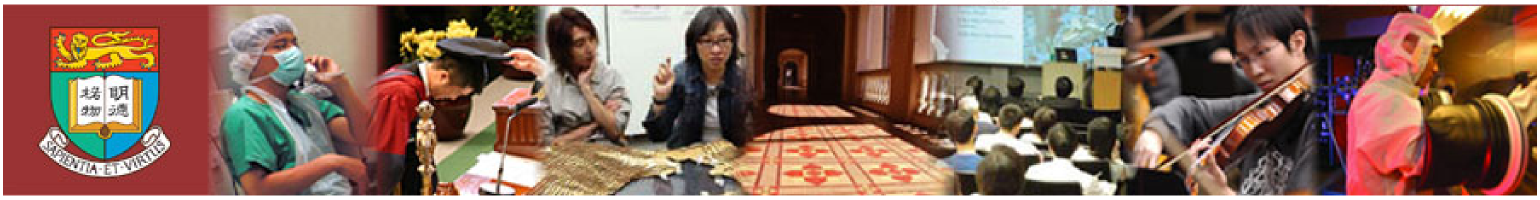

Title	**A Riemannian Fletcher-Reeves Conjugate Gradient Method for Doubly Stochastic Inverse Eigenvalue Problems**
Author(s)	**Yao, T; Bai, Z; Zhao, Z; Ching, WK**
Citation	**SIAM Journal on Matrix Analysis and Applications, 2016, v. 37 n. 1, p. 215-234**
Issued Date	**2016**
URL	**http://hdl.handle.net/10722/229223**
Rights	**SIAM Journal on Matrix Analysis and Applications. Copyright © Society for Industrial and Applied Mathematics.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.**

SIAM J. MATRIX ANAL. APPL.
Vol. 37, No. 1, pp. 215–234

A RIEMANNIAN FLETCHER–REEVES CONJUGATE GRADIENT METHOD FOR DOUBLY STOCHASTIC INVERSE EIGENVALUE PROBLEMS*

TENG-TENG YAO†, ZHENG-JIAN BAI‡, ZHI ZHAO§, AND WAI-KI CHING¶

Abstract. We consider the inverse eigenvalue problem of reconstructing a doubly stochastic matrix from the given spectrum data. We reformulate this inverse problem as a constrained nonlinear least squares problem over several matrix manifolds, which minimizes the distance between isospectral matrices and doubly stochastic matrices. Then a Riemannian Fletcher–Reeves conjugate gradient method is proposed for solving the constrained nonlinear least squares problem, and its global convergence is established. An extra gain is that a new Riemannian isospectral flow method is obtained. Our method is also extended to the case of prescribed entries. Finally, some numerical tests are reported to illustrate the efficiency of the proposed method.

Key words. inverse eigenvalue problem, doubly stochastic matrix, Riemannian Fletcher–Reeves conjugate gradient method, Riemannian isospectral flow

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

DOI. 10.1137/15M1023051

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 with each row summing to 1, i.e., $\sum_{j=1}^{n} A_{ij} = 1$ for all $i = 1, \ldots, n$. An n-by-n real matrix A is called a doubly stochastic matrix if it is nonnegative with each row and column summing to 1, i.e., $\sum_{i=1}^{n} A_{ij} = \sum_{j=1}^{n} A_{ij} = 1$ for all $i, j = 1, \ldots, n$. Stochastic matrices and doubly stochastic matrices arise in various applications such as probability theory, statistics, quantum mechanics, hypergroups, economics, computer science, graph theory, physical chemistry, and population genetics, etc. See for instance [8, 11, 15, 16, 18, 22, 24, 29, 33] and references therein.

This paper is concerned with the following doubly stochastic inverse eigenvalue problem (DSIEP).

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

In probability and combinatorics, the DSIEP aims to construct a special transition matrix (i.e., a doubly stochastic matrix) from the given spectral data. There exist

*Received by the editors May 26, 2015; accepted for publication (in revised form) by C.-H. Guo December 14, 2015; published electronically February 25, 2016.
http://www.siam.org/journals/simax/37-1/M102305.html

†School of Mathematical Sciences, Xiamen University, Xiamen 361005, People's Republic of China (yaotengteng718@163.com).

‡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) and the Fundamental Research Funds for the Central Universities (20720150001).

§Department of Mathematics, Hangzhou Dianzi University, Hangzhou, People's Republic of China (zhaozhi231@163.com).

¶Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Hong Kong (wching@hku.hk). The research of this author was supported in part by the Research Grants Council of Hong Kong under grant 17301214, HKU CRCG grants, and the HKU Strategic Research Theme Fund on Computational Science.

some solvability conditions for the DSIEP (see for instance [13, 14, 17, 19, 20, 23, 34] and references therein). However, there exist only a few numerical methods for solving the DSIEP. In particular, an idempotent system-based constructive method was proposed in [23] and Soules basis-based constructive methods were given in [21, 32]. An algorithm based on Householder transformations and rank 1 updating was presented in [39]. A fast Fourier transformation-based method was proposed in [27, 28]. However, all these constructive methods produce particular doubly stochastic matrices numerically under some special sufficient conditions.

Recently, there have been some Riemannian optimization methods for eigenproblems, which include a truncated conjugate gradient method for the symmetric generalized eigenvalue problem [1], a Riemannian trust-region method for the symmetric generalized eigenproblem [6], Newton's method and the conjugate gradient method for the symmetric eigenvalue problem [31], and a Riemannian Newton method for nonlinear eigenvalue problems [37].

In this paper, we propose a Riemannian nonlinear conjugate gradient method for solving the DSIEP. By using a real Schur matrix decomposition, the DSIEP is reformulated as a constrained nonlinear least squares problem over several matrix manifolds, where the cost function aims to minimize the distance between isospectral matrices and doubly stochastic matrices. The basic geometric properties of these matrix manifolds are studied and the Riemannian gradient of the cost function is derived. Then we propose a Riemannian Fletcher–Reeves conjugate gradient method for solving the constrained nonlinear least squares problem. This is sparked by a modified Fletcher–Reeves method proposed by Zhang, Zhou, and Li [36] and the recent development on Riemannian conjugate gradient methods [25, 26, 31]. The global convergence of the proposed method is established. An extra gain is that our model yields a new Riemannian isospectral flow method (similar to the isospectral flow method in [10] for solving the inverse stochastic spectrum problem). The proposed method is also extended to the DSIEP with prescribed entries. Finally, some numerical experiments are reported to show that the proposed geometric methods are effective for solving DSIEPs.

Throughout the paper, we use the following notation. The symbol A^T denotes the transpose of a matrix A. I_n is the identity matrix of order n. Let $\mathbb{R}^{n\times n}$ and $\mathcal{O}(n)$ be the set of all n-by-n real matrices and the set of all n-by-n orthogonal matrices, respectively. For any two matrices $A, B \in \mathbb{R}^{n\times n}$, $A \odot B$ and $[A, B] := AB - BA$ mean the Hadamard product and the Lie bracket product of A and B, respectively. For a matrix $A \in \mathbb{R}^{n\times n}$, $\mathrm{diag}(A)$ is a diagonal matrix with the same diagonal entries as A. Denote by $\mathrm{tr}(A)$ the sum of the diagonal entries of a square matrix A. Define the index set $\mathcal{N} := \{(i, j) : i, j = 1, \ldots, n\}$. For a matrix $A \in \mathbb{R}^{n\times n}$, we define $\mathrm{skew}(A) := \frac{1}{2}(A - A^T)$.

The rest of this paper is organized as follows. In section 2 we review some preliminary results on Riemannian manifolds. In section 3 we present a Riemannian Fletcher–Reeves conjugate gradient method for solving the DSIEP. In section 4 the global convergence of the proposed method is established. In section 5 we consider some extensions. Finally, some numerical tests are reported in section 6 and some concluding remarks are given in section 7.

2. Preliminaries.

In this section, we review some necessary concepts and basic results regarding Riemannian manifolds. One may refer to [2, 4] for more discussions.

Let $\mathcal{M}$ be a d-dimensional manifold. A curve c on $\mathcal{M}$ is defined as a smooth mapping from $\mathbb{R}$ to $\mathcal{M}$. Let $X \in \mathcal{M}$. Denote by $\mathfrak{F}_X(\mathcal{M})$ the set of all smooth realvalued functions defined on a neighborhood of X. A tangent vector ξ_X to $\mathcal{M}$ at X is defined as a mapping from $\mathfrak{F}_X(\mathcal{M})$ to $\mathbb{R}$ such that

$$\xi_X f = \dot{c}(0) f := \frac{d(f(c(t)))}{dt}\Big|_{t=0} \quad \forall f \in \mathfrak{F}_X(\mathcal{M}), \tag{2.1}$$

where $c : \mathbb{R} \to \mathcal{M}$ with $c(0) = X$ is any curve on $\mathcal{M}$ that realizes the tangent vector ξ_X. Let $T_X\mathcal{M}$ be the tangent space to $\mathcal{M}$ at X, which consists of all tangent vectors to $\mathcal{M}$ at X. Denote by $T\mathcal{M}$ the tangent bundle of $\mathcal{M}$:

$$T\mathcal{M} := \bigcup_{X \in \mathcal{M}} T_X\mathcal{M}.$$

A vector field on $\mathcal{M}$ is a smooth function $\xi : \mathcal{M} \to T\mathcal{M}$, which assigns a tangent vector $\xi_X \in T_X\mathcal{M}$ to each point $X \in \mathcal{M}$. A Riemannian metric g on $\mathcal{M}$ is a family of (positive definite) inner products

$$g_X : T_X\mathcal{M} \times T_X\mathcal{M} \to \mathbb{R}, \quad X \in \mathcal{M},$$

such that, for all smooth vector fields ξ, ζ on $\mathcal{M}$, $X \to g_X(\xi_X, \zeta_X)$ is a smooth function. The inner product $g_X(\cdot,\cdot)$ induces a norm $\|\xi_X\| = \sqrt{g_X(\xi_X, \xi_X)}$ on $T_X\mathcal{M}$. In this case, $(\mathcal{M}, g)$ is called a Riemannian manifold [4, p. 45]. In addition, if $\mathcal{M}$ is an embedded submanifold of a Riemannian manifold $(\overline{\mathcal{M}}, \bar{g})$, then $\mathcal{M}$ is also a Riemannian manifold if it is endowed with the Riemannian metric

$$g_X(\xi_X, \zeta_X) := \bar{g}_X(\xi_X, \zeta_X), \quad \xi_X, \zeta_X \in T_X\mathcal{M},$$

where ξ_X and ζ_X are viewed as elements in $T_X\overline{\mathcal{M}}$.

Let $\mathcal{M}$ and $\mathcal{N}$ be two manifolds. Suppose that $F : \mathcal{M} \to \mathcal{N}$ is a smooth mapping. Then the differential (or derivative) $\mathrm{D}F(X)$ of F at $X \in \mathcal{M}$ is defined as a mapping from $T_X\mathcal{M}$ to $T_{F(X)}\mathcal{N}$ such that

$$\mathrm{D}F(X)[\xi_X] \in T_{F(X)}\mathcal{N} \quad \forall \xi_X \in T_X\mathcal{M},$$

where the mapping $\mathrm{D}F(X)[\xi_X]$ from $\mathfrak{F}_{F(X)}(\mathcal{N})$ to $\mathbb{R}$ is defined by

$$\mathrm{D}F(X)[\xi_X] f = \xi_X(f \circ F) \quad \forall f \in \mathfrak{F}_{F(X)}(\mathcal{N}). \tag{2.2}$$

Let $\overline{F} : \overline{\mathcal{M}} \to \mathcal{N}$ be a smooth mapping between two Riemannian manifolds $\overline{\mathcal{M}}$ and $\mathcal{N}$. Suppose that F is the restriction of $\overline{F}$ to an embedded Riemannian submanifold $\mathcal{M}$. According to (2.1) and (2.2), it follows that

$$\begin{aligned} \mathrm{D}F(X)[\xi_X] f = \xi_X(f \circ F) &= \frac{d(f(F(c(t))))}{dt}\Big|_{t=0} = \frac{d(f(\overline{F}(c(t))))}{dt}\Big|_{t=0} \\ &= \xi_X(f \circ \overline{F}) = \mathrm{D}\overline{F}(X)[\xi_X] f \quad \forall f \in \mathfrak{F}_{F(X)}(\mathcal{N}), \end{aligned}$$

where $c : \mathbb{R} \to \mathcal{M}$ is a curve in $\mathcal{M}$ with $c(0) = X$ and $\dot{c}(0) = \xi_X$. Thus, we have

$$\mathrm{D}F(X)[\xi_X] = \mathrm{D}\overline{F}(X)[\xi_X] \quad \forall \xi_X \in T_X\mathcal{M},\ X \in \mathcal{M}. \tag{2.3}$$

Suppose that $(\mathcal{M}, g)$ is a Riemannian manifold and $f : \mathcal{M} \to \mathbb{R}$ is a smooth function. Then the Riemannian gradient $\operatorname{grad} f(X)$ of f at $X \in \mathcal{M}$ is defined as the unique element in $T_X\mathcal{M}$ such that

$$g_X(\operatorname{grad} f(X), \xi_X) = \mathrm{D}f(X)[\xi_X] \quad \forall \xi_X \in T_X\mathcal{M}.$$

Finally, we recall the definition of pullback. We need the concept of retraction, which was originally introduced in the field of algebraic topology [12]. Here, we use the definition of retraction in [4, 5, 30]. For a smooth mapping $F : \mathcal{M} \to \mathcal{N}$ between two manifolds $\mathcal{M}$ and $\mathcal{N}$ and a retraction R on $\mathcal{M}$, the pullback $\widehat{F}$ of F is a mapping from $T\mathcal{M}$ to $\mathcal{N}$ defined by

$$\widehat{F}(\xi) := F(R(\xi)) \quad \forall \xi \in T\mathcal{M}. \tag{2.4}$$

Let $\widehat{F}_X$ denote the restriction of $\widehat{F}$ to $T_X\mathcal{M}$, which is defined by

$$\widehat{F}_X(\xi_X) = F(R_X(\xi_X)) \quad \forall \xi_X \in T_X\mathcal{M}.$$

Then we have, by the definition of retraction [4],

$$\mathrm{D}F(X)[\xi_X] = \mathrm{D}F(R_X(0_X))[\xi_X] = \mathrm{D}F(R_X(0_X))[\mathrm{D}R_X(0_X)[\xi_X]] = \mathrm{D}\widehat{F}_X(0_X)[\xi_X]$$

for any $X \in \mathcal{M}$ and $\xi_X \in T_X\mathcal{M}$. Thus,

$$\mathrm{D}F(X) = \mathrm{D}\widehat{F}_X(0_X), \quad X \in \mathcal{M}. \tag{2.5}$$

3. Doubly stochastic inverse eigenvalue problem.

In this section, we propose a Riemannian Fletcher–Reeves conjugate gradient method for solving the DSIEP. The DSIEP is turned into a nonlinear least squares problem defined on some matrix manifolds. Then some basic geometric properties of these matrix manifolds are studied. Finally, a Riemannian Fletcher–Reeves conjugate gradient method is proposed for solving the nonlinear least squares problem.

3.1. Reformulation. Since the set of complex numbers $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ is closed under complex conjugation, without loss of generality, one may assume

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

where $a_j, b_j \in \mathbb{R}$ and $b_j \neq 0$ for $j = 1, \dots, s$. Define a block diagonal matrix Λ by

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

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

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

Define the set $\mathcal{U}$ by

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

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

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

Define by $\mathcal{Z}$ the set of all n-by-n doubly stochastic matrices, i.e.,

$$\mathcal{Z} := \{Z \odot Z \in \mathbb{R}^{n\times n} \mid Z \in \mathcal{OB}, Z^T \in \mathcal{OB}\},$$

where the set $\mathcal{OB}$ is the oblique manifold [3, 4]:

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

Then one may define a smooth manifold of isospectral matrices by

$$\mathcal{W}(\Lambda) := \{X \in \mathbb{R}^{n\times n} \mid X = Q(\Lambda + U)Q^T,\ Q \in \mathcal{O}(n), U \in \mathcal{U}\}.$$

The DSIEP has a solution if and only if $\mathcal{W}(\Lambda) \cap \mathcal{Z} \neq \emptyset$.

We assume that the DSIEP has at least one solution. Then the DSIEP aims to find a solution to the following nonlinear matrix equation:

$$H(Z,Q,U) := (H_1(Z,Q,U), H_2(Z)) = 0 \tag{3.1}$$

for $(Z,Q,U) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$. The mappings $H_1 : \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U} \to \mathbb{R}^{n\times n}$ and $H_2 : \mathcal{OB} \to \mathbb{R}^n$ are defined by

$$\begin{cases} H_1(Z,Q,U) = Z \odot Z - Q(\Lambda + U)Q^T, & (Z,Q,U) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}, \\ H_2(Z) = (Z \odot Z)^T \mathbf{e} - \mathbf{e}, & Z \in \mathcal{OB}, \end{cases}$$

where $\mathbf{e} \in \mathbb{R}^n$ is an n-vector of all ones.

It is easy to see that if $(\overline{Z}, \overline{Q}, \overline{U}) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ is a solution to $H(Z,Q,U) = 0$, then $\overline{C} = \overline{Z} \odot \overline{Z}$ is a solution to the DSIEP. Alternatively, one may solve the following nonlinear least squares problem:

$$\begin{array}{lll} \min & h(Z,Q,U) := \frac{1}{2}\|H(Z,Q,U)\|_F^2 \\ \text{subject to (s.t.)} & Z \in \mathcal{OB}, \quad Q \in \mathcal{O}(n), \quad U \in \mathcal{U}, \end{array} \tag{3.2}$$

where $\|\cdot\|_F$ is the matrix Frobenius norm.

3.2. Basic properties. In this section, we give some basic geometric properties of the product manifold $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$. We first note that the dimensions of $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{U}$ are given by [4, p. 27]

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

where $|\mathcal{J}|$ is the cardinality of the index subset $\mathcal{J} := \mathcal{N} \setminus \mathcal{I}$. Then we have

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

It follows that

$$\dim(\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}) > \dim(\mathbb{R}^{n\times n}) + \dim(\mathbb{R}^n) \quad \text{for } n \geq 4.$$

Hence, the nonlinear equation $H(Z,Q,U) = 0$ is an underdetermined system defined from the product manifold $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ to the Euclidean space $\mathbb{R}^{n\times n}$ for $n \geq 4$.

The tangent space of $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ at $(Z,Q,U) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ is given by

$$T_{(Z,Q,U)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U} = T_Z\mathcal{OB} \times T_Q\mathcal{O}(n) \times T_U\mathcal{U},$$

where $T_Z\mathcal{OB}$, $T_Q\mathcal{O}(n)$, and $T_U\mathcal{U}$ are the tangent spaces of $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{U}$ at $Z \in \mathcal{OB}$, $Q \in \mathcal{O}(n)$, and $U \in \mathcal{U}$, respectively ([3] and [4, p. 42]):

$$\begin{cases} T_Z\mathcal{OB} &= \{W \in \mathbb{R}^{n\times n} \mid \operatorname{diag}(ZW^T) = 0\}, \\ T_Q\mathcal{O}(n) &= \{QK \mid K^T = -K,\ K \in \mathbb{R}^{n\times n}\}, \\ T_U\mathcal{U} &= \mathcal{U}. \end{cases}$$

We now define a Riemannian metric on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$. 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

$$\langle (Z_1, Q_1, U_1), (Z_2, Q_2, U_2)\rangle := \operatorname{tr}(Z_1^T Z_2) + \operatorname{tr}(Q_1^T Q_2) + \operatorname{tr}(U_1^T U_2)$$

for all $(Z_1, Q_1, U_1), (Z_2, Q_2, U_2) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ and its induced norm $\|\cdot\|$. It is obvious that $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ is an embedded submanifold of the Euclidean space $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. Thus, $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ can be equipped with a induced Riemannian metric:

$$g_{(Z,Q,U)}((\xi_1, \zeta_1, \eta_1), (\xi_2, \zeta_2, \eta_2)) := \langle (\xi_1, \zeta_1, \eta_1), (\xi_2, \zeta_2, \eta_2)\rangle \tag{3.3}$$

for all $(Z, Q, U) \in \mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$ and $(\xi_1, \zeta_1, \eta_1), (\xi_2, \zeta_2, \eta_2) \in T_{(Z,Q,U)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$. Without causing any confusion, in what follows, let $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ stand for the Riemannian metric on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ and its induced norm. Therefore, $(\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}, g)$ is a Riemannian manifold. Then the orthogonal projections of any given points $\xi, \zeta, \eta \in \mathbb{R}^{n\times n}$ onto $T_Z\mathcal{OB}$, $T_Q\mathcal{O}(n)$, and $T_U\mathcal{U}$ are given, respectively, by ([3] and [4, p. 48])

$$\Pi_Z\xi = \xi - \operatorname{diag}(Z\xi^T)Z, \quad \Pi_Q\zeta = Q\operatorname{skew}(Q^T\zeta), \quad \Pi_U\eta = G \odot \eta, \tag{3.4}$$

where the matrix $G \in \mathbb{R}^{n\times n}$ is defined by

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

Hence, 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 $T_{(Z,Q,U)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ takes the form of

$$\Pi_{(Z,Q,U)}(\xi, \zeta, \eta) = (\Pi_Z\xi, \Pi_Q\zeta, \Pi_U\eta). \tag{3.5}$$

Next, we give a retraction R on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$. As in [3] and [4, p. 58], we can choose the retractions on $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{U}$ as follows:

$$\begin{cases} R_Z(\xi_Z) &= \Big(\operatorname{diag}((Z + \xi_Z)(Z + \xi_Z)^T)\Big)^{-1/2}(Z + \xi_Z) \quad \text{for } \xi_Z \in T_Z\mathcal{OB}, \\ R_Q(\zeta_Q) &= \operatorname{qf}(Q + \zeta_Q) \quad \text{for } \zeta_Q \in T_Q\mathcal{O}(n), \\ R_U(\eta_U) &= U + \eta_U \quad \text{for } \eta_U \in T_U\mathcal{U}. \end{cases}$$

Here, for a given nonsingular matrix $A \in \mathbb{R}^{n\times n}$, $\operatorname{qf}(A)$ means the Q factor of the QR decomposition of A in the form of $A = \widetilde{Q}\widetilde{R}$, where $\widetilde{Q} \in \mathcal{O}(n)$ and $\widetilde{R}$ is an upper triangular matrix with strictly positive diagonal elements. Therefore, a retraction R on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ can be defined by

$$R_{(Z,Q,U)}(\xi_Z, \xi_Q, \xi_U) = (R_Z(\xi_Z), R_Q(\xi_Q), R_U(\xi_U))$$

for all $(Z,Q,U) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ and $(\xi_Z, \xi_Q, \xi_U) \in T_{(Z,Q,U)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$.

We now define a vector transport on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$. We use the concept of vector transport in [4, p. 169], which is related to the concept of parallel translation (see for instance [4, p. 104]) and is easy to implement numerically. We see that $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{U}$ are embedded Riemannian submanifolds of $\mathbb{R}^{n\times n}$. By using orthogonal projections in (3.4), the vector transports on $\mathcal{OB}$, $\mathcal{O}(n)$, and $\mathcal{U}$ take the forms [4, p. 174]

$$\begin{cases} \mathcal{T}_{\eta_Z}\xi_Z = \Pi_{R_Z(\eta_Z)}\xi_Z &= \xi_Z - \mathrm{diag}(R_Z(\eta_Z)\xi_Z^T)R_Z(\eta_Z) \quad \text{for } \xi_Z, \eta_Z \in T_Z\mathcal{OB}, \\ \mathcal{T}_{\eta_Q}\xi_Q = \Pi_{R_Q(\eta_Q)}\xi_Q &= R_Q(\eta_Q)\mathrm{skew}\Big((R_Q(\eta_Q))^T\xi_Q\Big) \quad \text{for } \xi_Q, \eta_Q \in T_Q\mathcal{O}(n), \\ \mathcal{T}_{\eta_U}\xi_U = \Pi_{R_U(\eta_U)}\xi_U &= \xi_U \quad \text{for } \xi_U, \eta_U \in T_U\mathcal{U}. \end{cases}$$

Therefore, by (3.5), the vector transport on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ can be defined by

$$\mathcal{T}_{(\eta_Z,\eta_Q,\eta_U)}(\xi_Z, \xi_Q, \xi_U) := \Pi_{R_{(Z,Q,U)}(\eta_Z,\eta_Q,\eta_U)}(\xi_Z, \xi_Q, \xi_U) = (\mathcal{T}_{\eta_Z}\xi_Z, \mathcal{T}_{\eta_Q}\xi_Q, \mathcal{T}_{\eta_U}\xi_U)$$

for any $(\xi_Z, \xi_Q, \xi_U), (\eta_Z, \eta_Q, \eta_U) \in T_{(Z,Q,U)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$.

We now derive explicit expressions of the differential of H defined in (3.1) and the Riemannian gradient of the cost function h defined in (3.2). To do so, we define the mapping $\overline{H} : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} \times \mathbb{R}^n$ by

$$\overline{H}(Z,Q,U) = (\overline{H}_1(Z,Q,U), \overline{H}_2(Z)) := (Z \odot Z - Q(\Lambda + U)Q^T, (Z \odot Z)^T\mathbf{e} - \mathbf{e}) \tag{3.6}$$

for all $(Z,Q,U) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. Thus, H is the restriction of $\overline{H}$ onto $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$, i.e., $H = \overline{H}|_{\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}}$. For any given $(Z,Q,U) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ and $(\Delta Z, \Delta Q, \Delta U) \in T_{(Z,Q,U)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$, we get by (3.6),

$$\begin{aligned} & \overline{H}_1(Z + \Delta Z, Q + \Delta Q, U + \Delta U) - \overline{H}_1(Z,Q,U) \\ = \ & (Z + \Delta Z) \odot (Z + \Delta Z) - (Q + \Delta Q)(\Lambda + U + \Delta U)(Q + \Delta Q)^T \\ & \quad - (Z \odot Z - Q(\Lambda + U)Q^T) \\ = \ & 2Z \odot \Delta Z - Q(\Lambda + U)(\Delta Q)^T - \Delta Q(\Lambda + U)Q^T \\ & \quad - \Delta Q(\Lambda + U)(\Delta Q)^T - Q\Delta U Q^T + \Delta Z \odot \Delta Z - Q\Delta U(\Delta Q)^T \\ & \quad - \Delta Q \Delta U Q^T - \Delta Q \Delta U (\Delta Q)^T \\ = \ & 2Z \odot \Delta Z - Q(\Lambda + U)Q^T(\Delta Q Q^T)^T - (\Delta Q Q^T)Q(\Lambda + U)Q^T - Q\Delta U Q^T \\ & + \Delta Z \odot \Delta Z - \Delta Q(\Lambda + U)(\Delta Q)^T - Q\Delta U(\Delta Q)^T - \Delta Q\Delta U Q^T - \Delta Q \Delta U(\Delta Q)^T \\ = \ & 2Z \odot \Delta Z + [Q(\Lambda + U)Q^T, \Delta Q Q^T] - Q\Delta U Q^T \\ & + \Delta Z \odot \Delta Z - \Delta Q(\Lambda + U)(\Delta Q)^T - Q\Delta U(\Delta Q)^T - \Delta Q \Delta U Q^T - \Delta Q\Delta U(\Delta Q)^T \end{aligned}$$

and

$$\begin{aligned} & \overline{H}_2(Z + \Delta Z) - \overline{H}_2(Z) \\ = \ & ((Z + \Delta Z) \odot (Z + \Delta Z))^T\mathbf{e} - (Z \odot Z)^T\mathbf{e} \\ = \ & 2(Z \odot \Delta Z)^T\mathbf{e} + (\Delta Z \odot \Delta Z)^T\mathbf{e}, \end{aligned}$$

where the condition $(\Delta Q Q^T)^T = -\Delta Q Q^T$ is used. Hence, by using (2.3), the differential $\mathrm{D}H(Z,Q,U) : T_{(Z,Q,U)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U} \to T_{H(Z,Q,U)}\mathbb{R}^{n\times n} \times \mathbb{R}^n \simeq \mathbb{R}^{n\times n} \times \mathbb{R}^n$

of H at a point $(Z,Q,U)\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$ is determined by

$$\begin{aligned}&\mathrm{D}H(Z,Q,U)[(\Delta Z,\Delta Q,\Delta U)]=(\mathrm{D}H_1(Z,Q,U)[(\Delta Z,\Delta Q,\Delta U)],\mathrm{D}H_2(Z)[(\Delta Z])\\ =\;&(2Z\odot\Delta Z+[Q(\Lambda+U)Q^T,\Delta QQ^T]-Q\Delta UQ^T,2(Z\odot\Delta Z)^T\mathbf{e})\end{aligned}$$

for all $(\Delta Z,\Delta Q,\Delta U)\in T_{(Z,Q,U)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$. The adjoint $(\mathrm{D}H(Z,Q,U))^*:T_{H(Z,Q,U)}\mathbb{R}^{n\times n}\times\mathbb{R}^n\to T_{(Z,Q,U)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$ of the operator $\mathrm{D}H(Z,Q,U)$ is determined by

$$\begin{aligned}&\langle(\Delta W_1,\Delta W_2),\mathrm{D}H(Z,Q,U)[(\Delta Z,\Delta Q,\Delta U)]\rangle_F\\ =\;&g((\mathrm{D}H(Z,Q,U))^*[(\Delta W_1,\Delta W_2)],(\Delta Z,\Delta Q,\Delta U))\end{aligned}$$

for all $(\Delta Z,\Delta Q,\Delta U)\in T_{(Z,Q,U)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$ and $(\Delta W_1,\Delta W_2)\in T_{H(Z,Q,U)}\mathbb{R}^{n\times n}\times\mathbb{R}^n$, where $\langle\cdot,\cdot\rangle_F$ is the Frobenius inner product on $\mathbb{R}^{n\times n}\times\mathbb{R}^n$. Thus,

$$\begin{aligned}&(\mathrm{D}H(Z,Q,U))^*[\Delta W]\\ =\;&((\mathrm{D}H(Z,Q,U))_1^*[\Delta W],(\mathrm{D}H(Z,Q,U))_2^*[\Delta W],(\mathrm{D}H(Z,Q,U))_3^*[\Delta W])\end{aligned}\tag{3.7}$$

for all $\Delta W:=(\Delta W_1,\Delta W_2)\in T_{H(Z,Q,U)}\mathbb{R}^{n\times n}\times\mathbb{R}^n$, where for each $\Delta W:=(\Delta W_1,\Delta W_2)\in T_{H(Z,Q,U)}\mathbb{R}^{n\times n}\times\mathbb{R}^n$,

$$\begin{cases}(\mathrm{D}H(Z,Q,U))_1^*[\Delta W]&=\;2Z\odot(\Delta W_1+\mathbf{e}(\Delta W_2)^T)\\ &\quad-2\,\mathrm{diag}(Z(Z\odot(\Delta W_1+\mathbf{e}(\Delta W_2)^T))^T)Z,\\ (\mathrm{D}H(Z,Q,U))_2^*[\Delta W]&=\;\frac{1}{2}([Q(\Lambda+U)Q^T,(\Delta W_1)^T]+[Q(\Lambda+U)^TQ^T,\Delta W_1])Q,\\ (\mathrm{D}H(Z,Q,U))_3^*[\Delta W]&=\;-G\odot(Q^T\Delta W_1Q).\end{cases}$$

By using (3.7), the Riemannian gradient of h at a point $(Z,Q,U)\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$ is given by [4, p. 184]

$$\begin{aligned}&\mathrm{grad}\,h(Z,Q,U)=(\mathrm{D}H(Z,Q,U))^*[H(Z,Q,U)]\\ :=\;&((\mathrm{D}H)_1^*[H(Z,Q,U)],(\mathrm{D}H)_2^*[H(Z,Q,U)],(\mathrm{D}H)_3^*[H(Z,Q,U)]),\end{aligned}\tag{3.8}$$

where

$$\begin{cases}(\mathrm{D}H)_1^*[H(Z,Q,U)]&=\;2Z\odot(H_1(Z,Q,U)+\mathbf{e}H_2^T(Z,Q,U))\\ &\quad-2\,\mathrm{diag}(Z(Z\odot(H_1(Z,Q,U)+\mathbf{e}H_2^T(Z,Q,U)))^T)Z,\\ (\mathrm{D}H)_2^*[H(Z,Q,U)]&=\;\frac{1}{2}[Q(\Lambda+U)Q^T,H_1^T(Z,Q,U)]Q\\ &\quad+\frac{1}{2}[Q(\Lambda+U)^TQ^T,H_1(Z,Q,U)]Q,\\ (\mathrm{D}H)_3^*[H(Z,Q,U)]&=\;-G\odot(Q^TH_1(Z,Q,U)Q).\end{cases}$$

By using a similar idea in [9] and [10], a Riemannian gradient flow method for the cost function h over $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$ is given by

$$\frac{d(Z,Q,U)}{dt}=-\,\mathrm{grad}\,h(Z,Q,U).\tag{3.9}$$

Given a starting point $(Z(0),Q(0),U(0))\in\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$, this yields an initial value problem for the DSIEP (3.2). Then one may find a solution to the DSIEP

(3.2) by using existing ODE solvers for the differential equation (3.9). The numerical experiments in section 6 show that this extra isospectral flow method is effective for small- and medium-scale problems.

Finally, as in (2.4), the pullback mappings $\widehat{H} : T\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U} \to \mathbb{R}^{n\times n} \times \mathbb{R}^n$ and $\widehat{h} : T\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U} \to \mathbb{R}$ of H and h through the retraction R on $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ are given, respectively, by

$$\widehat{H}(\xi,\zeta,\eta) = H(R(\xi,\zeta,\eta)) \quad \text{and} \quad \widehat{h}(\xi,\zeta,\eta) = h(R(\xi,\zeta,\eta)) \tag{3.10}$$

for all $(\xi,\zeta,\eta) \in T\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$, where $\widehat{H}_{(Z,Q,U)} = \widehat{H}|_{T_{(Z,Q,U)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}}$ and $\widehat{h}_{(Z,Q,U)} = \widehat{h}|_{T_{(Z,Q,U)}\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}}$. Moreover, we have by (3.8) [4, p. 56]

$$\operatorname{grad} h(Z,Q,U) = \operatorname{grad} \widehat{h}_{(Z,Q,U)}(0_{(Z,Q,U)}) = (\mathrm{D}\widehat{H}_{(Z,Q,U)}(0_{(Z,Q,U)}))^*[\widehat{H}_{(Z,Q,U)}(0_{(Z,Q,U)})], \tag{3.11}$$

where $0_{(Z,Q,U)}$ is the origin of $T_{(Z,Q,U)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$.

3.3. A Riemannian Fletcher–Reeves conjugate gradient method. In this section, we present a Riemannian Fletcher–Reeves conjugate gradient method for solving the DSIEP (3.2). This is motivated by a modified Fletcher–Reeves method with Armijo-type line search in [36], which always provides a descent direction of the objective function, while the direction generated by a standard Fletcher–Reeves method with Armijo-type line search or Wolfe-type line search is not necessarily one of descent (see [36]). The corresponding geometric algorithm is described as follows.

ALGORITHM 3.1 (A Riemannian Fletcher–Reeves conjugate gradient method).

Step 0. *Given* $(Z^0,Q^0,U^0) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$, $\alpha \geq 1$, $\rho,\delta_1 \in (0,1)$, $\delta_2 > 0$. $k := 0$.

Step 1. *If* $h(Z^k,Q^k,U^k) = 0$, *then stop. Otherwise, go to* **Step 2**.

Step 2. *Set*

$$\begin{aligned} &(\Delta Z_k, \Delta Q_k, \Delta U_k) \\ := \; & -\operatorname{grad} h(Z^k,Q^k,U^k) + \beta_k \Delta Y_k - \theta_k \operatorname{grad} h(Z^k,Q^k,U^k), \end{aligned} \tag{3.12}$$

where

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

$$\beta_k := \frac{\|\operatorname{grad} h(Z^k,Q^k,U^k)\|^2}{\|\operatorname{grad} h(Z^{k-1},Q^{k-1},U^{k-1})\|^2}, \tag{3.13}$$

$$\theta_k := \frac{\langle \operatorname{grad} h(Z^k,Q^k,U^k), \Delta Y_k\rangle}{\|\operatorname{grad} h(Z^{k-1},Q^{k-1},U^{k-1})\|^2}. \tag{3.14}$$

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

$$\begin{aligned} & h(R_{(Z^k,Q^k,U^k)}(\alpha_k(\Delta Z_k,\Delta Q_k,\Delta U_k))) - h(Z^k,Q^k,U^k) \\ \leq \; & \delta_1\alpha_k\langle \operatorname{grad} h(Z^k,Q^k,U^k), (\Delta Z_k,\Delta Q_k,\Delta U_k)\rangle \\ & -\, \delta_2\alpha_k^2\|(\Delta Z_k,\Delta Q_k,\Delta U_k)\|^2. \end{aligned} \tag{3.15}$$

Set

$$(Z^{k+1},Q^{k+1},U^{k+1}) := R_{(Z^k,Q^k,U^k)}(\alpha_k(\Delta Z_k,\Delta Q_k,\Delta U_k)). \tag{3.16}$$

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

In the following, we have several remarks on the above algorithm.

- In Algorithm 3.1, we set

$$(\Delta Z_0, \Delta Q_0, \Delta U_0) := -\operatorname{grad} h(Z^0, Q^0, U^0).$$

- From (3.12), (3.13), and (3.14), we get for all $k \geq 1$,

$$\langle (\Delta Z_k, \Delta Q_k, \Delta U_k), \operatorname{grad} h(Z^k, Q^k, U^k)\rangle = -\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^2. \tag{3.17}$$

  This shows that $(\Delta Z_k, \Delta Q_k, \Delta U_k)$ is a descent direction of h.

- In Algorithm 3.1, we observe that for all $k \geq 1$,

$$\|\Delta Y_k\| \leq \|(\Delta Z_{k-1}, \Delta Q_{k-1}, \Delta U_{k-1})\|. \tag{3.18}$$

- By (3.14) and (3.18), we have that for all $k \geq 1$,

$$\begin{aligned} \theta_k &\leq \frac{\|\operatorname{grad} h(Z^k, Q^k, U^k)\| \cdot \|\Delta Y_k\|}{\|\operatorname{grad} h(Z^{k-1}, Q^{k-1}, U^{k-1})\|^2} \\ &\leq \frac{\|\operatorname{grad} h(Z^k, Q^k, U^k)\| \cdot \|(\Delta Z_{k-1}, \Delta Q_{k-1}, \Delta U_{k-1})\|}{\|\operatorname{grad} h(Z^{k-1}, Q^{k-1}, U^{k-1})\|^2}. \end{aligned}$$

- In Step 3 of Algorithm 3.1, the initial step length is set to be α. As mentioned in [35], the line-search process may not be very efficient. As in [38], one can derive a reasonable initial step length:

$$t_k = \left| \frac{\langle \operatorname{grad} h(Z^k, Q^k, U^k), (\Delta Z_k, \Delta Q_k, \Delta U_k)\rangle}{\|\mathrm{D}H(Z^k, Q^k, U^k)[(\Delta Z_k, \Delta Q_k, \Delta U_k)]\|^2} \right|. \tag{3.19}$$

  Sparked by [35], the line-search step (i.e., Step 3) of Algorithm 3.1 can be modified such that if

$$\|\mathrm{D}H(Z^k, Q^k, U^k)[(\Delta Z_k, \Delta Q_k, \Delta U_k)]\| > 0$$

  and

$$\begin{aligned} &h(R_{(Z^k,Q^k,U^k)}(t_k(\Delta Z_k, \Delta Q_k, \Delta U_k))) - h(Z^k, Q^k, U^k) \\ &\leq \delta_1 t_k \langle \operatorname{grad} h(Z^k, Q^k, U^k), (\Delta Z_k, \Delta Q_k, \Delta U_k)\rangle - \delta_2 t_k^2 \|(\Delta Z_k, \Delta Q_k, \Delta U_k)\|^2, \end{aligned}$$

  then we set $\alpha_k = t_k$; otherwise, the step length α_k can be selected by Step 3 of Algorithm 3.1. The numerical tests in section 6 show that the initial step length (3.19) is very effective.

Finally, we point out that Algorithm 3.1 has some advantages over the constructive methods in [23, 21, 32, 39, 27, 28]. A least squares approach is proposed for solving the general DSIEP, where there is no additional sufficient condition imposed on the prescribed eigenvalues. As shown in section 5, our method can be used to construct a doubly stochastic matrix with both prescribed entries and a spectrum. Even if the DSIEP is not solvable, our method generates a least squares solution.

4. Global convergence. In this section, we establish the global convergence of Algorithm 3.1. First, we can easily derive the following result. We omit the proof here.

LEMMA 4.1. *For any given point $(Z^0, Q^0, U^0) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$, the level set*

$$\Omega := \{(Z, Q, U) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U} \mid h(Z, Q, U) \leq h(Z^0, Q^0, U^0)\}$$

is compact.

Remark 4.2. We see that $\mathcal{OB}\times\mathcal{O}(n)\times\mathcal{U}$ 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 $T_{(Z,Q,U)}\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U} \subset \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. Thus, the Riemannian gradient $\operatorname{grad} h$ defined in (3.8) can be viewed as a continuous nonlinear mapping between $\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ and $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. By Lemma 4.1, Ω is compact. Then there exists a constant $\gamma > 0$ such that [4, pp. 151–152]

$$\| \operatorname{grad} h(Z, Q, U)\| \leq \gamma \quad \forall (Z, Q, U) \in \Omega.$$

Moreover, for any points $(Z_1, Q_1, U_1), (Z_2, Q_2, U_2) \in \mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$, the operation $\operatorname{grad} h(Z_2, Q_2, U_2) - \operatorname{grad} h(Z_1, Q_1, U_1)$ is meaningful since both $\operatorname{grad} h(Z_1, Q_1, U_1)$ and $\operatorname{grad} h(Z_2, Q_2, U_2)$ can be treated as vectors in $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$. By Lemma 4.1, Ω is compact. Then $\operatorname{grad} h$ is Lipschitz continuous on Ω, i.e., there exists a constant $\beta_{L_1} > 0$ such that

$$\| \operatorname{grad} h(Z_2, Q_2, U_2) - \operatorname{grad} h(Z_1, Q_1, U_1)\| \leq \beta_{L_1} \operatorname{dist}((Z_1, Q_1, U_1), (Z_2, Q_2, U_2))$$

for all $(Z_1, Q_1, U_1), (Z_2, Q_2, U_2) \in \Omega$.

LEMMA 4.3. *There exists a constant $\nu > 0$ such that for all k sufficiently large,*

$$\alpha_k \geq \nu \frac{\| \operatorname{grad} h(Z^k, Q^k, U^k)\|^2}{\|(\Delta Z_k, \Delta Q_k, \Delta U_k)\|^2}. \tag{4.1}$$

Proof. It follows from (3.15), (3.16), and (3.17) that the sequence $\{h(Z^k, Q^k, U^k)\}$ is decreasing and bounded below, and is thus convergent. Hence, we have

$$\sum_{k=0}^{\infty}(-\delta_1\alpha_k\langle \operatorname{grad}\ h(Z^k, Q^k, U^k), (\Delta Z_k, \Delta Q_k, \Delta U_k)\rangle + \delta_2\alpha_k^2\|(\Delta Z_k, \Delta Q_k, \Delta U_k)\|^2) < \infty.$$

This, together with (3.17), leads to

$$\sum_{k=0}^{\infty} \alpha_k^2\|(\Delta Z_k, \Delta Q_k, \Delta U_k)\|^2 < \infty \tag{4.2}$$

and

$$\begin{aligned} &-\sum_{k=0}^{\infty} \alpha_k \langle \operatorname{grad} h(Z^k, Q^k, U^k), (\Delta Z_k, \Delta Q_k, \Delta U_k)\rangle \\ &= \sum_{k=0}^{\infty} \alpha_k \| \operatorname{grad} h(Z^k, Q^k, U^k)\|^2 < \infty. \end{aligned}$$

Then

$$\lim_{k\to\infty} \alpha_k \| \operatorname{grad} h(Z^k, Q^k, U^k)\|^2 = 0 \quad \text{and} \quad \lim_{k\to\infty} \alpha_k \|(\Delta Z_k, \Delta Q_k, \Delta U_k)\| = 0. \tag{4.3}$$

Next, we show (4.1). By (3.17), one has, for all $k \geq 1$,

$$\| \operatorname{grad} h(Z^k, Q^k, U^k)\| \leq \|(\Delta Z_k, \Delta Q_k, \Delta U_k)\|.$$

Hence, if $\alpha_k \geq 1$ for all k, then (4.1) holds for $\nu = 1$.

We now suppose that $\alpha_k < 1$ for all k sufficiently large. According to Step 3 of Algorithm 3.1, one has, for all k sufficiently large,

$$\begin{aligned} &h\big(R_{(Z^k,Q^k,U^k)}(\rho^{-1}\alpha_k(\Delta Z_k,\Delta Q_k,\Delta U_k))\big) - h\big(Z^k,Q^k,U^k\big) \\ &\quad > \delta_1\rho^{-1}\alpha_k\big\langle \operatorname{grad} h(Z^k,Q^k,U^k),(\Delta Z_k,\Delta Q_k,\Delta U_k)\big\rangle \\ &\qquad - \delta_2\rho^{-2}\alpha_k^2\|(\Delta Z_k,\Delta Q_k,\Delta U_k)\|^2. \end{aligned} \tag{4.4}$$

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 $\kappa > 0$ and $\beta_{L_2} > 0$ such that

$$\big\| \operatorname{grad} \widehat{h}_X(\eta_X) - \operatorname{grad} \widehat{h}_X(\xi_X)\big\| \le \beta_{L_2}\|\eta_X - \xi_X\| \tag{4.5}$$

for any $X \in \Omega$ and $\xi_X, \eta_X \in T_X\mathcal{OB} \times \mathcal{O}(n) \times \mathcal{U}$ with $\|\eta_X\|, \|\xi_X\| \le \kappa$. Also, we have by (4.3), for all k sufficiently large,

$$\alpha_k\|(\Delta Z_k,\Delta Q_k,\Delta U_k)\| \le \kappa. \tag{4.6}$$

By the mean-value theorem, (3.11), (3.17), (4.5), and (4.6), there is a $\omega_k \in (0,1)$ such that for all k sufficiently large,

$$\begin{aligned} &h\big(R_{(Z^k,Q^k,U^k)}(\rho^{-1}\alpha_k(\Delta Z_k,\Delta Q_k,\Delta U_k))\big) - h(Z^k,Q^k,U^k) \\ &= h\big(R_{(Z^k,Q^k,U^k)}(\rho^{-1}\alpha_k(\Delta Z_k,\Delta Q_k,\Delta U_k))\big) - h\big(R_{(Z^k,Q^k,U^k)}(0_{(Z^k,Q^k,U^k)})\big) \\ &= \widehat{h}_{(Z^k,Q^k,U^k)}\big(\rho^{-1}\alpha_k(\Delta Z_k,\Delta Q_k,\Delta U_k)\big) - \widehat{h}_{(Z^k,Q^k,U^k)}\big(0_{(Z^k,Q^k,U^k)}\big) \\ &= \rho^{-1}\alpha_k\big\langle \operatorname{grad}\widehat{h}_{(Z^k,Q^k,U^k)}\big(\omega_k\rho^{-1}\alpha_k(\Delta Z_k,\Delta Q_k,\Delta U_k)\big),(\Delta Z_k,\Delta Q_k,\Delta U_k)\big\rangle \\ &= \rho^{-1}\alpha_k\big\langle \operatorname{grad}\widehat{h}_{(Z^k,Q^k,U^k)}(0_{(Z^k,Q^k,U^k)}),(\Delta Z_k,\Delta Q_k,\Delta U_k)\big\rangle \\ &\quad + \rho^{-1}\alpha_k\big\langle(\Delta Z_k,\Delta Q_k,\Delta U_k),\operatorname{grad}\widehat{h}_{(Z^k,Q^k,U^k)}\big(\omega_k\rho^{-1}\alpha_k(\Delta Z_k,\Delta Q_k,\Delta U_k)\big)\big\rangle \\ &\quad - \rho^{-1}\alpha_k\big\langle(\Delta Z_k,\Delta Q_k,\Delta U_k),\operatorname{grad}\widehat{h}_{(Z^k,Q^k,U^k)}\big(0_{(Z^k,Q^k,U^k)}\big)\big\rangle \\ &\le \rho^{-1}\alpha_k\big\langle \operatorname{grad}\widehat{h}_{(Z^k,Q^k,U^k)}(0_{(Z^k,Q^k,U^k)}),(\Delta Z_k,\Delta Q_k,\Delta U_k)\big\rangle \\ &\quad + \beta_{L_2}\omega_k\rho^{-2}\alpha_k^2\|(\Delta Z_k,\Delta Q_k,\Delta U_k)\|^2 \\ &\le -\rho^{-1}\alpha_k\|\operatorname{grad} h(Z^k,Q^k,U^k)\|^2 + \beta_{L_2}\rho^{-2}\alpha_k^2\|(\Delta Z_k,\Delta Q_k,\Delta U_k)\|^2. \end{aligned} \tag{4.7}$$

Combining (3.17) and (4.4) with (4.7), we have for all k sufficiently large,

$$\alpha_k > \frac{(1-\delta_1)\rho\|\operatorname{grad} h(Z^k,Q^k,U^k)\|^2}{(\delta_2+\beta_{L_2})\|(\Delta Z_k,\Delta Q_k,\Delta U_k)\|^2}.$$

Then we get (4.1) by setting $\nu = \min\{1, ((1-\delta_1)\rho)/(\delta_2+\beta_{L_2})\}$. ∎

By (4.1) and (4.2), we have the following Riemannian analogy of the Zoutendijk condition (see [26] for more details).

LEMMA 4.4. *One has*

$$\sum_{k=0}^{\infty}\frac{\|\operatorname{grad} h(Z^k,Q^k,U^k)\|^4}{\|(\Delta Z_k,\Delta Q_k,\Delta U_k)\|^2} < \infty. \tag{4.8}$$

We now establish the global convergence of Algorithm 3.1. The proof can be viewed as a generalization of [36, Theorem 3.3].

THEOREM 4.5. *Let $\{(Z^k, Q^k, U^k)\}$ be the sequence generated by Algorithm* 3.1. *Then we have*

$$\liminf_{k\to\infty} \|\operatorname{grad} h(Z^k, Q^k, U^k)\| = 0.$$

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

$$\|\operatorname{grad} h(Z^k, Q^k, U^k)\| \geq \epsilon \quad \forall k. \tag{4.9}$$

We have by (3.12), (3.13), (3.14), and (3.18), for all $k \geq 1$,

$$\begin{aligned} &\|(\Delta Z_k, \Delta Q_k, \Delta U_k)\|^2 \\ &= \beta_k^2 \|\Delta Y_k\|^2 - (\theta_k^2 - 1)\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^2 \\ &\leq \beta_k^2 \|(\Delta Z_{k-1}, \Delta Q_{k-1}, \Delta U_{k-1})\|^2 - (\theta_k^2 - 1)\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^2. \end{aligned} \tag{4.10}$$

We get by (3.13), (4.9), and (4.10),

$$\begin{aligned} &\frac{\|\Delta Z_k, \Delta Q_k, \Delta U_k)\|^2}{\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^4} \\ &\leq \frac{\beta_k^2 \|(\Delta Z_{k-1}, \Delta Q_{k-1}, \Delta U_{k-1})\|^2}{\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^4} - \frac{\theta_k^2 - 1}{\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^2} \\ &\leq \frac{\|(\Delta Z_{k-1}, \Delta Q_{k-1}, \Delta U_{k-1})\|^2}{\|\operatorname{grad} h(Z^{k-1}, Q^{k-1}, U^{k-1})\|^4} - \frac{\theta_k^2}{\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^2} + \frac{1}{\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^2} \\ &\leq \frac{\|(\Delta Z_{k-1}, \Delta Q_{k-1}, \Delta U_{k-1})\|^2}{\|\operatorname{grad} h(Z^{k-1}, Q^{k-1}, U^{k-1})\|^4} + \frac{1}{\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^2} \\ &\leq \sum_{j=0}^{k} \frac{1}{\|\operatorname{grad} h(Z^j, Q^j, U^j)\|^2} \leq \frac{k+1}{\epsilon^2}, \end{aligned}$$

where the last inequality implies

$$\sum_{k=0}^{\infty} \frac{\|\operatorname{grad} h(Z^k, Q^k, U^k)\|^4}{\|(\Delta Z_k, \Delta Q_k, \Delta U_k)\|^2} \geq \epsilon^2 \sum_{k=0}^{\infty} \frac{1}{k+1} = \infty.$$

This contradicts the Riemannian Zoutendijk condition (4.8). This completes the proof. ◻

5. DSIEP with prescribed entries. In this section, we consider the DSIEP with prescribed entries (DSIEP-PE). In many applications, the underlying structure of a desired stochastic matrix is often characterized by the prescribed entries at arbitrary locations. The DSIEP-PE can be stated as follows.

DSIEP-PE. *Given a self-conjugate set of complex numbers $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, find a doubly stochastic matrix $C \in \mathbb{R}^{n\times n}$ such that its eigenvalues are $\lambda_1, \lambda_2, \ldots, \lambda_n$ exactly and*

$$C_{ij} = (C_a)_{ij} \ \ \forall\, (i, j) \in \mathcal{L},$$

where $C_a \in \mathbb{R}^{n\times n}$ is a prescribed nonnegative matrix and $\mathcal{L} \subset \mathcal{N}$ is a given index subset.

As in section 3, let $\Lambda \in \mathbb{R}^{n\times n}$ be a block diagonal matrix with the given eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Define the matrix $\widehat{G} \in \mathbb{R}^{n\times n}$ by

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

Then we define a nonnegative matrix $\widehat{C}_a$ and a diagonal matrix $\widehat{I}_n$ by

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

where $\mathrm{Diag}(\mathbf{a})$ is a diagonal matrix with the vector $\mathbf{a}$ as the diagonal. In addition, we assume that the given index subset $\mathcal{L}$ is such that $\sum_{j=1}^{n} (\widehat{C}_a)_{ij} < 1$ for $i = 1, \ldots, n$. In this case, the diagonal matrix $\widehat{I}_n$ is nonsingular. Thus, the DSIEP-PE aims to solve the nonlinear matrix equation

$$\Phi(Z, Q, U) := (\Phi_1(Z, Q, U), \Phi_2(Z)) = 0 \tag{5.1}$$

for $(Z, Q, U) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$, where the set $\widehat{\mathcal{OB}}$ is a manifold defined by

$$\widehat{\mathcal{OB}} := \{Z \in \mathbb{R}^{n\times n} \mid \mathrm{diag}(ZZ^T) = \widehat{I}_n,\ \widehat{G} \odot Z = 0\},$$

and the mappings $\Phi_1 : \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U} \to \mathbb{R}^{n\times n}$ and $\Phi_2 : \widehat{\mathcal{OB}} \to \mathbb{R}^n$ are defined by

$$\begin{cases} \Phi_1(Z, Q, U) := \widehat{C}_a + Z \odot Z - Q(\Lambda + U)Q^T, & (Z, Q, U) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}, \\ \Phi_2(Z) := (\widehat{C}_a + Z \odot Z)^T \mathbf{e} - \mathbf{e}, & Z \in \widehat{\mathcal{OB}}. \end{cases}$$

Notice that the dimension of $\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$ is given by

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

We see that the nonlinear equation $\Phi(Z, Q, U) = 0$ is an underdetermined system defined from the product manifold $\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$ to the Euclidean space $\mathbb{R}^{n\times n} \times \mathbb{R}^n$ if n is large and the number $|\mathcal{L}|$ of prescribed entries is not large.

If we find a solution $(\overline{Z}, \overline{Q}, \overline{U}) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$ to $\Phi(Z, Q, U) = 0$, then $\overline{C} = \widehat{G} \odot C_a + \overline{Z} \odot \overline{Z}$ is a solution to the DSIEP-PE. Alternatively, one may solve the DSIEP-PE by finding a global solution to the following nonlinear least squares problem:

$$\begin{array}{ll} \min & \phi(Z, Q, U) := \frac{1}{2}\|\Phi(Z, Q, U)\|_F^2 \\ \text{s.t.} & Z \in \widehat{\mathcal{OB}}, \quad Q \in \mathcal{O}(n), \quad U \in \mathcal{U}. \end{array} \tag{5.2}$$

Let $\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$ be equipped with the Riemannian metric defined as in (3.3). It follows that the tangent space of $\widehat{\mathcal{OB}}$ at a point $Z \in \widehat{\mathcal{OB}}$ is given by

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

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

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

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

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

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

$$\widehat{\mathcal{T}}_{\eta_Z}(\xi_Z) := \Pi_{\widehat{R}_Z(\eta_Z)}(\xi_Z) = \xi_Z - \Big(\operatorname{diag}((Z+\eta_Z)(Z+\eta_Z)^T)\Big)^{-1}\operatorname{diag}((Z+\eta_Z)\xi_Z^T)(Z+\eta_Z)$$

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

Next, we establish explicit formulas for the differential of the smooth mapping Φ defined in (5.1) and the Riemannian gradient of the cost function ϕ defined in Problem (5.2). As in section 3, the differential

$$\mathrm{D}\Phi(Z,Q,U) : T_{(Z,Q,U)}\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U} \to T_{\Phi(Z,Q,U)}\mathbb{R}^{n\times n} \times \mathbb{R}^n \simeq \mathbb{R}^{n\times n} \times \mathbb{R}^n$$

of Φ at a point $(Z,Q,U) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$ is determined by

$$\begin{aligned} & \mathrm{D}\Phi(Z,Q,U)[(\Delta Z,\Delta Q,\Delta U)] = (\mathrm{D}\Phi_1(Z,Q,U)[(\Delta Z,\Delta Q,\Delta U)], \mathrm{D}\Phi_2(Z)[\Delta Z]) \\ = \quad & (2Z\odot\Delta Z + [Q(\Lambda+U)Q^T, \Delta Q Q^T] - Q\Delta U Q^T, 2(Z\odot\Delta Z)^T\mathbf{e}) \end{aligned}$$

for all $(\Delta Z,\Delta Q,\Delta U) \in T_{(Z,Q,U)}\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$. The adjoint

$$(\mathrm{D}\Phi(Z,Q,U))^* : T_{\Phi(Z,Q,U)}\mathbb{R}^{n\times n} \times \mathbb{R}^n \to T_{(Z,Q,U)}\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$$

of $\mathrm{D}\Phi(Z,Q,U)$ is determined by

$$(\mathrm{D}\Phi(Z,Q,U))^*[\Delta W] = ((\mathrm{D}\Phi)_1^*[\Delta W], (\mathrm{D}\Phi)_2^*[\Delta W], (\mathrm{D}\Phi)_3^*[\Delta W]) \tag{5.3}$$

for all

$$\Delta W := (\Delta W_1, \Delta W_2) \in T_{\Phi(Z,Q,U)}\mathbb{R}^{n\times n} \times \mathbb{R}^n,$$

where for each $\Delta W := (\Delta W_1, \Delta W_2) \in T_{\Phi(Z,Q,U)}\mathbb{R}^{n\times n} \times \mathbb{R}^n$,

$$\begin{cases} (\mathrm{D}\Phi)_1^*[\Delta W] &= 2Z\odot(\Delta W_1 + \mathbf{e}(\Delta W_2)^T) \\ & \quad - 2\widehat{I}_n^{-1}\operatorname{diag}(Z(Z\odot(\Delta W_1+\mathbf{e}(\Delta W_2)^T))^T)Z, \\ (\mathrm{D}\Phi)_2^*[\Delta W] &= \dfrac{1}{2}([Q(\Lambda+U)Q^T, (\Delta W_1)^T] + [Q(\Lambda+U)^TQ^T, \Delta W_1])Q, \\ (\mathrm{D}\Phi)_3^*[\Delta W] &= -G\odot(Q^T\Delta W_1 Q). \end{cases}$$

Then the Riemannian gradient of ϕ at a point $(Z,Q,U) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$ is given by [4, p. 184]

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

where $(\mathrm{D}\Phi(Z,Q,U))^*[\cdot]$ is given by (5.3).

Similarly, a Riemannian gradient flow method for the cost function ϕ over the product manifold $\widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$ is given by

$$\frac{d(Z,Q,U)}{dt} = -\operatorname{grad}\phi(Z,Q,U), \tag{5.4}$$

which is an initial value problem if given an initial value $(Z(0), Q(0), U(0)) \in \widehat{\mathcal{OB}} \times \mathcal{O}(n) \times \mathcal{U}$. Then one may get a solution to the DSIEP-PE (5.2) by using existing ODE solvers for the above differential equation.

As in section 3.3, one may develop a Riemannian Fletcher–Reeves conjugate gradient algorithm for solving the DSIEP-PE, i.e., one may solve the DSIEP-PE (5.2) by using Algorithm 3.1. The corresponding global convergence can also be established.

6. Numerical tests. In this section, we report the numerical performance of Algorithm 3.1 for solving the DSIEP (3.2) and the DSIEP-PE (5.2). All numerical tests are carried out by using MATLAB 7.1 running on a workstation with an Intel Xeon CPU E5-2687W at 3.10 GHz and 32 GB of RAM. To generate a random doubly stochastic matrix, we may use Birkhoff's theorem [7], which says that any n-by-n doubly stochastic matrix A can be written as a convex combination of m permutation matrices for $m \leq (n-1)^2 + 1$.

For Algorithm 3.1, the starting points are generated randomly by the built-in functions `rand` and `schur`: for the DSIEP,

$$(6.1)\qquad \begin{cases} \widehat{Z} \odot \widehat{Z} = \texttt{rand}\,(n,n),\ Z^0 = (\mathrm{diag}(\widehat{Z}\widehat{Z}^T))^{-\frac{1}{2}}\widehat{Z} = Z(0) \in \mathcal{OB}, \\ [Q^0, \widehat{U}] = \texttt{schur}\,(Z^0 \odot Z^0, '\mathrm{real}') = [Q(0), \widehat{U}], \quad U^0 = G \odot \widehat{U} = U(0), \end{cases}$$

and for the DSIEP-PE,

$$(6.2)\qquad \begin{cases} \widehat{Z} \odot \widehat{Z} = \texttt{rand}\,(n,n),\ Z^0 = \widehat{I}_n^{\frac{1}{2}}(\mathrm{diag}((\widehat{W} \odot \widehat{Z})(\widehat{W} \odot \widehat{Z})^T))^{-\frac{1}{2}}(\widehat{W} \odot \widehat{Z}) = Z(0) \in \widehat{\mathcal{OB}}, \\ [Q^0, \widehat{U}] = \texttt{schur}\,(\widehat{C}_a + Z^0 \odot Z^0, '\mathrm{real}') = [Q(0), \widehat{U}], \quad U^0 = G \odot \widehat{U} = U(0), \end{cases}$$

where $\widehat{W} := E - \widehat{G}$, with E being an n-by-n matrix of all ones.

In our numerical tests, the stopping criterion for Algorithm 3.1 is set to be

$$\|H(Z^k, Q^k, U^k)\|_F \leq 10^{-12} \quad \text{or} \quad \|\Phi(Z^k, Q^k, U^k)\|_F \leq 10^{-12}.$$

As in [36], we also set $\alpha = 1.4$, $\rho = 0.5$, $\delta_1 = 10^{-3}$, and $\delta_2 = 10^{-8}$.

In our tests, "`CT.`", "`IT.`", "`NF.`", "`Err.`", and "`Res.`" mean the total computing time, the number of iterations, the number of function evaluations, the error $\|H(Z^k, Q^k, U^k)\|_F$ or $\|\Phi(Z^k, Q^k, U^k)\|_F$, and the residual $\|\operatorname{grad} h(Z^k, Q^k, U^k)\|$ or $\|\operatorname{grad} \phi(Z^k, Q^k, U^k)\|$ at the final iterate of the corresponding algorithm accordingly.

We consider two examples with different problem sizes n.

Example 6.1. We consider the DSIEP with varying n. Let $\widetilde{C} = \sum_{j=1}^n c_j P_j$ be a random $n \times n$ doubly stochastic matrix where $\{c_j \geq 0\}$ are generated randomly such that $\sum_{j=1}^n c_j = 1$ and $\{P_j\}$ are n random permutation matrices. We choose the eigenvalues of $\widehat{C}$ as the prescribed spectra.

Example 6.2. We consider the DSIEP-PE with varying n. Let $\widetilde{C} = \sum_{j=1}^n c_j P_j$ be a random $n \times n$ doubly stochastic matrix where $\{c_j \geq 0\}$ are generated randomly such that $\sum_{j=1}^n c_j = 1$ and $\{P_j\}$ are n random permutation matrices. We choose the eigenvalues of $\widehat{C}$ as the prescribed spectra. Also, we choose the index subset $\mathcal{L} := \{(i,j) \mid 0.02 \leq (\widehat{C})_{ij} \leq 0.03,\ i,j = 1, \ldots, n\}$. The prescribed nonnegative matrix $C_a \in \mathbb{R}^{n \times n}$ is such that $(C_a)_{ij} = \widehat{C}_{ij}$ if $(i,j) \in \mathcal{L}$ and $(C_a)_{ij} = 0$ otherwise.

For demonstration purposes, in Tables 1 and 2 we report the numerical results for Examples 6.1 and 6.2 with different problem sizes n, where the initial step length guess (3.19) may be used in Algorithm 3.1.

TABLE 1
Numerical results for Example 6.1.

Alg.	n	CT.	IT.	NF.	Err.	Res.
Alg. 3.1	100	2.2 s	742	747	9.6×10^{-13}	1.6×10^{-13}
	200	14 s	1454	1460	9.9×10^{-13}	1.2×10^{-13}
	500	02 m 22 s	1548	1564	9.8×10^{-13}	1.3×10^{-13}
	800	23 m 17 s	4848	4855	9.9×10^{-13}	9.5×10^{-14}
	1000	46 m 13 s	5327	5334	9.9×10^{-13}	1.3×10^{-13}
	1500	03 h 46 m 13 s	3820	21211	9.9×10^{-13}	6.3×10^{-13}
	2000	07 h 47 m 51 s	8507	9110	9.9×10^{-13}	7.5×10^{-13}
	3000	≥ 15 h	*	*	*	*
Alg. 3.1 with (3.19)	100	1.0 s	278	281	9.8×10^{-13}	5.7×10^{-13}
	200	4.4 s	346	349	8.6×10^{-13}	5.8×10^{-13}
	500	01 m 00 s	419	422	9.3×10^{-13}	5.4×10^{-13}
	800	03 m 36 s	461	464	9.3×10^{-13}	3.5×10^{-13}
	1000	06 m 44 s	477	480	9.3×10^{-13}	5.3×10^{-13}
	1500	20 m 50 s	521	524	9.8×10^{-13}	7.4×10^{-13}
	2000	58 m 37 s	558	561	9.8×10^{-13}	5.1×10^{-13}
	3000	03 h 17 m 12 s	629	632	9.9×10^{-13}	7.6×10^{-13}

TABLE 2
Numerical results for Example 6.2.

Alg.	n	CT.	IT.	NF.	Err.	Res.
Alg. 3.1	100	3.9 s	855	2521	9.8×10^{-13}	4.0×10^{-13}
	200	9.8 s	996	1001	9.8×10^{-13}	1.4×10^{-13}
	500	05 m 45 s	1817	7929	9.9×10^{-13}	4.4×10^{-13}
	800	21 m 48 s	4694	4701	9.9×10^{-13}	9.1×10^{-13}
	1000	39 m 57 s	5152	5159	9.9×10^{-13}	1.3×10^{-13}
	1500	02 h 19 m 36 s	2736	13136	9.9×10^{-13}	6.4×10^{-13}
	2000	07 h 32 m 05 s	8966	9713	9.9×10^{-13}	3.8×10^{-13}
	3000	≥ 15 h	*	*	*	*
Alg. 3.1 with (3.19)	100	1.3 s	397	400	9.3×10^{-13}	4.3×10^{-13}
	200	36 s	406	409	9.4×10^{-13}	5.2×10^{-13}
	500	48 s	431	434	9.9×10^{-13}	1.8×10^{-13}
	800	02 m 34 s	451	454	9.3×10^{-13}	3.8×10^{-13}
	1000	04 m 36 s	480	483	9.8×10^{-13}	3.6×10^{-13}
	1500	14 m 46 s	520	523	9.8×10^{-13}	5.8×10^{-13}
	2000	35 m 53 s	562	565	9.7×10^{-13}	5.4×10^{-13}
	3000	02 h 02 m 58 s	630	633	9.9×10^{-13}	6.0×10^{-13}

We see from Tables 1 and 2 that Algorithm 3.1 is very efficient for solving large-scale problems. Moreover, the initial step length guess (3.19) substantially reduces the number of iterations and thus improves the efficiency.

Finally, to illustrate the efficiency of our algorithm, we compare Algorithm 3.1 with the ODE solver `ode113` provided by MATLAB for solving the differential equations (3.9) and (5.4) (as in [10]). For Algorithm 3.1 and the ODE solver `ode113` for solving (3.9) and (5.4), the starting points are randomly generated as in (6.1). For comparison purposes, in our numerical tests, the stopping criteria for Algorithm 3.1 and the ODE solver `ode113` for solving (3.9) and (5.4) are, respectively, set to be

$$\|H(Z^k, Q^k, U^k)\|_F \leq 10^{-8} \quad \text{and} \quad \|\Phi(Z^k, Q^k, U^k)\|_F \leq 10^{-8}.$$

TABLE 3
Comparison of Algorithm 3.1 *and* `ode113` *for Example* 6.1.

Alg.	n	CT.	IT.	NF.	Err.	Res.
`ode113` for (3.9)	10	0.98 s	2034	4336	8.48×10^{-9}	1.63×10^{-9}
	20	1.09 s	1989	4304	7.44×10^{-9}	1.29×10^{-9}
	50	4.35 s	3373	7376	9.98×10^{-9}	1.26×10^{-9}
	80	11.0 s	4596	10138	9.32×10^{-9}	9.71×10^{-10}
	100	19.6 s	5667	12520	9.31×10^{-9}	8.97×10^{-10}
	150	59.0 s	7520	16760	9.64×10^{-9}	7.52×10^{-10}
	200	02 m 15 s	9812	21929	9.87×10^{-9}	6.85×10^{-10}
	300	09 m 26 s	13518	30332	9.92×10^{-9}	5.53×10^{-10}
	400	26 m 43 s	17313	39011	9.78×10^{-9}	4.74×10^{-10}
	500	53 m 40 s	21106	47739	9.92×10^{-9}	4.36×10^{-10}
Alg. 3.1	10	0.22 s	332	1557	9.84×10^{-9}	1.07×10^{-8}
	20	0.31 s	377	1732	9.69×10^{-9}	1.11×10^{-8}
	50	0.76 s	394	1402	9.51×10^{-9}	9.04×10^{-9}
	80	0.81 s	387	392	9.71×10^{-9}	1.76×10^{-9}
	100	5.13 s	886	4117	9.89×10^{-9}	5.70×10^{-9}
	150	3.57 s	633	639	9.88×10^{-9}	1.42×10^{-9}
	200	14.4 s	764	3137	9.82×10^{-9}	8.17×10^{-9}
	300	30.6 s	1294	1300	9.90×10^{-9}	9.60×10^{-10}
	400	02 m 19 s	1163	5441	9.95×10^{-9}	4.02×10^{-9}
	500	03 m 11 s	1048	4442	9.98×10^{-9}	4.14×10^{-9}

TABLE 4
Comparison of Algorithm 3.1 *and* `ode113` *for Example* 6.2.

Alg.	n	CT.	IT.	NF.	Err.	Res.
`ode113` for (5.4)	10	0.37 s	1127	2394	6.88×10^{-9}	1.89×10^{-9}
	20	1.15 s	2487	5404	7.93×10^{-9}	1.24×10^{-9}
	50	6.91 s	6019	13330	9.79×10^{-9}	8.37×10^{-10}
	80	20.4 s	10338	23030	9.84×10^{-9}	5.95×10^{-10}
	100	25.9 s	9013	20186	9.76×10^{-9}	6.51×10^{-10}
	150	01 m 09 s	10583	23655	9.70×10^{-9}	6.02×10^{-10}
	200	02 m 13 s	11674	26119	9.64×10^{-9}	5.82×10^{-10}
	300	07 m 47 s	13830	31022	9.98×10^{-9}	5.53×10^{-10}
	400	22 m 49 s	17312	39008	9.93×10^{-9}	4.94×10^{-10}
	500	46 m 39 s	21323	48168	9.92×10^{-9}	4.51×10^{-10}
Alg. 3.1	10	0.19 s	330	1477	9.70×10^{-9}	1.14×10^{-8}
	20	0.25 s	355	1341	9.72×10^{-9}	7.43×10^{-9}
	50	0.86 s	440	1700	9.73×10^{-9}	7.25×10^{-9}
	80	1.84 s	551	1636	9.77×10^{-9}	4.22×10^{-9}
	100	2.04 s	505	1491	9.90×10^{-9}	4.15×10^{-9}
	150	6.89 s	657	2530	9.94×10^{-9}	6.40×10^{-9}
	200	5.82 s	642	647	9.99×10^{-9}	1.44×10^{-9}
	300	47.2 s	997	4472	9.92×10^{-9}	8.84×10^{-9}
	400	03 m 12 s	1664	7780	9.74×10^{-9}	8.87×10^{-9}
	500	03 m 14 s	1096	4805	9.97×10^{-9}	6.06×10^{-9}

For the ODE solver `ode113`, we evaluate the output values at time intervals of 10. The integration terminates automatically when the above stopping criteria are satisfied. In addition, the other parameters in Algorithm 3.1 are set as above.

Tables 3 and 4 display the numerical results for Examples 6.1–6.2. We observe from Tables 3 and 4 that the ODE solver `ode113` for (3.9) and (5.4) works acceptably for smalland medium-scale problems, while Algorithm 3.1 performs more effectively in terms of computing time.

We must point out the fact that the proposed algorithm converges to different solutions for different initial guesses, and the values of the cost function are almost zero at such different solutions. This can be observed from these and many other numerical tests.

7. Concluding remarks. In this paper, we have considered the doubly stochastic inverse eigenvalue problem (DSIEP) for constructing a doubly stochastic matrix from the given spectral information. We reformulate the inverse problem as a nonlinear least squares problem over several matrix manifolds. By exploiting the Riemannian gradient of the cost function and basic geometric properties (e.g., tangent space, retraction, and vector transport) of these matrix manifolds, we present a Riemannian Fletcher–Reeves conjugate gradient algorithm for the DSIEP. We have established the global convergence of the proposed algorithm. Moreover, we get an extra gain, i.e., our model yields a new Riemannian isospectral flow method. The proposed algorithm is also extended to the DSIEP with prescribed entries. Numerical experiments show that the proposed geometric algorithm is very effective for solving large-scale problems, while our new Riemannian isospectral flow methods work acceptably for small- and medium-scale problems. Since the proposed geometric algorithm converges to different solutions for different initial guesses, an interesting question is how to find an optimal approximation to a given doubly stochastic matrix. This needs further investigation.

Acknowledgments. We would like to thank the editor and the anonymous referees for their valuable comments and suggestions.

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), Vol. 5, 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] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub, *Newton's method on Riemannian manifolds and a geometric model for the human spine*, IMA J. Numer. Anal., 22 (2002), pp. 359–390.

[6] 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 et al., eds., Springer, New York, 2006, pp. 210–217.

[7] G. Birkhoff, *Three observations on linear algebra*, Univ. Nac. Tucumán. Rev. Ser. A, 5 (1946), pp. 147–151.

[8] R. A. Brualdi, *Some applications of doubly stochastic matrices*, Linear Algebra Appl., 107 (1988), pp. 77–100.

[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.

[10] 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.

[11] 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, MA, 1998.
[12] M. W. Hirsch, *Differential Topology*, Graduate Texts in Mathematics 33, Springer, New York, 1976.
[13] S. G. Hwang and S. S. Pyo, *The inverse eigenvalue problem for symmetric doubly stochastic matrices*, Linear Algebra Appl., 379 (2004), pp. 77–83.
[14] Y. J. Lei, W. R. Xu, Y. Lu, Y. R. Niu, and X. M. Gu, *On the symmetric doubly stochastic inverse eigenvalue problem*, Linear Algebra Appl., 445 (2014), pp. 181–205.
[15] J. D. Louck, *Doubly stochastic matrices in quantum mechanics*, Found. Phys., 27 (1997), pp. 1085–1104.
[16] M. Marcus, *Some properties and applications of doubly stochastic matrices*, Amer. Math. Monthly, 67 (1960), pp. 215–221.
[17] L. F. Martignon, *Doubly stochastic matrices with prescribed positive spectrum*, Linear Algebra Appl., 61 (1984), pp. 11–13.
[18] H. Minc, *Nonnegative Matrices*, Wiley, New York, 1988.
[19] B. Mourad, *An inverse problem for symmetric doubly stochastic matrices*, Inverse Problems, 19 (2003), pp. 821–831.
[20] B. Mourad, *On a spectral property of doubly stochastic matrices and its application to their inverse eigenvalue problem*, Linear Algebra Appl., 436 (2012), pp. 3400–3412.
[21] B. Mourad, H. Abbas, A. Mourad, A. Ghaddar, and I. Kaddoura, *An algorithm for constructing doubly stochastic matrices for the inverse eigenvalue problem*, Linear Algebra Appl., 439 (2013), pp. 1382–1400.
[22] 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.
[23] H. Perfect and L. Mirsky, *Spectral properties of doubly-stochastic matrices*, Monatsh. Math., 69 (1965), pp. 35–37.
[24] S. Petoukhov and M. He, *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications*, IGI Global, Hershey, PA, 2010.
[25] H. Sato, *A Dai-Yuan-type Riemannian conjugate gradient method with the weak Wolfe conditions*, Comput. Optim. Appl., to appear.
[26] H. Sato and T. Iwai, *A new, globally convergent Riemannian conjugate gradient method*, Optimization, 64 (2015), pp. 1011–1031.
[27] O. Rojo and H. Rojo, *Constructing symmetric nonnegative matrices via the fast Fourier transform*, Comput. Math. Appl., 45 (2003), pp. 1655–1672.
[28] O. Rojo and H. Rojo, *Some results on symmetric circulant matrices and on symmetric centrosymmetric matrices*, Linear Algebra Appl., 392 (2004), pp. 211–233.
[29] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer, New York, 1981.
[30] M. Shub, *Some remarks on dynamical systems and numerical analysis*, in Dynamical Systems and Partial Differential Equations, Proceedings of the 7th ELAM, L. Lara-Carrero and J. Lewowicz, eds., pp. 69–92, Equinoccio, Univ. Simón Bolívar, Caracas, 1986.
[31] S. T. Smith, *Optimization techniques on Riemannian manifolds*, Fields Inst. Commun., 3 (1994), pp. 113–136.
[32] G. W. Soules, *Constructing symmetric nonnegative matrices*, Linear Multilinear Algebra, 13 (1983), pp. 241–251.
[33] V. S. Sunder and N. J. Wildberger, *Action of finite hypergroups and examples*, in Harmonic Analysis and Hypergroups, Bikhoüser, Boston, MA, 1998.
[34] W. R. Xu, Y. J. Lei, X. M. Gu, Y. Lu, and Y. R. Niu, *Comment on: "A note on the inverse eigenvalue problem for symmetric doubly stochastic matrices"*, Linear Algebra Appl., 439 (2013), pp. 2256–2262.
[35] 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.
[36] L. Zhang, W. J. Zhou, and D. H. Li, *Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo-type line search*, Numer. Math., 104 (2006), pp. 561–572.
[37] 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.
[38] Z. Zhao, X. Q. Jin, and Z. J. Bai, *A geometric nonlinear conjugate gradient method for stochastic inverse eigenvalue problems*, SIAM J. Numer. Anal., submitted.
[39] S. Zhu, T. Gu, and X. Liu, *Solving inverse eigenvalue problems via Householder and rank-one matrices*, Linear Algebra Appl., 430 (2009), pp. 318–334.