedded submanifold of $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$; it is natural to define the vector transport by orthogonal projection. As shown in section 3, the proposed geometric nonlinear conjugate gradient method with Armijo-type line search is globally convergent. From the numerical tests in section 5, one can see that Algorithm 2.1 is easy to implement and computationally effective for solving the StIEP.

3. Global convergence. In this section, we establish the global convergence of Algorithm 2.1. We first derive the following result.

LEMMA 3.1. *For any point* $X^0 := (S^0, P^0, V^0) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$, *the level set*

$$\mathcal{Z} := \left\{ X := (S, P, V) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V} \mid h(X) \le h(X^0) \right\} \tag{3.1}$$

is compact.

Proof. For any point $(S, P, V) \in \mathcal{Z}$, we have

$$\|S\|_F \le \sqrt{n}, \quad \|S \odot S\|_F \le \sqrt{n}, \quad \|P\|_F = \sqrt{n}\ , \tag{3.2}$$

and

$$\|S \odot S - P(\Lambda + V)P^T\|_F \le \|S^0 \odot S^0 - P^0(\Lambda + V^0)(P^0)^T\|_F = \sqrt{2h(X^0)}.$$

Note that the Frobenius norm is invariant under orthogonal transformation. It follows from (3.2) that

$$\begin{aligned} \|V\|_F = \|PVP^T\|_F &\le \sqrt{2h(X^0)} + \|P\Lambda P^T\|_F + \|S \odot S\|_F \\ &\le \sqrt{2h(X^0)} + \|\Lambda\|_F + \sqrt{n}. \end{aligned}$$

This, together with (3.2), shows that the level set $\mathcal{Z}$ is bounded. In addition, since h is a continuous function, $\mathcal{Z}$ is also closed. Thus the level set $\mathcal{Z}$ is compact. ∎

On the relation between the sequences $\{\|Y^{k-1}\|\}$ and $\{\|\Delta X_{k-1}\|\}$ defined in (2.15)–(2.16), we have the following result.

PROPOSITION 3.2. *Let* $\{Y^{k-1}\}$ *and* $\{\Delta X_{k-1}\}$ *be the sequences generated by Algorithm* 2.1*; then there exists a constant* $\nu > 0$ *such that for all* k *sufficiently large,*

$$\|Y^{k-1}\| \le \nu \alpha_{k-1} \|\Delta X_{k-1}\|,$$

where α_{k-1} *is the stepsize defined in Algorithm* 2.1.

Proof. By (2.7), (2.16), and (2.20), we get

$$\begin{aligned} Y^{k-1} &= \operatorname{grad} h(X^k) - \mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}} \operatorname{grad} h(X^{k-1}) \\ &= \operatorname{grad} h(X^k) - \Pi_{R_{X^{k-1}}(\alpha_{k-1}\Delta X_{k-1})} \operatorname{grad} h(X^{k-1}) \\ &= \operatorname{grad} h(X^k) - \Pi_{X^k} \operatorname{grad} h(X^{k-1}) \\ &= \Pi_{X^k}\big(\operatorname{grad} h(X^k) - \operatorname{grad} h(X^{k-1})\big). \end{aligned} \tag{3.3}$$

Since $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ is an embedded Riemannian submanifold of $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$, one may rely on the natural inclusion $T_X\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V} \subset \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ for all $X \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$. Thus, the Riemannian gradient $\operatorname{grad} h$ defined in (2.9) can be seen as a continuous mapping between $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ and $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. For any points $X_1, X_2 \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$, the operation $\operatorname{grad} h(X_2) - \operatorname{grad} h(X_1)$ is meaningful since both $\operatorname{grad} h(X_1)$ and $\operatorname{grad} h(X_2)$ can be treated as vectors in $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. We note that $\operatorname{grad} h$ is Lipschitz continuous on the compact region $\mathcal{Z}$ defined in (3.1); i.e., there exists a constant $\beta_{L_1} > 0$ such that

$$\|\operatorname{grad} h(X_2) - \operatorname{grad} h(X_1)\| \le \beta_{L_1} \operatorname{dist}(X_1, X_2) \tag{3.4}$$

for all $X_1, X_2 \in \mathcal{Z} \subset \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$, where "dist" defines the Riemannian distance on $(\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}, g)$ [4, p. 46]. Since $\mathcal{V}$ is a Euclidean space, for the retraction R defined on $\mathcal{V}$, we have

$$\|\gamma_V\| = \operatorname{dist}\big(R_V(\gamma_V), V\big) \quad \forall\, V \in \mathcal{V},\ \gamma_V \in T_V\mathcal{V}. \tag{3.5}$$

We observe that the manifolds $\mathcal{OB}$ and $\mathcal{O}(n)$ are both compact manifolds. By (3.5), for the retraction R defined on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$, there exist two constants $\mu > 0$ and $\delta_\mu > 0$ such that [4, p. 149]

$$\mu\|\xi_X\| \geq \operatorname{dist}\big(R_X(\xi_X), X\big) \tag{3.6}$$

for all $X \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$, and $\xi_X \in T_X\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ with $\|\xi_X\| \leq \delta_\mu$. We get by (2.22) for all k sufficiently large,

$$\|\alpha_{k-1}\Delta X_{k-1}\| \leq \delta_\mu. \tag{3.7}$$

We have by (3.3), (3.4), (3.6), and (3.7), for all k sufficiently large,

$$\begin{aligned}\|Y^{k-1}\| &= \big\|\Pi_{X^k}\big(\operatorname{grad} h(X^k) - \operatorname{grad} h(X^{k-1})\big)\big\| \\ &\leq \|\operatorname{grad} h(X^k) - \operatorname{grad} h(X^{k-1})\| \\ &\leq \beta_{L_1}\operatorname{dist}\big(X^k, X^{k-1}\big) = \beta_{L_1}\operatorname{dist}\big(R_{X^{k-1}}(\alpha_{k-1}\Delta X_{k-1}), X^{k-1}\big) \\ &\leq \beta_{L_1}\mu\|\alpha_{k-1}\Delta X_{k-1}\| \equiv \nu\alpha_{k-1}\|\Delta X_{k-1}\|,\end{aligned}$$

where $\nu = \beta_{L_1}\mu$. The proof is complete. ∎

In Algorithm 2.1, we observe that for all $k \geq 1$,

$$\|\mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}}\Delta X_{k-1}\| = \|\Pi_{X^k}\Delta X_{k-1}\| \leq \|\Delta X_{k-1}\|. \tag{3.8}$$

From (2.17), (2.18), and (3.8), for all $k \geq 1$,

$$\left\{\begin{aligned}\beta_k &= \frac{\langle \operatorname{grad} h(X^k), Y^{k-1}\rangle}{\|\operatorname{grad} h(X^{k-1})\|^2} \leq \frac{\|\operatorname{grad} h(X^k)\| \cdot \|Y^{k-1}\|}{\|\operatorname{grad} h(X^{k-1})\|^2}, \\ \theta_k &= \frac{\langle \operatorname{grad} h(X^k), \mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}}\Delta X_{k-1}\rangle}{\|\operatorname{grad} h(X^{k-1})\|^2} \\ &\leq \frac{\|\operatorname{grad} h(X^k)\| \cdot \|\mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}}\Delta X_{k-1}\|}{\|\operatorname{grad} h(X^{k-1})\|^2} \leq \frac{\|\operatorname{grad} h(X^k)\| \cdot \|\Delta X_{k-1}\|}{\|\operatorname{grad} h(X^{k-1})\|^2}.\end{aligned}\right. \tag{3.9}$$

By Lemma 3.1, $\mathcal{Z}$ is compact. In addition, h is continuously differentiable and $\operatorname{grad} h$ can be seen as a continuous nonlinear mapping defined between $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ and $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. Then there exists a constant $\gamma > 0$ such that

$$\|\operatorname{grad} h(X)\| \leq \gamma \quad \forall X \in \mathcal{Z}. \tag{3.10}$$

We have known that the sequence $\{h(X^k)\}$ is monotone decreasing. Thus the sequence $\{X^k\}$ generated by Algorithm 2.1 is contained in $\mathcal{Z}$.

Similar to the proof of Lemma 3.1 in [39], we can establish the following result.

LEMMA 3.3. *If there exists a constant* $\epsilon > 0$ *such that*

$$\|\operatorname{grad} h(X^k)\| \geq \epsilon \quad \forall k, \tag{3.11}$$

then there exists a constant $\Theta > 0$ *such that*

$$\|\Delta X_k\| \leq \Theta \quad \forall k.$$

Proof. By Proposition 3.2, (2.15), (3.8), (3.9), (3.10), and (3.11), there exists an integer $k_1 > 0$ such that for all $k > k_1$,

$$\begin{aligned} \|\Delta X_k\| &\le \|\operatorname{grad} h(X^k)\| + \|\beta_k \mathcal{T}_{\alpha_{k-1}\Delta X_{k-1}} \Delta X_{k-1}\| + \|\theta_k Y^{k-1}\| \\ &\le \gamma + \frac{\|\operatorname{grad} h(X^k)\| \cdot \|Y^{k-1}\|}{\|\operatorname{grad} h(X^{k-1})\|^2}\|\Delta X_{k-1}\| + \frac{\|\operatorname{grad} h(X^k)\| \cdot \|\Delta X_{k-1}\|}{\|\operatorname{grad} h(X^{k-1})\|^2}\|Y^{k-1}\| \\ &= \gamma + 2\frac{\|\operatorname{grad} h(X^k)\| \cdot \|Y^{k-1}\|}{\|\operatorname{grad} h(X^{k-1})\|^2}\|\Delta X_{k-1}\| \\ &\le \gamma + 2\frac{\gamma\nu\alpha_{k-1}\|\Delta X_{k-1}\|}{\epsilon^2}\|\Delta X_{k-1}\|. \end{aligned} \tag{3.12}$$

By (2.22), we get

$$\lim_{k\to\infty} \frac{2\gamma\nu}{\epsilon^2}\alpha_{k-1}\|\Delta X_{k-1}\| = 0.$$

Then there exist a constant $r \in (0,1)$ and an integer $k_2 > 0$ such that for all $k > k_2$, we have the following inequality:

$$\frac{2\gamma\nu}{\epsilon^2}\alpha_{k-1}\|\Delta X_{k-1}\| \le r. \tag{3.13}$$

From (3.12) and (3.13), it follows that for all $k > k_0 := \max\{k_1, k_2\}$,

$$\begin{aligned} \|\Delta X_k\| &\le \gamma + r\|\Delta X_{k-1}\| \\ &\le \gamma(1 + r + r^2 + \cdots + r^{k-k_0-1}) + r^{k-k_0}\|\Delta X_{k_0}\| \\ &\le \tfrac{\gamma}{1-r} + \|\Delta X_{k_0}\|. \end{aligned}$$

Let $\Theta_1 := \max\left\{\|\Delta X_1\|, \|\Delta X_2\|, \ldots, \|\Delta X_{k_0}\|\right\}$. We have

$$\|\Delta X_k\| \le \max\left\{\Theta_1, \frac{\gamma}{1-r} + \|\Delta X_{k_0}\|\right\} \equiv \Theta \quad \forall k. \qquad \square$$

We now establish the global convergence of Algorithm 2.1. The proof can be viewed as a generalization of [39, Theorem 3.2].

THEOREM 3.4. *Let $\{X^k\}$ be the sequence generated by Algorithm* 2.1. *Then we have*

$$\liminf_{k\to\infty} \|\operatorname{grad} h(X^k)\| = 0.$$

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

$$\|\operatorname{grad} h(X^k)\| \ge \epsilon \quad \forall k. \tag{3.14}$$

We have by (2.21) that for all $k \ge 1$,

$$\alpha_k \|\operatorname{grad} h(X^k)\|^2 = -\alpha_k \langle \Delta X_k, \operatorname{grad} h(X^k)\rangle \le \alpha_k \|\Delta X_k\| \cdot \|\operatorname{grad} h(X^k)\|,$$

which implies that

$$0 \le \alpha_k \|\operatorname{grad} h(X^k)\| \le \alpha_k \|\Delta X_k\| \quad \forall k \ge 1.$$

We get by (2.22),

$$\lim_{k\to\infty} \alpha_k \|\operatorname{grad} h(X^k)\| = 0.$$

Hence, if $\liminf_{k\to\infty} \alpha_k > 0$, then $\liminf_{k\to\infty} \|\operatorname{grad} h(X^k)\| = 0$. This contradicts (3.14).

We now suppose that $\liminf_{k\to\infty} \alpha_k = 0$. By taking a subsequence if necessary, we may assume that

$$\lim_{k\to\infty} \alpha_k = 0. \tag{3.15}$$

According to Step 3 of Algorithm 2.1, we have for all k sufficiently large,

$$h\big(R_{X^k}(\rho^{-1}\alpha_k \Delta X_k)\big) - h(X^k) \geq -\delta\rho^{-2}\alpha_k^2 \|\Delta X_k\|^2. \tag{3.16}$$

We note that the pullback function $\widehat{h}$ is continuously differentiable since both the cost function h and the retraction mapping R are continuously differentiable. Then there exist two constants $\delta_2 > 0$ and $\beta_{L_2} > 0$ such that

$$\big\| \operatorname{grad} \widehat{h}_X(\eta_X) - \operatorname{grad} \widehat{h}_X(\xi_X) \big\| \leq \beta_{L_2} \|\eta_X - \xi_X\| \tag{3.17}$$

for any $X \in \mathcal{Z}$ and $\xi_X, \eta_X \in T_X\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ with $\|\eta_X - \xi_X\| < \delta_2$. By using the mean-value theorem, Lemma 3.3, (2.14), (2.21), (2.22), (3.16), and (3.17), there exists a constant $\omega_k \in (0,1)$ such that for all k sufficiently large,

$$\begin{aligned}
&h\big(R_{X^k}(\rho^{-1}\alpha_k \Delta X_k)\big) - h(X^k) \\
&= h\big(R_{X^k}(\rho^{-1}\alpha_k \Delta X_k)\big) - h\big(R_{X^k}(\mathbf{0}_{X^k})\big) \\
&= \widehat{h}_{X^k}\big(\rho^{-1}\alpha_k \Delta X_k\big) - \widehat{h}_{X^k}\big(\mathbf{0}_{X^k}\big) \\
&= \rho^{-1}\alpha_k \langle \operatorname{grad} \widehat{h}_{X^k}\big(\omega_k \rho^{-1}\alpha_k \Delta X_k\big), \Delta X_k \rangle \\
&= \rho^{-1}\alpha_k \langle \operatorname{grad} \widehat{h}_{X^k}(\mathbf{0}_{X^k}), \Delta X_k \rangle + \rho^{-1}\alpha_k \langle \Delta X_k, \operatorname{grad} \widehat{h}_{X^k}\big(\omega_k \rho^{-1}\alpha_k \Delta X_k\big) \rangle \\
&\quad -\rho^{-1}\alpha_k \langle \Delta X_k, \operatorname{grad} \widehat{h}_{X^k}(\mathbf{0}_{X^k}) \rangle \\
&\leq \rho^{-1}\alpha_k \langle \operatorname{grad} \widehat{h}_{X^k}(\mathbf{0}_{X^k}), \Delta X_k \rangle + \beta_{L_2}\omega_k^2 \rho^{-2}\alpha_k^2 \|\Delta X_k\|^2 \\
&\leq -\rho^{-1}\alpha_k \|\operatorname{grad} h(X^k)\|^2 + \beta_{L_2}\rho^{-2}\alpha_k^2 \|\Delta X_k\|^2.
\end{aligned} \tag{3.18}$$

Combining (3.16) with (3.18) yields for all k sufficiently large,

$$\|\operatorname{grad} h(X^k)\|^2 \leq (\beta_{L_2} + \delta)\rho^{-1}\alpha_k \|\Delta X_k\|^2.$$

By Lemma 3.3, we know that the sequence $\{\Delta X_k\}$ is bounded. This, together with (3.15), gives rise to

$$\lim_{k\to\infty} \|\operatorname{grad} h(X^k)\| = 0.$$

This is also a contradiction. Thus the proof is complete. ∎

Next, we establish the relationship between a stationary point of h in (2.3) and a solution to the StIEP via Algorithm 2.1. According to (2.2) and (2.9), we have

$$\begin{aligned}
&\Pi_P \frac{\partial}{\partial P} \overline{h}(S,P,V) \\
&= \frac{1}{2}\big([S \odot S - H(S,P,V), H(S,P,V)^T] + [(S \odot S - H(S,P,V))^T, H(S,P,V)]\big)P \\
&= \frac{1}{2}\big([S \odot S, H(S,P,V)^T] + [(S \odot S)^T, H(S,P,V)]\big)P.
\end{aligned}$$

Then a stationary point $(S, P, V) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ of h is determined by $\operatorname{grad} h(S, P, V) = \mathbf{0}_{(S,P,V)}$, i.e.,

$$\begin{cases} S \odot H(S,P,V) - \operatorname{diag}\big(S(S \odot H(S,P,V))^T\big)S = \mathbf{0}_{n\times n}, \\ [S \odot S, (H(S,P,V))^T] + [(S \odot S)^T, H(S,P,V)] = \mathbf{0}_{n\times n}, \\ W \odot (P^T H(S,P,V)P) = \mathbf{0}_{n\times n}. \end{cases}$$

Define $\widehat{T}, \widetilde{T} \in \mathbb{R}^{n^2\times n^2}$ by

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

for any matrix $A \in \mathbb{R}^{n\times n}$, where vec(A) is an n^2-vector obtained by stacking the columns of A on top of one another. Then one has [7, p. 448]

$$\widehat{T}(A \otimes B) = (B \otimes A)\widehat{T} \quad \forall A, B \in \mathbb{R}^{n\times n}.$$

Hence, we get by vectorizing the above matrix equation,

$$\begin{cases} \big(I_{n^2} - (S^T \otimes I_n)\widetilde{T}\widehat{T}(S \otimes I_n)\big)\operatorname{Diag}(\operatorname{vec}(S))\operatorname{vec}(H(S,P,V)) = \mathbf{0}_{n^2}, \\ (I_{n^2} - \widehat{T})\big((S \odot S) \otimes I_n - I_n \otimes (S \odot S)^T\big)\operatorname{vec}(H(S,P,V)) = \mathbf{0}_{n^2}, \\ \operatorname{Diag}\big(\operatorname{vec}(W)\big)(P \otimes P)^T\operatorname{vec}(H(S,P,V)) = \mathbf{0}_{n^2}, \end{cases}$$

where $\mathbf{0}_{n^2}$ is an n^2-vector of all zeros. Therefore, we have the following result on the relationship between a stationary point of h and a solution to the StIEP.

THEOREM 3.5. *Let $\overline{X} := (\overline{S}, \overline{P}, \overline{V}) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{V}$ be an accumulation point of the sequence $\{X^k := (S^k, P^k, V^k)\}$ generated by Algorithm* 2.1 *such that* $\operatorname{grad} h(\overline{S}, \overline{P}, \overline{V}) = \mathbf{0}_{(\overline{S},\overline{P},\overline{V})}$. *If the matrix*

$$\begin{pmatrix} \big(I_{n^2} - (\overline{S}^T \otimes I_n)\widetilde{T}\widehat{T}(\overline{S} \otimes I_n)\big)\operatorname{Diag}(\operatorname{vec}(\overline{S})) \\ (I_{n^2} - \widehat{T})\big((\overline{S} \odot \overline{S}) \otimes I_n - I_n \otimes (\overline{S} \odot \overline{S})^T\big) \\ \operatorname{Diag}\big(\operatorname{vec}(W)\big)(\overline{P} \otimes \overline{P})^T \end{pmatrix}$$

is of full rank, then $\overline{C} := \overline{S} \otimes \overline{S}$ is a solution to the StIEP.

4. StIEP with prescribed entries. In this section, we consider the StIEP-PE. In many applications, the underlying structure of a desired stochastic matrix is often characterized by the prescribed entries at arbitrary locations. Here, we present a one-dimensional random walk model as an illustrative example (see, for instance, [6, p. 45] or [35, Chap. VI]). A particle moves along a line every second. From state i, the particle jumps a step to the right (i.e., state $i+1$) with probability p_i, jumps to the left (i.e., state $i-1$) with probability q_i, or stays in state i with probability $r_i = 1 - p_i - q_i$. This random process can be defined as a Markov chain whose state space (possible positions) can be restricted to $\{0, 1, 2, \ldots\}$. The transition probabilities (the probability $\widehat{P}_{i,j}$ of moving from state i to state j) are given by

$$\widehat{P}_{i,j} = \begin{cases} p_i & \text{if } i - j = -1, \\ r_i = 1 - p_i - q_i & \text{if } i - j = 0, \\ q_i & \text{if } i - j = 1, \\ 0 & \text{if } |i - j| \geq 2. \end{cases}$$

Suppose that state 0 is a reflecting barrier (i.e., $q_0 = 0$) and that $p_{n-1} = 0$ (i.e., the state space is $\{0, 1, 2, \ldots, n-1\}$). Then the transition matrix is described by

$$\widehat{P} \;=\; \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \\ n-1 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & n-1 \end{array} \\ \left(\begin{array}{cccccc} r_0 & p_0 & 0 & 0 & \cdots & 0 \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 \\ 0 & q_2 & r_2 & p_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & q_{n-2} & r_{n-2} & p_{n-2} \\ 0 & 0 & \cdots & 0 & q_{n-1} & r_{n-1} \end{array} \right) \end{array}.$$

Then the structure of the transition matrix $\widehat{P}$ can be characterized by the prescribed entries $\widehat{P}_{ij} = 0$ for all $(i, j) \in \mathcal{L}$, where $\mathcal{L} = \{(i, j) : \widehat{P}_{ij} = 0,\ i, j = 0, 1, \ldots, n-1\}$. For more details on random walk and its applications, see, for instance, [15, 22, 28, 35].

Define a diagonal matrix $\widehat{I}_n$ and $\widehat{W} \in \mathbb{R}^{n\times n}$ by

$$\widehat{I}_n := I_n - \mathrm{Diag}\big(\widehat{C}_a \mathbf{e}\big) \quad \text{and} \quad \widehat{W}_{ij} := \begin{cases} 1 & \text{if } (i, j) \in \mathcal{L}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{e} = (1, \ldots, 1)^T \in \mathbb{R}^n$. Here, we assume that the given index subset $\mathcal{L}$ is such that $\sum_{j=1}^n (\widehat{C}_a)_{ij} < 1$ for $i = 1, \ldots, n$. Then $\widehat{I}_n$ is nonsingular. Thus the StIEP-PE aims to solve the following nonlinear matrix equation:

$$\Psi(S, P, V) = \mathbf{0}_{n\times n}$$

for $(S, P, V) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$, where the set $\widehat{\mathcal{OB}}$ is a manifold defined by

$$\widehat{\mathcal{OB}} := \big\{ S \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(SS^T) = \widehat{I}_n,\ \widehat{W} \odot S = \mathbf{0}_{n\times n} \big\}.$$

The mapping $\Psi : \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V} \to \mathbb{R}^{n\times n}$ is defined by

$$\Psi(S, P, V) = \widehat{C}_a + S \odot S - P(\Lambda + V)P^T, \quad (S, P, V) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}. \tag{4.1}$$

Note that the dimension of the product manifold $\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$ is given by

$$\dim(\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}) = n(n-1) - |\mathcal{L}| + \frac{n(n-1)}{2} + |\mathcal{J}|.$$

We see that the nonlinear equation $\Psi(S, P, V) = \mathbf{0}_{n\times n}$ is an under-determined system defined from $\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$ to $\mathbb{R}^{n\times n}$ if n is large and the number $|\mathcal{L}|$ of prescribed entries is small.

If we find a solution $(\overline{S}, \overline{P}, \overline{V}) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$ to $\Psi(S, P, V) = \mathbf{0}_{n\times n}$, then $\overline{C} = \widehat{C}_a + \overline{S} \odot \overline{S}$ is a solution to the StIEP-PE. Alternatively, one may solve the StIEP-PE by finding a global solution to the following minimization problem:

$$\begin{array}{lll} \min & \psi(S, P, V) := \dfrac{1}{2}\|\Psi\|_F^2 = \dfrac{1}{2}\|\widehat{C}_a + S \odot S - P(\Lambda + V)P^T\|_F^2 \\ \text{s.t.} & S \in \widehat{\mathcal{OB}}, \quad P \in \mathcal{O}(n), \quad V \in \mathcal{V}. \end{array} \tag{4.2}$$

Let $\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$ be equipped with the Riemannian metric defined as in (2.4). It follows that the tangent space of $\widehat{\mathcal{OB}}$ at a point $S \in \widehat{\mathcal{OB}}$ is given by

$$T_S\widehat{\mathcal{OB}} = \big\{ T \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(ST^T) = \mathbf{0}_{n\times n},\ \widehat{W} \odot T = \mathbf{0}_{n\times n} \big\}.$$

Then the orthogonal projection of a point $\xi \in \mathbb{R}^{n\times n}$ onto $T_S\widehat{\mathcal{OB}}$ is given by

$$\widehat{\Pi}_S\xi = (E - \widehat{W}) \odot \xi - \widehat{I}_n^{-1} \operatorname{diag}\big(S((E - \widehat{W}) \odot \xi)^T\big)S, \tag{4.3}$$

where E is an n-by-n matrix of ones. The retraction on $\widehat{\mathcal{OB}}$ at a point $S \in \widehat{\mathcal{OB}}$ can be defined by

$$\widehat{R}_S(\xi_S) = \widehat{I}_n^{\frac{1}{2}}\Big(\operatorname{diag}\big((S+\xi_S)(S+\xi_S)^T\big)\Big)^{-\frac{1}{2}}(S+\xi_S) \quad \forall \xi_S \in T_S\widehat{\mathcal{OB}}.$$

The vector transport on $\widehat{\mathcal{OB}}$ is given by

$$\widehat{\mathcal{T}}_{\eta_S}(\xi_S) = \xi_S - \Big(\operatorname{diag}\big((S+\eta_S)(S+\eta_S)^T\big)\Big)^{-1}\operatorname{diag}\big((S+\eta_S)\xi_S^T\big)(S+\eta_S)$$

for any $S \in \widehat{\mathcal{OB}}$ and $\xi_S, \eta_S \in T_S\widehat{\mathcal{OB}}$.

Next, we establish explicit formulas for the differential of the smooth mapping Ψ defined in (4.1) and the Riemannian gradient of the cost function ψ defined in problem (4.2). Define the mapping $\overline{\Psi} : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and the cost function $\overline{\psi} : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}$ by

$$\overline{\Psi}(S,P,V) := \widehat{C}_a + S \odot S - P(\Lambda + V)P^T \quad \text{and} \quad \overline{\psi}(S,P,V) := \frac{1}{2}\|\overline{\Psi}(S,P,V)\|_F^2$$

for all $(S,P,V) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. As in section 2, by simple calculation, the Riemannian gradient of ψ at a point $(S,P,V) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$ is given by [4, p. 48]

$$\operatorname{grad}\psi(S,P,V) = \left(\widehat{\Pi}_S\Big(\frac{\partial}{\partial S}\overline{\psi}(S,P,V)\Big), \widehat{\Pi}_P\Big(\frac{\partial}{\partial P}\overline{\psi}(S,P,V)\Big), \widehat{\Pi}_V\Big(\frac{\partial}{\partial V}\overline{\psi}(S,P,V)\Big)\right),$$

where

$$\begin{cases} \widehat{\Pi}_S\big(\frac{\partial}{\partial S}\overline{\psi}(S,P,V)\big) = 2S \odot \Psi(S,P,V) - 2\widehat{I}_n^{-1}\operatorname{diag}\big(S(S\odot\Psi(S,P,V))^T\big)S, \\ \widehat{\Pi}_P\big(\frac{\partial}{\partial P}\overline{\psi}(S,P,V)\big) = \dfrac{1}{2}\big([P(\Lambda+V)P^T, (\Psi(S,P,V))^T] + [P(\Lambda+V)^TP^T, \Psi(S,P,V)]\big)P, \\ \widehat{\Pi}_V\big(\frac{\partial}{\partial V}\overline{\psi}(S,P,V)\big) = -W \odot \big(P^T\Psi(S,P,V)P\big). \end{cases}$$

Then the steepest descent flow for the cost function ψ over the product manifold $\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$ is given by

$$\frac{d(S,P,V)}{dt} = -\operatorname{grad}\psi(S,P,V). \tag{4.4}$$

Given an initial value $(S(0),P(0),V(0)) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$, this reformulates problem (4.2) as an initial value problem. Then one may get a solution to problem (4.2) by using existing ODE solvers for the above differential equation. This extra geometric isospectral flow method is effective for small- and medium-scale problems (see section 5).

Similarly, the differential $\mathrm{D}\Psi(S,P,V) : T_{(S,P,V)}\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V} \to T_{\Psi(S,P,V)}\mathbb{R}^{n\times n} \simeq \mathbb{R}^{n\times n}$ of Ψ at a point $(S,P,V) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$ is determined by

$$\mathrm{D}\Psi(S,P,V)[(\Delta S, \Delta P, \Delta V)] = 2S \odot \Delta S + [P(\Lambda + V)P^T, \Delta P P^T] - P\Delta V P^T$$

for all $(\Delta S, \Delta P, \Delta V) \in T_{(S,P,V)}\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$.

As in section 2.3, one may solve problem (4.2) by using Algorithm 2.1, where a reasonable initial step length can be set to

$$t_k = \frac{\big|\langle \operatorname{grad} \psi(X^k), \Delta X_k\rangle\big|}{\|\mathrm{D}\Psi(X^k)[\Delta X_k]\|^2}. \tag{4.5}$$

The global convergence can also be established. By following similar arguments in Theorem 3.5, one has the following result on the relationship between a stationary point of ψ and a solution to the StIEP-PE.

THEOREM 4.1. *Let* $\overline{X} := (\overline{S}, \overline{P}, \overline{V}) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{V}$ *be an accumulation point of the sequence* $\{X^k := (S^k, P^k, V^k)\}$ *generated by Algorithm* 2.1 *for solving problem* (4.2) *such that* $\operatorname{grad} \psi(\overline{S}, \overline{P}, \overline{V}) = \mathbf{0}_{(\overline{S},\overline{P},\overline{V})}$. *If the matrix*

$$\begin{pmatrix} \big(I_{n^2} - (\overline{S}^T \otimes \widehat{I}_n^{-1})\widetilde{T}\widehat{T}(\overline{S} \otimes I_n)\big)\mathrm{Diag}(\mathrm{vec}(\overline{S})) \\ (I_{n^2} - \widehat{T})\big((\widehat{C}_a + \overline{S} \odot \overline{S}) \otimes I_n - I_n \otimes (\widehat{C}_a + \overline{S} \odot \overline{S})^T\big) \\ \mathrm{Diag}\big(\mathrm{vec}(W)\big)(\overline{P} \otimes \overline{P})^T \end{pmatrix}$$

is of full rank, then $\overline{C} = \widehat{C}_a + \overline{S} \odot \overline{S}$ *is a solution to the StIEP-PE.*

5. Numerical tests. We report the numerical performance of Algorithm 2.1 for solving problems (2.3) and (4.2). All numerical tests are carried out using MATLAB on a Linux server having 8 cores with 4 Intel Xeon E7-8837 processors at 2.67GHz and 256G RAM.

For Algorithm 2.1, the starting points are generated randomly by the built-in functions `rand` and `schur`. In particular, for problem (2.3),

$$\begin{aligned} &\widehat{S} \odot \widehat{S} = \texttt{rand}\,(n,n), \quad S^0 = \big(\,\mathrm{diag}(\widehat{S}\widehat{S}^T)\big)^{-\frac{1}{2}}\widehat{S} = S(0) \in \mathcal{OB}, \\ &[P^0, \widehat{V}] = \texttt{schur}\,(S^0 \odot S^0, \text{`real'}) = [P(0), \widehat{V}], \quad V^0 = W \odot \widehat{V} = V(0). \end{aligned} \tag{5.1}$$

For problem (4.2),

$$\begin{aligned} &S^0 = (\widehat{I}_n)^{\frac{1}{2}}\big(\,\mathrm{diag}(((E - \widehat{W}) \odot \widehat{S})((E - \widehat{W}) \odot \widehat{S})^T)\big)^{-\frac{1}{2}}((E - \widehat{W}) \odot \widehat{S}) = S(0) \in \widehat{\mathcal{OB}}, \\ &[P^0, \widehat{V}] = \texttt{schur}\,(\widehat{C}_a + S^0 \odot S^0, \text{`real'}) = [P(0), \widehat{V}], \quad V^0 = W \odot \widehat{V} = V(0). \end{aligned} \tag{5.2}$$

In our numerical tests, the stopping criterion for Algorithm 2.1 is set to be

$$\|H(S^k, P^k, V^k)\|_F < 10^{-12} \quad \text{or} \quad \|\Psi(S^k, P^k, V^k)\|_F < 10^{-12}.$$

We also set $\alpha = 1.4$, $\delta = 10^{-4}$, and $\rho = 0.5$.

In our numerical tests, '`CT.`', '`IT.`', '`NF.`', '`Err.`', and '`Res.`' mean the total computing time, the number of iterations, the number of function evaluations, the error $\|H(S^k, P^k, V^k)\|_F$ or $\|\Psi(S^k, P^k, V^k)\|_F$, and the residual $\|\operatorname{grad} h(S^k, P^k, V^k)\|$ or $\|\operatorname{grad} \psi(S^k, P^k, V^k)\|$ at the final iterate of the corresponding algorithms, accordingly.

We consider the following two examples with different problem size n.

Example 5.1. We consider the StIEP 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

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

TABLE 1
Numerical results for Example 5.1.

Alg.	n	CT.	IT.	NF.	Err.	Res.
Alg. 2.1	200	12.4 s	376	436	8.71×10^{-13}	4.90×10^{-13}
	400	52.7 s	544	604	9.84×10^{-13}	5.84×10^{-13}
	600	02 m 19 s	665	729	9.64×10^{-13}	2.54×10^{-13}
	800	05 m 29 s	775	879	9.81×10^{-13}	1.85×10^{-13}
	1000	13 m 14 s	1169	1611	9.71×10^{-13}	2.16×10^{-13}
	2000	01 h 38 m 48 s	1453	1431	9.73×10^{-13}	3.25×10^{-13}
Alg. 2.1 with (2.25)	200	8.4 s	204	205	8.54×10^{-13}	3.37×10^{-13}
	400	36.4 s	265	266	9.94×10^{-13}	1.53×10^{-13}
	600	01 m 36 s	291	292	9.13×10^{-13}	3.53×10^{-13}
	800	03 m 29 s	321	322	9.98×10^{-13}	2.93×10^{-13}
	1000	05 m 56 s	308	309	9.96×10^{-13}	5.75×10^{-13}
	2000	57 m 43 s	357	358	9.98×10^{-13}	1.43×10^{-13}

TABLE 2
Numerical results for Example 5.2.

Alg.	n	$\|\mathcal{L}\|$	CT.	IT.	NF.	Err.	Res.
Alg. 2.1	200	4103	14.1 s	487	537	9.99×10^{-13}	1.08×10^{-13}
	400	16013	01 m 01 s	698	805	9.66×10^{-13}	3.20×10^{-13}
	600	36308	02 m 52 s	882	961	9.63×10^{-13}	5.00×10^{-13}
	800	64035	06 m 22 s	953	1072	8.92×10^{-13}	1.25×10^{-13}
	1000	99302	11 m 56 s	1093	1209	9.38×10^{-13}	1.51×10^{-13}
	2000	398989	01 h 58 m 21 s	2001	2214	8.78×10^{-13}	3.48×10^{-13}
Alg. 2.1 with (4.5)	200	4103	8.71 s	270	271	8.15×10^{-13}	3.33×10^{-13}
	400	16013	31.1 s	323	324	9.96×10^{-13}	2.63×10^{-13}
	600	36308	01 m 32 s	413	414	9.82×10^{-13}	4.36×10^{-13}
	800	64035	03 m 16 s	429	430	9.38×10^{-13}	2.30×10^{-13}
	1000	99302	05 m 19 s	418	419	9.35×10^{-13}	3.53×10^{-13}
	2000	398989	34 m 52 s	506	507	9.46×10^{-13}	2.46×10^{-13}

We choose the eigenvalues of $\widehat{C}$ as the prescribed spectrum.

Example 5.2. We consider the StIEP-PE with varying n. Let $\widehat{C}$ be a random stochastic matrix generated as in Example 5.1. We choose the eigenvalues of $\widehat{C}$ as the prescribed spectrum. Also, we choose the index subset $\mathcal{L} := \big\{(i,j) \mid 3/(5n) < (\widehat{C})_{ij} < 4/(5n),\ i,j = 1,\ldots,n\big\}$. The prescribed nonnegative matrix $\widehat{C}_a \in \mathbb{R}^{n\times n}$ is defined by $(\widehat{C}_a)_{ij} = (\widehat{C})_{ij}$ if $(i,j) \in \mathcal{L}$ and $(\widehat{C}_a)_{ij} = 0$ if $(i,j) \notin \mathcal{L}$.

For demonstration purposes, in Tables 1–2 we report the numerical results for Examples 5.1–5.2 with different problem size n, where the initial step length guess (2.25) or (4.5) may be used in Algorithm 2.1.

We see from Tables 1–2 that Algorithm 2.1 is very efficient for solving large-scale problems. Moreover, Algorithm 2.1 with the initial step length guess (2.25) or (4.5) can reduce the number of iterations and thus improve the effectiveness of the proposed method. Finally, from these and many other numerical tests, we observe the fact that the proposed algorithm converges to different solutions for different starting points.

Finally, to illustrate the efficiency of our algorithm, we compare Algorithm 2.1 with the MATLAB ODE solver `ode113` for solving the differential equations (2.10) and (4.4), and the ODE solver `ode113`–based steepest descent flow method in [12] for

TABLE 3
Comparison of Algorithm 2.1 *and* `ode113` *for Example* 5.1.

Alg.	n	CT.	IT.	NF.	Err.	Res.
ode113 in [12]	10	26.155 s	4.82×10^4	1.04×10^5	9.82×10^{-9}	1.34×10^{-9}
	20	01 m 07 s	8.29×10^4	1.80×10^5	9.98×10^{-9}	1.30×10^{-9}
	30	03 m 48 s	1.74×10^5	3.79×10^5	9.96×10^{-9}	1.21×10^{-9}
	40	25 m 41 s	4.36×10^5	9.49×10^5	9.98×10^{-9}	8.63×10^{-10}
	50	41 m 23 s	5.55×10^5	1.20×10^6	9.98×10^{-9}	1.19×10^{-10}
ode113 for (2.10)	10	0.2591 s	5.38×10^2	1.13×10^3	6.65×10^{-9}	2.13×10^{-9}
	20	0.6477 s	1.03×10^3	2.21×10^3	8.71×10^{-9}	1.76×10^{-9}
	30	1.0228 s	1.12×10^3	2.42×10^3	7.27×10^{-9}	1.34×10^{-9}
	40	3.5612 s	1.53×10^3	3.33×10^3	9.40×10^{-9}	1.35×10^{-9}
	50	4.9585 s	1.66×10^3	3.63×10^3	8.59×10^{-9}	1.18×10^{-9}
Alg. 2.1	10	0.0359 s	75	144	9.52×10^{-9}	7.63×10^{-9}
	20	0.0594 s	107	156	8.00×10^{-9}	3.11×10^{-9}
	30	0.0680 s	92	122	9.58×10^{-9}	3.16×10^{-9}
	40	0.2669 s	131	164	9.96×10^{-9}	4.24×10^{-9}
	50	0.3381 s	135	171	9.56×10^{-9}	4.66×10^{-9}

solving the following minimization problem:

$$\begin{array}{ll} \min & \varphi(G,P) := \dfrac{1}{2}\|G \odot G - P\Lambda P^{-1}\|_F^2 \\ \text{s.t.} & G \in \mathbb{R}^{n\times n}, \quad P \in \mathbb{R}^{n\times n} \text{ is nonsingular,} \end{array} \tag{5.3}$$

where the inverse matrix P^{-1} involved in the flow method was further characterized by an ASVD [12]. Then a solution to the StIEP is obtained by using Theorem 1.1.

For Algorithm 2.1, the ODE solver `ode113` for solving (2.10) and (4.4) and the ODE solver `ode113`–based steepest descent flow method in [12] are used. The starting points are randomly generated as in (5.1), (5.2), and

$$G(0) = S(0), \quad P(0) = \texttt{rand}\,(n,n), \quad [X(0), S(0), Y(0)] = \texttt{svd}\,(P(0)).$$

For comparison purposes, in our numerical tests the stopping criteria for Algorithm 2.1, the ODE solver `ode113` for solving (2.10) and (4.4), and the ODE solver `ode113`–based steepest descent flow method in [12] are, respectively, set to be

$$\|H(S^k, P^k, V^k)\|_F < 10^{-8}, \quad \|\Psi(S^k, P^k, V^k)\|_F < 10^{-8},$$

and

$$\|G^k \odot G^k - P^k\Lambda(P^k)^{-1}\|_F < 10^{-8}.$$

For the ODE solver `ode113`, we evaluate the output values at a time interval of 10. The integration terminates automatically when the above stopping criteria are satisfied. In addition, the other parameters in Algorithm 2.1 are set as above.

Tables 3–4 display the numerical results for Examples 5.1 and 5.2. Here, for the ODE solver `ode113`–based steepest descent flow method in [12], we still use '`Err.`' and '`Res.`' to denote the error $\|G^k \odot G^k - P^k\Lambda(P^k)^{-1}\|_F$ and the residual $\|\operatorname{grad}\varphi(G^k, P^k)\|_F$ at the final iterate of the corresponding algorithm, accordingly.

We observe from Table 3 that the ODE solver `ode113` for (2.10) works much better than the ODE solver `ode113`–based steepest descent flow method in [12] in terms of both computing time and the number of iterations. In terms of computing time, Algorithm 2.1 is the most effective.

We also see from Table 4 that the ODE solver `ode113` for (4.4) works acceptably while Algorithm 2.1 performs more effectively in terms of computing time.

TABLE 4
Comparison of Algorithm 2.1 *and* `ode113` *for Example* 5.2.

Alg.	n	$\|\mathcal{L}\|$	CT.	IT.	NF.	Err.	Res.
`ode113` for (4.4)	20	40	1.2123 s	1.82×10^3	3.94×10^3	8.52×10^{-9}	1.19×10^{-9}
	40	181	6.5945 s	2.66×10^3	5.85×10^3	9.01×10^{-9}	9.22×10^{-10}
	60	390	11.121 s	3.10×10^3	6.82×10^3	9.69×10^{-9}	9.15×10^{-10}
	80	626	15.432 s	3.43×10^3	7.54×10^3	9.80×10^{-9}	8.61×10^{-10}
	100	951	24.841 s	3.86×10^3	8.50×10^3	9.57×10^{-9}	7.97×10^{-10}
Alg. 2.1	20	40	0.1031 s	148	207	8.51×10^{-9}	3.46×10^{-9}
	40	181	0.3399 s	160	182	8.64×10^{-9}	4.79×10^{-9}
	60	390	0.6174 s	188	229	9.43×10^{-9}	6.55×10^{-9}
	80	626	0.8228 s	203	244	7.92×10^{-9}	4.83×10^{-9}
	100	951	1.2415 s	204	211	9.13×10^{-9}	4.15×10^{-9}

6. Concluding remarks. In this paper, we have proposed a geometric nonlinear conjugate gradient algorithm for solving the stochastic inverse eigenvalue problem. The global convergence of the proposed algorithm is established. The proposed algorithm is also extended to the case of prescribed entries. Moreover, an extra advantage is that our models yield new isospectral flow methods for both StIEP and StIEP-PE. Numerical experiments show the effectiveness of the proposed algorithm for large-scale problems, while our new isospectral flow methods work acceptably well for small- and medium-scale problems. As noted in [11, p. 104], an interesting question is how to generalize the proposed algorithm to the stochastic inverse eigenvalue problem with a specified stationary vector. This needs further investigation.

Acknowledgments. We would like to thank the associate editor and the referees for their valuable comments and suggestions which have considerably improved this paper.

REFERENCES

[1] P.-A. Absil, C. G. Baker, and K. A. Gallivan, *A truncated-CG style method for symmetric generalized eigenvalue problems*, J. Comput. Appl. Math., 189 (2006), pp. 274–285.
[2] P.-A. Absil, C. G. Baker, and K. A. Gallivan, *Trust-region methods on Riemannian manifolds*, Found. Comput. Math., 7 (2007), pp. 303–330.
[3] P.-A. Absil and K. A. Gallivan, *Joint diagonalization on the oblique manifold for independent component analysis*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Volume 5, Toulouse, 2006, pp. V-945–V-948.
[4] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.
[5] C. G. Baker, P.-A. Absil, and K. A. Gallivan, *An implicit Riemannian trust-region method for the symmetric generalized eigenproblem*, in Computational Science (ICCS 2006), Lecture Notes in Comput. Sci. 3991, V. N. Alexandrov, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, eds., Springer, New York, 2006, pp. 210–217.
[6] R. B. Bapat and T. E. S. Raghavan, *Nonnegative Matrices and Applications*, Cambridge University Press, Cambridge, UK, 1997.
[7] D. S. Bernstein, *Matrix Mathematics – Theory, Facts, and Formulas*, 2nd ed., Princeton University Press, Princeton, NJ, 2009.
[8] M. T. Chu, *Inverse eigenvalue problems*, SIAM Rev., 40 (1998), pp. 1–39, doi:10.1137/S0036144596303984.
[9] M. T. Chu and K. R. Driessel, *Constructing symmetric nonnegative matrices with prescribed eigenvalues by differential equations*, SIAM J. Math. Anal., 22 (1991), pp. 1372–1387, doi:10.1137/0522088.
[10] M. T. Chu and G. H. Golub, *Structured inverse eigenvalue problems*, Acta Numer., 11 (2002), pp. 1–71.
[11] M. T. Chu and G. H. Golub, *Inverse Eigenvalue Problems: Theory, Algorithms, and Applications*, Oxford University Press, Oxford, UK, 2005.

[12] M. T. Chu and Q. Guo, *A numerical method for the inverse stochastic spectrum problem*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 1027–1039, doi:10.1137/S0895479896292418.

[13] J. E. Cohen, J. H. B. Kemperman, and G. Zbăganu, *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population Sciences*, Birkhäuser, Boston, 1998.

[14] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353, doi:10.1137/S0895479895290954.

[15] N. W. Goel and N. Richter-Dyn, *Stochastic Models in Biology*, Academic Press, New York, 1974.

[16] X. Q. Jin, Y. M. Wei, and Z. Zhao, *Numerical Linear Algebra and Its Applications*, 2nd ed., Science Press, Beijing, 2015.

[17] F. I. Karpelevič, *On the characteristic roots of matrices with nonnegative elements*, Izvestiya Akad. Nauk SSSR Ser. Mat., 15 (1951), pp. 361–383 (in Russian).

[18] M. M. Lin, *An algorithm for constructing nonnegative matrices with prescribed real eigenvalues*, Appl. Math. Comput., 256 (2015), pp. 582–590.

[19] H. Minc, *Nonnegative Matrices*, Wiley-Intersci. Ser. Discrete Math. Optim., John Wiley & Sons, Inc., New York, 1988.

[20] A. V. Nazin and B. T. Polyak, *Randomized algorithm to determine the eigenvector of a stochastic matrix with application to the PageRank problem*, Autom. Remote Control, 72 (2011), pp. 342–352.

[21] R. Orsi, *Numerical methods for solving inverse eigenvalue problems for nonnegative matrices*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 190–212, doi:10.1137/050634529.

[22] K. Pearson, *The problem of the random walk*, Nature, 72 (1905), p. 294.

[23] H. Perfect, *Methods of constructing certain stochastic matrices*, Duke Math. J., 20 (1953), pp. 395–404.

[24] H. Perfect, *Methods of constructing certain stochastic matrices.* II, Duke Math. J., 22 (1955), pp. 305–311.

[25] S. Petoukhov and M. He, *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications*, IGI Global, Hershey, PA, 2010.

[26] E. Polak and G. Ribière, *Note surla convergence de méthodes de directions conjuguées*, ESAIM Math. Model. Numer. Anal., 3 (1969), pp. 35–43.

[27] B. T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comput. Math. Math. Phys., 9 (1969), pp. 94–112.

[28] P. Révész, *Random Walk in Random and Non-random Environments*, World Scientific, Singapore, 1990.

[29] W. Ring and B. Wirth, *Optimization methods on Riemannian manifolds and their application to shape space*, SIAM J. Optim., 22 (2012), pp. 596–627, doi:10.1137/11082885X.

[30] H. Sato, *A Dai-Yuan-type Riemannian conjugate gradient method with the weak Wolfe conditions*, Comput. Optim. Appl., 64 (2016), pp. 101–118.

[31] H. Sato and T. Iwai, *A new, globally convergent Riemannian conjugate gradient method*, Optimization, 64 (2015), pp. 1011–1031.

[32] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer-Verlag, New York, 1981.

[33] S. T. Smith, *Optimization techniques on Riemannian manifolds*, in Hamiltonian and Gradient Flows, Algorithms and Control, A. Bloch, ed., Fields Inst. Commun. 3, AMS, Providence, RI, 1994, pp. 113–136.

[34] G. W. Soules, *Constructing symmetric nonnegative matrices*, Linear Multilinear Algebra, 13 (1983), pp. 241–251.

[35] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed., Elsevier, Amsterdam, 2007.

[36] S. F. Xu, *An Introduction to Inverse Algebraic Eigenvalue Problems*, Peking University Press, Beijing; Friedr. Vieweg & Sohn, Braunschweig, 1998.

[37] S. J. Yang and C. Q. Xu, *Row stochastic inverse eigenvalue problem*, J. Inequalities Appl., 24 (2011), pp. 137–141.

[38] T.-T. Yao, Z.-J. Bai, Z. Zhao, and W.-K. Ching, *A Riemannian Fletcher–Reeves conjugate gradient method for doubly stochastic inverse eigenvalue problems*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 215–234, doi:10.1137/15M1023051.

[39] L. Zhang, W. J. Zhou, and D. H. Li, *A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence*, IMA J. Numer. Anal., 26 (2006), pp. 629–640.

[40] Z. Zhao, Z.-J. Bai, and X.-Q. Jin, *A Riemannian Newton algorithm for nonlinear eigenvalue problems*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 752–774, doi:10.1137/140967994.