A RIEMANNIAN NEWTON ALGORITHM FOR NONLINEAR EIGENVALUE PROBLEMS

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

Abstract. We give the formulation of a Riemannian Newton algorithm for solving a class of nonlinear eigenvalue problems by minimizing a total energy function subject to the orthogonality constraint. Under some mild assumptions, we establish the global and quadratic convergence of the proposed method. Moreover, the positive definiteness condition of the Riemannian Hessian of the total energy function at a solution is derived. Some numerical tests are reported to illustrate the efficiency of the proposed method for solving large-scale problems.

Keywords. Nonlinear eigenvalue problem, Riemannian Newton algorithm, Stiefel manifold, Grassmann manifold

AMS subject classifications. 15A18, 65F15, 49M15, 47J10

1. Introduction.

We consider the following total energy minimization problem:

$$\min_{X\in\mathbb{R}^{n\times k}} \overline{E}(X) := \frac{1}{2}\mathrm{tr}(X^TLX) + \frac{\alpha}{4}\rho(X)^TL^{-1}\rho(X) \quad \text{s.t.} \quad X^TX = I_k, \tag{1.1}$$

where X^T denotes the transpose of X, L is a discrete Laplacian operator, $\alpha > 0$ is a given constant, $\rho(X) := \mathrm{diag}(XX^T)$, "s.t." means "subject to", and I_k is the identity matrix of order k. We point out that the matrix L may be singular with different boundary conditions (see [45]). In this case, we may replace L^{-1} by the Moore-Penrose generalized inverse $L^\dagger$. The symbol $\mathrm{diag}(M) := (m_{11}, m_{22}, \ldots, m_{nn})^T$ denotes a vector containing the diagonal elements of an $n \times n$ matrix $M = [m_{ij}]$. Obviously, the first order necessary conditions for the total energy minimization problem (1.1) are given by [33]

$$H(X)X = X\widetilde{\Lambda}_k, \qquad X^TX = I_k,$$

where the k-by-k real symmetric matrix $\widetilde{\Lambda}_k$ is a Lagrange multiplier. We note that the global minimizer of the constrained minimization problem (1.1) is not unique. If X is a solution, then XQ is also a solution for any $k \times k$ real orthogonal matrix Q. Thus a necessary condition for the global minimum of problem (1.1) takes the form of a nonlinear eigenvalue problem (NEP) [46]:

$$H(X)X = X\Lambda_k, \qquad X^TX = I_k, \tag{1.2}$$

where the diagonal matrix $\Lambda_k \in \mathbb{R}^{k\times k}$ contains the k smallest eigenvalues of the symmetric matrix $H(X) = L + \alpha\mathrm{Diag}(L^{-1}\rho(X)) \in \mathbb{R}^{n\times n}$. The symbol $\mathrm{Diag}(\mathbf{x})$ is a diagonal matrix with a vector $\mathbf{x}$ on its diagonal. Note that the meaning of the notation $\mathrm{diag}(\cdot)$ is different from that of the notation $\mathrm{Diag}(\cdot)$.

*Department of Mathematics, University of Macau, Macao, People's Republic of China (zhaozhi231@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 is partially supported by the National Natural Science Foundation of China grant 11271308 and NCET.

‡Department of Mathematics, University of Macau, Macao, People's Republic of China (xqjin@umac.mo). The research of this author is supported by the research grant MYRG098(Y2-L3)-FST13-JXQ from University of Macau.

The total energy minimization problem (1.1) is a simplified version of the Hartree-Fock (HF) total energy minimization problem and the Kohn-Sham (KS) total energy minimization problem in electronic structure calculations (see for instance [30, 38, 44, 45]). Moreover, the NEP (1.2) is a simplified version of the associated HF and KS equations. The self-consistent field (SCF) iteration is widely used for solving the HF and KS equations, which calculates the k smallest eigenvalues and associated eigenvectors of the NEP (1.2) iteratively: Given the current iterate X^j, compute X^{j+1} such that

$$H(X^j)X^{j+1} = X^{j+1}\Lambda_k^{j+1} \quad \text{and} \quad (X^{j+1})^T X^{j+1} = I_k,$$

where Λ_k^{j+1} contains the k smallest eigenvalues of $H(X^j)$. However, the original version of the SCF iteration often fails to converge [11]. In past decades, different heuristics have been developed to accelerate and stabilize the SCF iteration [24, 25]. On the convergence of the SCF iteration, one may refer to [13, 27, 46]. In [45], the SCF iteration is used as an indirect way to solve problem (1.1) by minimizing a sequence of quadratic surrogate functions.

Recently, there are several optimization methods for solving the minimization problem (1.1) directly [5, 7, 25, 26, 32, 34, 35, 40, 41]. Because of the orthogonality constraint $X^T X = I_k$, those methods only use the gradient of the total energy and often converge slowly. In [44], a constrained optimization algorithm is proposed for minimizing the total energy by projecting the total energy into a sequence of subspaces and seeking the minimum point of the total energy over each subspace. In [42], a projected gradient-type method is given for minimizing a general function with the orthogonality constraint. In [16], Newton's method and the conjugate gradient (CG) method are developed on the Grassmann and Stiefel manifolds. In [28], modified steepest descent-type method with Armijo's line search and modified Newton method are presented on the Grassmann and Stiefel manifolds. Also, in [31], Line-search, trust-region, and Newton algorithms are well-studied on matrix manifolds. The SCF iteration with various trust-region techniques is employed to minimize the total energy [17, 18, 39, 45]. In [19], a Newton method is presented for solving a class of nonlinear eigenvalue problems arising from electronic structure calculation, which is only efficient for small-scale problems.

In this paper, we propose a Riemannian Newton algorithm for solving the total energy minimization problem (1.1) over the Grassmann manifold related to the Stiefel manifold $\mathrm{St}(k,n) := \{X \in \mathbb{R}^{n\times k} \mid X^T X = I_k\}$. This is sparked by two recent papers [27] and [19]. In [27], the convergence condition of the SCF iteration is related to the Hessian of the total energy. In [19], the NEP is viewed as a system of nonlinear equations, and then a Newton method is used for solving it. Therefore, in this paper, we first construct the Grassmann manifold from the Stiefel manifold $\mathrm{St}(k,n)$ based on an orthogonal equivalence relation and a Riemannian metric. Then we propose a Riemannian Newton algorithm for solving problem (1.1) over the Grassmann manifold. In particular, we combine the Riemannian Newton algorithm with the Riemannian linear search technique. Sparked by [2, 16, 28], we use the CG method [20, Algorithm 10.2.1] to solving each Newton equation inexactly, where we do not need the inversion of the Riemannian Hessian of the total energy function and thus the computational complexity is reduced. Also, the Riemannian linear search guarantees that the proposed method will converge to a local minimum [28]. Under some mild conditions, we show that the proposed Riemannian Newton algorithm converges globally and quadratically. Moreover, we give the positive

definiteness condition of the Riemannian Hessian of the total energy function at a solution. Some numerical experiments are reported to demonstrate the efficiency of our method for solving large-scale problems.

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 Newton algorithm for solving the minimization problem (1.1) over the Grassmann manifold related to the Stiefel manifold $\mathrm{St}(k,n)$. In section 4 we give a convergence analysis. In section 5 we investigate the positive definiteness condition of the Riemannian Hessian of the total energy function in problem (1.1) over the Grassmann manifold. In section 6 we report some numerical results and finally give some concluding remarks in section 7.

2. Preliminaries. In this section, we recall some basic concepts and results on Riemannian manifolds [1, 2]. Let $\mathcal{M}$ be a d-dimensional manifold. Let $\mathcal{R}_x(\mathcal{M})$ be the set of all smooth real-valued functions defined on a neighborhood of a point $x \in \mathcal{M}$. A tangent vector ξ_x to $\mathcal{M}$ at x is defined as a mapping from $\mathcal{R}_x(\mathcal{M})$ to $\mathbb{R}$ such that

$$\xi_x f = \dot{\gamma}(0)f := \left.\frac{d\big(f(\gamma(t))\big)}{dt}\right|_{t=0}, \quad \forall f \in \mathcal{R}_x(\mathcal{M}),$$

for some smooth curve γ on $\mathcal{M}$ with $\gamma(0) = x$. The tangent space $T_x\mathcal{M}$ to $\mathcal{M}$ at x is consisted 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 $\xi : \mathcal{M} \to T\mathcal{M}$ such that $\xi(x) = \xi_x \in T_x\mathcal{M}$ for all $x \in \mathcal{M}$. A Riemannian metric g on $\mathcal{M}$ is a family of inner products

$$g_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}, \quad x \in \mathcal{M},$$

where the inner product $g_x(\cdot,\cdot)$ varies smoothly and induces a norm $\|\xi_x\| = \sqrt{g_x(\xi_x,\xi_x)}$ on $T_x\mathcal{M}$. Thus, $(\mathcal{M}, g)$ is a Riemannian manifold [2, p.45].

Let $\mathcal{M}$ and $\mathcal{L}$ be two manifolds. Let $G : \mathcal{M} \to \mathcal{L}$ be a smooth mapping. Then the differential $\mathrm{D}G(x)$ of G at $x \in \mathcal{M}$ is a mapping from $T_x\mathcal{M}$ to $T_{G(x)}\mathcal{L}$ such that

$$\mathrm{D}G(x)[\xi_x] \in T_{G(x)}\mathcal{L}, \quad \forall \xi_x \in T_x\mathcal{M},$$

where $\mathrm{D}G(x)[\xi_x]$ is a tangent vector to $\mathcal{L}$ at $G(x) \in \mathcal{L}$, which is a mapping from $\mathcal{R}_{G(x)}(\mathcal{L})$ to $\mathbb{R}$ defined by:

$$\mathrm{D}G(x)[\xi_x]f = \xi_x(f \circ G), \quad \forall f \in \mathcal{R}_{G(x)}(\mathcal{L}).$$

Given a Riemannian manifold $(\mathcal{M}, g)$ with a Riemannian connection ∇ (see for instance [2, 10]), let $f : \mathcal{M} \to \mathbb{R}$ be a smooth function. Then the Riemannian gradient $\mathrm{grad}\, f(x)$ of f at $x \in \mathcal{M}$ is defined as the unique element in $T_x\mathcal{M}$ such that

$$g_x(\mathrm{grad}\, f(x), \xi_x) = \mathrm{D}f(x)[\xi_x], \quad \forall \xi_x \in T_x\mathcal{M}.$$

The Riemannian Hessian of f at $x \in \mathcal{M}$ is defined as the linear mapping from $T_x\mathcal{M}$ to $T_x\mathcal{M}$ such that [2, Definition 5.5.1],

$$\mathrm{Hess}\, f(x)[\xi_x] = \nabla_{\xi_x}\, \mathrm{grad}\, f(x), \quad \forall \xi_x \in T_x\mathcal{M}.$$

The concept of retraction originally appears in the field of algebraic topology [21]. Here, we adopt the following definition of retraction [2, 4, 37].

DEFINITION 2.1. *Let $\mathcal{M}$ be a manifold. Let R be a mapping from $T\mathcal{M}$ onto $\mathcal{M}$. Let R_x denote the restriction of R to $T_x\mathcal{M}$. We say that R is a retraction on $\mathcal{M}$ if*

(i) *R is smooth.*
(ii) *$R_x(0_X) = x$, where 0_X is the origin of $T_x\mathcal{M}$.*
(iii) *$\mathrm{D}R_x(0_X) = \mathrm{id}_{T_x\mathcal{M}}$, where $\mathrm{id}_{T_x\mathcal{M}}$ is the identity mapping on $T_x\mathcal{M}$ with the canonical identification $T_{0_X}T_x\mathcal{M} \simeq T_x\mathcal{M}$.*

For a real-valued function f on the manifold $\mathcal{M}$ and a retraction R on $\mathcal{M}$, we define the pullback $\widehat{f}$ of f as the mapping from $T\mathcal{M}$ to $\mathbb{R}$ such that

$$\widehat{f}(\xi) = f(R(\xi)), \quad \forall \xi \in T\mathcal{M}, \tag{2.1}$$

and let $\widehat{f}_x$ mean the restriction of $\widehat{f}$ to $T_x\mathcal{M}$, which is defined by

$$\widehat{f}_x(\xi_x) = f(R_x(\xi_x)), \quad \forall \xi_x \in T_x\mathcal{M}.$$

On the Riemannian distance to a nondegenerate local minimizer $\overline{x}$ of a smooth real-valued function f on $(\mathcal{M}, g)$, we have the following lemma [2, Lemma 7.4.8].

LEMMA 2.2. *Let $x^* \in \mathcal{M}$ and let $f : \mathcal{M} \to \mathbb{R}$ be a C^2 function (its first and second derivatives are continuous) such that $\operatorname{grad} f(x^*) = 0$ and $\operatorname{Hess} f(x^*)$ is positive-definite with maximal and minimal eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$. Then given two positive scalars τ_0, τ_1 with $\tau_0 < \lambda_{\min}$ and $\tau_1 > \lambda_{\max}$, there exists a neighborhood $\mathcal{N}(x^*)$ of x^* such that*

$$\tau_0 \operatorname{dist}(x^*, x) \leq \| \operatorname{grad} f(x)\| \leq \tau_1 \operatorname{dist}(x^*, x), \quad \forall x \in \mathcal{N}(x^*),$$

where $\operatorname{dist}(\cdot, \cdot)$ means the Riemannian distance on $(\mathcal{M}, g)$ [2, p. 46].

On a relation between the Riemannian gradient of a smooth function f on $\mathcal{M}$ at $R_x(\xi)$ and the gradient of $\widehat{f}_x$ at $\xi \in T_x\mathcal{M}$ with $\|\xi\| \leq \delta$ for some $\delta > 0$, we have the following special result [2, Lemma 7.4.9].

LEMMA 2.3. *Let R be a retraction on $\mathcal{M}$ and let f be a continuously differentiable cost function on $\mathcal{M}$. Then for any given $x^* \in \mathcal{M}$ and a scalar $\tau_2 > 1$, there exist a neighborhood $\mathcal{N}(x^*)$ of x^* and $\delta > 0$ such that*

$$\| \operatorname{grad} f(R_x(\xi))\| \leq \tau_2 \| \operatorname{grad} \widehat{f}_x(\xi)\|,$$

for all $x \in \mathcal{N}(x^*)$ and all $\xi \in T_x\mathcal{M}$ with $\|\xi\| \leq \delta$, where $\widehat{f}$ is defined as in (2.1).

3. Riemannian Newton Algorithm. In this section, we propose a Riemannian Newton algorithm for solving the total energy minimization problem (1.1). We first construct a Grassmann manifold from the Stiefel manifold $\mathrm{St}(k, n)$. Then, based on the induced Grassmann manifold, we give a matrix-form Riemannian Newton algorithm for solving problem (1.1).

3.1. The Grassmann manifold. We observe the fact that the function $\overline{E} : \mathrm{St}(k, n) \to \mathbb{R}$ defined in problem (1.1) is such that for any given $\overline{X} \in \mathrm{St}(k, n)$, $\overline{E}(\overline{X}) = \overline{E}(\overline{X}Q)$ for all $Q \in O_k$, where O_k is the set of all $k \times k$ orthogonal matrices. Thus, the global minimizer of problem (1.1) is not unique and is not isolated. The Riemannian Hessian of $\overline{E}$ must be singular, which causes a trouble for applying a Riemannian Newton algorithm to problem (1.1). To overcome this difficulty,

we construct a Grassmann manifold $\mathcal{Q}$ from the Stiefel manifold $\mathrm{St}(k,n)$ under the operation of orthogonal group O_k. We define a quotient manifold by

$$\mathcal{Q} := \mathrm{St}(k,n)/O_k, \tag{3.1}$$

based on the following equivalence relation on $\mathrm{St}(k,n)$:

$$\overline{X} \sim \overline{Y} \iff \big\{\overline{X}Q \mid Q \in O_k\big\} = \big\{\overline{Y}Q \mid Q \in O_k\big\}.$$

Then we have $\mathcal{Q} := \big\{[\overline{X}] : \overline{X} \in \mathrm{St}(k,n)\big\}$, where

$$[\overline{X}] := \big\{\overline{Y} \in \mathrm{St}(k,n) \mid \overline{Y} = \overline{X}Q,\ Q \in O_k\big\}$$

is the equivalent class containing $\overline{X}$. The *natural projection* is defined as the mapping from $\mathrm{St}(k,n)$ to $\mathcal{Q}$ such that

$$\pi(\overline{X}) = [\overline{X}], \quad \forall \overline{X} \in \mathrm{St}(k,n).$$

Moreover, we have $[\overline{X}] = \pi^{-1}(\pi(\overline{X}))$ and $\dim \pi^{-1}(\pi(\overline{X})) = \dim O_k = 1/2k(k-1)$. Since $\mathrm{St}(k,n)$ is the total space of $\mathcal{Q}$, we have [2, Proposition 3.4.4],

$$\begin{aligned} \dim \mathcal{Q} &= \dim \mathrm{St}(k,n) - \dim \pi^{-1}(\pi(X)) \\ &= nk - \tfrac{1}{2}k(k+1) - \tfrac{1}{2}k(k-1) = k(n-k). \end{aligned}$$

The function $\overline{E} : \mathrm{St}(k,n) \to \mathbb{R}$ induces a unique function $E : \mathcal{Q} \to \mathbb{R}$ such that

$$E\big([\overline{X}]\big) = \overline{E}\big(\pi^{-1}([\overline{X}])\big) \quad \text{and} \quad \overline{E}(\overline{X}) = E\big(\pi(\overline{X})\big).$$

Hence, problem (1.1) can be written as the following minimization problem:

$$\min E(X) \quad \text{s.t.} \quad X \in \mathcal{Q}. \tag{3.2}$$

For any $X \in \mathcal{Q}$, let ξ_X be an element of $T_X\mathcal{Q}$, and let $\overline{X}$ be an element in the equivalent class $\pi^{-1}(X)$, which is an embedded submanifold of $\mathrm{St}(k,n)$. Any element $\overline{\xi}_{\overline{X}} \in T_{\overline{X}}\mathrm{St}(k,n)$ that satisfies $\mathrm{D}\pi(\overline{X})[\overline{\xi}_{\overline{X}}] = \xi_X$ can be considered a representation of ξ_X. For any smooth function $f : \mathcal{Q} \to \mathbb{R}$, the function $\overline{f} := f \circ \pi : \mathrm{St}(k,n) \to \mathbb{R}$ is smooth [2, Proposition 3.4.5]. Moreover,

$$\mathrm{D}\overline{f}(\overline{X})[\overline{\xi}_{\overline{X}}] = \mathrm{D}f\big(\pi(\overline{X})\big)\big[\mathrm{D}\pi(\overline{X})[\overline{\xi}_{\overline{X}}]\big] = \mathrm{D}f(X)[\xi_X].$$

Since there are infinitely many valid representations $\overline{\xi}_{\overline{X}}$ of ξ_X at $\overline{X}$, we need to define the vertical space and horizontal space at the point $\overline{X}$ [2, p.48]. Note that the tangent space to $\mathrm{St}(k,n)$ at $X \in \mathrm{St}(k,n)$ is given by [2, p.42],

$$\begin{aligned} T_{\overline{X}}\mathrm{St}(k,n) &= \{Z \in \mathbb{R}^{n\times k} : \overline{X}^T Z + Z^T \overline{X} = 0\} \\ &= \{\overline{X}\Omega + \overline{X}_\perp K : \Omega^T = -\Omega,\ K \in \mathbb{R}^{(n-k)\times k}\}, \end{aligned}$$

where $\overline{X}_\perp \in \mathbb{R}^{n\times(n-k)}$ such that $\mathrm{span}(\overline{X}_\perp)$ is the orthogonal complement of $\mathrm{span}(\overline{X})$. Also, a Riemannian metric $\overline{g}$ on $\mathrm{St}(k,n)$ is defined by

$$\overline{g}_{\overline{X}}(\overline{Z}_1, \overline{Z}_2) := \mathrm{tr}(\overline{Z}_1^T \overline{Z}_2), \quad \forall \overline{Z}_1, \overline{Z}_2 \in T_{\overline{X}}\mathrm{St}(k,n), \quad \overline{X} \in \mathrm{St}(k,n),$$

and its induced Frobenius norm $\|\cdot\|_{\overline{X}}$. Thus, the *vertical space* at $\overline{X}$ is defined as

$$\mathcal{V}_{\overline{X}} := T_{\overline{X}}\big(\pi^{-1}(X)\big) = \big\{\overline{X}\Omega : \Omega^T = -\Omega,\ \Omega \in \mathbb{R}^{k\times k}\big\}.$$

We can set the *horizontal space* $\mathcal{H}_{\overline{X}}$ at $\overline{X}$ to be

$$\begin{aligned}\mathcal{H}_{\overline{X}} : &= \mathcal{V}_{\overline{X}}^{\perp} = \big\{\xi_{\overline{X}} \in T_{\overline{X}}\mathrm{St}(k,n) : \overline{g}_{\overline{X}}(\xi_{\overline{X}}, \nu_{\overline{X}}) = 0 \text{ for all } \nu_{\overline{X}} \in \mathcal{V}_{\overline{X}}\big\}\\ &= \big\{\xi_{\overline{X}} \in T_{\overline{X}}\mathrm{St}(k,n) : \overline{X}^T\xi_{\overline{X}} = 0\big\} = \big\{\overline{X}_{\perp}K : K \in \mathbb{R}^{(n-k)\times k}\big\}.\end{aligned}$$

Then for any $\xi_X \in T_X\mathcal{Q}$, there exists a unique element $\overline{\xi}_{\overline{X}} \in \mathcal{H}_{\overline{X}}$ such that $\mathrm{D}\pi(\overline{X})[\overline{\xi}_{\overline{X}}] = \xi_X$, and $\overline{\xi}_{\overline{X}}$ is called the *horizontal lift* of ξ_X at $\overline{X}$. Moreover, the orthogonal projection of any element $\eta_{\overline{X}} \in T_{\overline{X}}\mathrm{St}(k,n)$ onto $\mathcal{H}_{\overline{X}}$ at $\overline{X}$ is given by

$$\mathrm{P}^h_{\overline{X}}\eta_{\overline{X}} = \big(I_n - \overline{X}\,\overline{X}^T\big)\eta_{\overline{X}}.$$

Now, we define a Riemannian metric g on the quotient manifold $\mathcal{Q}$ by

$$g_X(\xi_X, \zeta_X) := g_{\overline{X}}(\overline{\xi}_{\overline{X}}, \overline{\zeta}_{\overline{X}}), \quad \xi_X, \zeta_X \in T_X\mathcal{Q},\ \ X \in \mathcal{Q}, \tag{3.3}$$

where $\overline{\xi}_{\overline{X}}, \overline{\zeta}_{\overline{X}} \in \mathcal{H}_{\overline{X}}$ are the unique horizontal lifts of ξ_X, ζ_X at $\overline{X}$ respectively. Since $\overline{X} \in \pi^{-1}(X)$, $\overline{X}Q$ is in $\pi^{-1}(X)$ for any $Q \in O_k$, we need to show that

$$\overline{g}_{\overline{X}Q}(\overline{\xi}_{\overline{X}Q}, \overline{\zeta}_{\overline{X}Q}) = \overline{g}_{\overline{X}}(\overline{\xi}_{\overline{X}}, \overline{\zeta}_{\overline{X}}), \quad \forall Q \in O_k. \tag{3.4}$$

To verify (3.4), we first establish the following result. The proof is similar to that of Proposition 3.6.1 in [1], and we therefore omit it.

PROPOSITION 3.1. *Let $\overline{X} \in \mathrm{St}(k,n)$, $X = \pi(\overline{X})$, and $\xi_X \in T_X\mathcal{Q}$. Then it holds*

$$\overline{\xi}_{\overline{X}Q} = \overline{\xi}_{\overline{X}} \cdot Q$$

for all $Q \in O_k$, where the center dot denotes matrix multiplication, and

$$\overline{g}_{\overline{X}Q}(\overline{\xi}_{\overline{X}Q}, \overline{\zeta}_{\overline{X}Q}) = \overline{g}_{\overline{X}}(\overline{\xi}_{\overline{X}}, \overline{\zeta}_{\overline{X}}),$$

for all $\xi_X, \zeta_X \in T_X\mathcal{Q}$.

Thus the quotient manifold $\mathcal{Q}$ endowed with the Riemannian metric g defined in (3.3) is a Grassmann manifold.

Next, we define a second-order retraction R on $(\mathcal{Q}, g)$ as follows:

$$R_X\big(\xi_X\big) := \pi\big(\overline{R}_{\overline{X}}(\overline{\xi}_{\overline{X}})\big), \quad \forall \xi_X \in T_X\mathcal{Q}, \tag{3.5}$$

where $X = \pi(\overline{X}) \in \mathcal{Q}$, $\overline{\xi}_{\overline{X}} \in \mathcal{H}_{\overline{X}}$ is the horizontal lift of a $\xi_X \in T_X\mathcal{Q}$ at $\overline{X}$, and $\overline{R}$ is a second-order retraction on $\mathrm{St}(k,n)$, which is defined by [1, 3]:

$$\overline{R}_{\overline{X}}(\overline{Z}) = \sum_{i=1}^{k} \bar{\mathbf{u}}_{\mathbf{i}}\bar{\mathbf{v}}_{\mathbf{i}}{}^T, \quad \forall\ \overline{Z} \in T_{\overline{X}}\mathrm{St}(k,n), \tag{3.6}$$

where $\{\bar{\mathbf{u}}\}_{\mathbf{i=1}}^{\mathbf{k}}$ and $\{\bar{\mathbf{v}}\}_{\mathbf{i=1}}^{\mathbf{k}}$ are the left and right singular vectors corresponding to the largest k singular values of $\overline{X} + \overline{Z}$, which admits the singular value decomposition [20]:

$$\overline{X} + \overline{Z} = \overline{U}\,\overline{\Sigma}\,\overline{V}^T, \quad \overline{\Sigma} = \mathrm{Diag}(\bar{\sigma}_1(\overline{X} + \overline{Z}), \ldots, \bar{\sigma}_k(\overline{X} + \overline{Z})) \in \mathbb{R}^{n\times k}.$$

Here, $\bar{\sigma}_1(\overline{X}+\overline{Z}) \geq \bar{\sigma}_2(\overline{X}+\overline{Z}) \geq \cdots \geq \bar{\sigma}_k(\overline{X}+\overline{Z}) > 0$ and $\overline{U} = [\bar{\mathbf{u}}_\mathbf{1}, \ldots, \bar{\mathbf{u}}_\mathbf{n}] \in \mathbb{R}^{n\times n}$ and $\overline{V} = [\bar{\mathbf{v}}_\mathbf{1}, \ldots, \bar{\mathbf{v}}_\mathbf{k}] \in \mathbb{R}^{k\times k}$ are orthogonal matrices. The retraction $\overline{R}$ on $\mathrm{St}(k,n)$ may reduce the computational complexity and accelerate convergence [28, 29]. In our numerical experiments, we use the retraction $\overline{R}$. Obviously, for the retractions $\overline{R}$ defined in (3.6), we have $\pi(\overline{R}_{\overline{X}_a}(\bar{\xi}_{\overline{X}_a})) = \pi(\overline{R}_{\overline{X}_b}(\bar{\xi}_{\overline{X}_b}))$ for all $\overline{X}_a, \overline{X}_b \in \pi^{-1}(X)$. Thus R defined by (3.5) is a retraction on $\mathcal{Q}$ [2, Proposition 4.1.3].

For the function $\overline{E} : \mathrm{St}(k,n) \to \mathbb{R}$ defined in problem (1.1), we have $\overline{E} = E \circ \pi$, where $E : \mathcal{Q} \to \mathbb{R}$ is defined in problem (3.2). By the smoothness of $\overline{E}$ on $\mathrm{St}(k,n)$, we know that E is smooth on $\mathcal{Q}$.

3.2. Riemannian gradient and Riemannian Hessian of E. We give explicit formulas of Riemannian gradient and Riemannian Hessian of the cost function E defined in problem (3.2). To do so, we define the extended function $\widetilde{E} : \mathbb{R}^{n\times k} \to \mathbb{R}$ by

$$\widetilde{E}(\overline{X}) = \frac{1}{2}\mathrm{tr}(\overline{X}^T L \overline{X}) + \frac{\alpha}{4}\rho(\overline{X})L^{-1}\rho(\overline{X}), \quad \forall \overline{X} \in \mathbb{R}^{n\times k}.$$

Then $\overline{E}$ is the restriction of $\widetilde{E}$ onto $\mathrm{St}(k,n)$, i.e., $\overline{E} = \widetilde{E}|_{\mathrm{St}(k,n)}$. By simple calculation, the gradient of $\widetilde{E}$ at $\overline{X} \in \mathbb{R}^{n\times k}$ is given by [2, p.48]

$$\mathrm{grad}\, \widetilde{E}(\overline{X}) = H(\overline{X})\overline{X}.$$

Since $\mathrm{St}(k,n)$ is a Riemannian submanifold of $\mathbb{R}^{n\times k}$, the Riemannian gradient of $\widetilde{E}$ at $\overline{X} \in \mathrm{St}(k,n)$ is given by

$$\mathrm{grad}\, \overline{E}(\overline{X}) = \mathrm{P}^h_{\overline{X}}\big(\mathrm{grad}\, \widetilde{E}(\overline{X})\big) = \mathrm{P}^h_{\overline{X}}\big(H(\overline{X})\overline{X}\big).$$

where $\mathrm{P}_{\overline{X}}$ means the orthogonal projection onto $T_{\overline{X}}\mathcal{M}$, which is given by

$$\mathrm{P}_{\overline{X}}\overline{Z} = (I_n - \overline{X}\,\overline{X}^T)\overline{Z} + \overline{X}\mathrm{skew}(\overline{X}^T\overline{Z}) = \overline{Z} - \overline{X}\mathrm{sym}(\overline{X}^T\overline{Z}), \quad \forall \overline{Z} \in \mathbb{R}^{n\times k}.$$

Here, $\mathrm{skew}(A) := (A - A^T)/2$ and $\mathrm{sym}(A) := (A + A^T)/2$. Therefore, for any $X \in \mathcal{Q}$ and $\overline{X} \in \pi^{-1}(X)$, the unique horizontal lift of the Riemannian gradient $\mathrm{grad}\, E(X)$ of E at $\overline{X} \in \mathrm{St}(k,n)$ is given by

$$\overline{\mathrm{grad}\, E(X)}_{\overline{X}} = \mathrm{grad}\, \overline{E}(\overline{X}) = \mathrm{P}^h_{\overline{X}}\big(H(\overline{X})\overline{X}\big) = (I_n - \overline{X}\,\overline{X}^T)H(\overline{X})\overline{X}. \tag{3.7}$$

Let ∇ and $\overline{\nabla}$ be the Riemannian connections on $\mathcal{Q}$ and $\mathrm{St}(k,n)$. The Riemannian Hessian of E at $X \in \mathcal{Q}$ is given by

$$\mathrm{Hess}\, E(X)[Z_X] = \nabla_{Z_X}\, \mathrm{grad}\, E(X), \quad \forall Z_X \in T_X\mathcal{Q}.$$

Since $P^h_{\overline{X}}P_{\overline{X}}Z = P^h_{\overline{X}}Z$ for all $Z \in T_{\overline{X}}\mathrm{St}(k,n)$, we have [2, Eqn. (5.15) and Proposition 5.3.3],

$$\begin{aligned}\overline{\mathrm{Hess}\, E(X)[Z_X]}_{\overline{X}} &= \overline{\nabla_{Z_X}\, \mathrm{grad}\, E(X)}_{\overline{X}} = \mathrm{P}^h_{\overline{X}}\big(\overline{\nabla}_{\overline{Z}_{\overline{X}}}\overline{\mathrm{grad}\, E(X)}_{\overline{X}}\big)\\ &= \mathrm{P}^h_{\overline{X}}\Big(\mathrm{P}_{\overline{X}}\big(\mathrm{D}\,\mathrm{grad}\, \overline{E}(\overline{X})[\overline{Z}_{\overline{X}}]\big)\Big) = \mathrm{P}^h_{\overline{X}}\big(\mathrm{D}\,\mathrm{grad}\, \overline{E}(\overline{X})[\overline{Z}_{\overline{X}}]\big).\end{aligned}$$

where $\mathrm{D}\,\mathrm{grad}\, f(x)[\xi_x]$ means the classical directional derivative. We get by (3.7),

$$\begin{aligned}\mathrm{D}\,\mathrm{grad}\, \overline{E}(\overline{X})[\overline{Z}_{\overline{X}}] = &-\big(\overline{X}\,\overline{Z}_{\overline{X}}^T + \overline{Z}_{\overline{X}}\,\overline{X}^T\big)H(\overline{X})\overline{X}\\ &+2\alpha\big(I_n - \overline{X}\,\overline{X}^T\big)\mathrm{Diag}\big(L^{-1}\mathrm{diag}(\overline{X}\,\overline{Z}_{\overline{X}}^T)\big)\overline{X}\\ &+\big(I_n - \overline{X}\,\overline{X}^T\big)H(\overline{X})\overline{Z}_{\overline{X}}.\end{aligned}$$

Thus

$$\begin{aligned}&\overline{\mathrm{Hess}\ E(X)[Z_X]}_{\overline{X}}\\ &= \mathrm{P}^h_{\overline{X}}\Big(-\overline{Z}_{\overline{X}}\overline{X}^T H(\overline{X})\overline{X} + 2\alpha \mathrm{Diag}\big(L^{-1}\mathrm{diag}(\overline{X}\ \overline{Z}_{\overline{X}}^T)\big)\overline{X} + H(\overline{X})\overline{Z}_{\overline{X}}\Big),\end{aligned} \tag{3.8}$$

where the fact of $\mathrm{P}^h_{\overline{X}}\overline{X} = 0$ is used.

We remark that the Newton equation on the Grassmann manifold $\mathcal{Q}$ at the point $X \in \mathcal{Q}$ is given by [2, p.113],

$$\mathrm{Hess}\, E(X)[Z_X] = -\,\mathrm{grad}\, E(X), \quad Z_X \in T_X\mathcal{Q}.$$

Taking the horizontal lift yields

$$\overline{\mathrm{Hess}E(X)[Z_X]}_{\overline{X}} = -\overline{\mathrm{grad}\, E(X)}_{\overline{X}},$$

or

$$\mathrm{P}^h_{\overline{X}}\big(-\overline{Z}_{\overline{X}}\overline{X}^T H(\overline{X})\overline{X} + 2\alpha \mathrm{Diag}\big(L^{-1}\mathrm{diag}(\overline{X}\ \overline{Z}_{\overline{X}}^T)\big)\overline{X} + H(\overline{X})\overline{Z}_{\overline{X}}\big) = -\mathrm{P}^h_{\overline{X}}\big(H(\overline{X})\overline{X}\big),$$

for $\overline{Z}_{\overline{X}} \in \mathcal{H}_{\overline{X}}$.

3.3. Riemannian Newton algorithm. Without causing any confusion, we use $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ to denote the Riemannian metrics and their induced norms on $\mathrm{St}(k,n)$ and $\mathcal{Q}$ respectively. Based on the discussion in section 3.2, we describe a matrix-form Riemannian Newton algorithm for solving the minimization problem (3.2).

ALGORITHM 3.2. (A matrix-form Riemannian Newton algorithm)

Step 0. *Given* $\overline{X}^0 \in \mathrm{St}(k,n)$, $\beta,\eta \in (0,1)$, $\sigma \in (0,1/2]$, *and* $j := 0$.

Step 1. *Apply the CG method* [20, Algorithm 10.2.1] *to solving*

$$\mathrm{P}^h_{\overline{X}^j}\big(\mathrm{D}\,\mathrm{grad}\,\overline{E}(\overline{X}^j)[\Delta\overline{X}^j]\big) + \mathrm{grad}\,\overline{E}(\overline{X}^j) = 0, \tag{3.9}$$

for $\Delta\overline{X}^j \in \mathcal{H}_{\overline{X}^j}$ *such that*

$$\big\|\mathrm{P}^h_{\overline{X}^j}\big(\mathrm{D}\,\mathrm{grad}\,\overline{E}(\overline{X}^j)[\Delta\overline{X}^j]\big) + \mathrm{grad}\,\overline{E}(\overline{X}^j)\big\| \le \eta_j \big\|\mathrm{grad}\,\overline{E}(\overline{X}^j)\big\|, \tag{3.10}$$

and

$$\big\langle \mathrm{grad}\,\overline{E}(\overline{X}^j), \Delta\overline{X}^j\big\rangle \le -\eta_j\big\langle\Delta\overline{X}^j, \Delta\overline{X}^j\big\rangle, \tag{3.11}$$

where $\eta_j := \min\{\eta, \big\|\mathrm{grad}\,\overline{E}(\overline{X}^j)\big\|\}$. *If* (3.10) *and* (3.11) *are not attainable, then let*

$$\Delta\overline{X}^j := -\,\mathrm{grad}\,\overline{E}(\overline{X}^j).$$

Step 2. *Let* l_j *be the smallest nonnegative integer* l *such that*

$$\overline{E}(\overline{R}_{\overline{X}^j}(\beta^l\Delta\overline{X}^j)) - \overline{E}(\overline{X}^j) \le \sigma\beta^l\big\langle \mathrm{grad}\,\overline{E}(\overline{X}^j), \Delta\overline{X}^j\big\rangle. \tag{3.12}$$

Set

$$\overline{X}^{j+1} := \overline{R}_{\overline{X}^j}(\beta^{l_j}\Delta\overline{X}^j)Q^j, \quad \textit{for some } Q^j \in O_k.$$

Step 3. *Replace j by $j+1$ and go to* **Step 1**.

We remark that Algorithm 3.2 is a numerically realizable Riemannian Newton algorithm for solving the minimization problem (3.2). Suppose that $\{\overline{X}^j\}$ and $\{\overline{Y}^j\}$ are two sequences generated by Algorithm 3.2. If $[\overline{X}^0] = [\overline{Y}^0]$, then $[\overline{X}^j] = [\overline{Y}^j]$ for all j. Thus, Algorithm 3.2 returns a sequence $\{[\overline{X}^j]\} \in \mathcal{Q}$ by taking $X^0 = [\overline{X}^0] \in \mathcal{Q}$, where $\overline{X}^0 \in \mathrm{St}(k,n)$. We also point out that our method has some advantages over classical equality-constrained optimization methods: (1) A nice feature is that the generated iterates are all feasible. (2) As shown in section 4, our method converges globally and quadratically as an unconstrained optimization on a constrained set. (3) No additional Lagrange multipliers or penalty functions are required. Finally, numerical tests in section 6 show the efficiency of our method over the classical interior-point method [12].

4. Convergence Analysis. In this section, we establish the global and quadratic convergence of Algorithm 3.2. As in (2.1), we have the following equality on the Riemannian gradient of E and its pullback function $\widehat{E}$ through the retraction R defined in (3.5) [2, p.56]:

$$\operatorname{grad} E(X) = \operatorname{grad} \widehat{E}_X(0_X), \quad \forall X \in \mathcal{Q}, \quad 0_X \in T_X\mathcal{Q}. \tag{4.1}$$

For the second-order retraction R on $\mathcal{Q}$ defined in (3.5), we have [2, Proposition 5.5.5]:

$$\operatorname{Hess} E(X) = \operatorname{Hess} \widehat{E}(0_X), \quad \forall X \in \mathcal{Q}, \quad 0_X \in T_X\mathcal{Q}. \tag{4.2}$$

4.1. Global convergence. On the global convergence of Algorithm 3.2, we have the following result. The proof follows that of Theorem 11(a) in [15].

THEOREM 4.1. *Any accumulation point $\overline{X}_*$ of the sequence $\{\overline{X}^j\}$ generated by Algorithm* 3.2 *produces a stationary point $X_* := [\overline{X}_*]$ of the cost function E defined in problem* (3.2).

Proof. Suppose $\{\overline{X}^j\} \to \overline{X}_*$, renumbering if necessary. If there exists a subsequence $\{\Delta\overline{X}^j\}_{\mathcal{J}}$ such that $\Delta\overline{X}^j = -\operatorname{grad}\overline{E}(\overline{X}^j)$ for all $k \in \mathcal{J}$, then $\overline{X}_*$ is a stationary point of $\overline{E}$. We note that $\mathrm{D}\pi(\overline{X}_*)[\operatorname{grad}\overline{E}(\overline{X}_*)] = \operatorname{grad} E(X_*)$. Hence, $[\overline{X}_*]$ is a stationary point of E. Therefore, without loss of generality, to prove the theorem we only need to consider the case in which the direction is always given by (3.9). To verify that $\operatorname{grad} E(X_*) = 0$, we only need to show that $\operatorname{grad}\overline{E}(\overline{X}_*) = 0$. By contradiction, we assume that $\operatorname{grad}\overline{E}(\overline{X}_*) \neq 0$. Let $X^j := [\overline{X}^j]$ for all j. By (3.3) and (3.10), we have

$$\begin{aligned}
\|\operatorname{grad}\overline{E}(\overline{X}^j)\| &\le \|\overline{\operatorname{Hess} E(X^j)[\Delta X^j]}_{\overline{X}^j}\| + \|\overline{\operatorname{Hess} E(X^j)[\Delta X^j]}_{\overline{X}^j} + \operatorname{grad}\overline{E}(\overline{X}^j)\| \\
&= \|\operatorname{Hess} E(X^j)[\Delta X^j]\| + \|\overline{\operatorname{Hess} E(X^j)[\Delta X^j]}_{\overline{X}^j} + \operatorname{grad}\overline{E}(\overline{X}^j)\| \\
&\le \|\operatorname{Hess} E(X^j)\| \cdot \|\Delta X^j\| + \eta_j \|\operatorname{grad}\overline{E}(\overline{X}^j)\| \\
&\le \|\operatorname{Hess} E(X^j)\| \cdot \|\Delta\overline{X}^j\| + \eta \|\operatorname{grad}\overline{E}(\overline{X}^j)\|, \qquad (4.3)
\end{aligned}$$

where $0 < \eta_j \le \eta < 1$, and $\|\operatorname{Hess} E(X^j)\|$ denotes the operator norm defined by

$$\|\operatorname{Hess} E(X^j)\| := \sup\big\{\|\operatorname{Hess} E(X^j)[\Delta X^j]\| : \Delta X^j \in T_{X^j}\mathrm{St}(k,n),\ \|\Delta X^j\| = 1\big\}.$$

It follows from (4.3) that

$$\|\Delta\overline{X}^j\| \ge (1-\eta)\frac{\|\operatorname{grad}\overline{E}(\overline{X}^j)\|}{\|\operatorname{Hess} E(X^j)\|}, \tag{4.4}$$

where $\|\mathrm{Hess}\, E(X^j)\| > 0$ for all j. Otherwise, if $\|\mathrm{Hess}\, E(X^j)\| = 0$ for some j, then by (4.3), we have $\|\operatorname{grad} \overline{E}(\overline{X}^j)\| = 0$. Thus X^j is a stationary point of E and the algorithm stops.

Now, we note that there exist two constants $c_1, c_2 > 0$ such that

$$0 < c_1 \leq \|\Delta \overline{X}^j\| \leq c_2,$$

for all j. In fact, if there exists some subsequence $\{\|\Delta \overline{X}^j\|\}_{\mathcal{K}} \to 0$, then we have by (4.4) that $\{\|\operatorname{grad} \overline{E}(\overline{X}^j)\|\}_{\mathcal{K}} \to 0$, since $\|\mathrm{Hess}\, E(X^j)\|$ is bounded for the bounded sequence $\{\overline{X}^j\}_{\mathcal{K}}$. By continuity, we get $\operatorname{grad} \overline{E}(\overline{X}_*) = 0$, a contradiction. On the other hand, $\{\overline{X}^j\}$ can not be unbounded because, taking into account of the boundedness of $\{\|\operatorname{grad} \overline{E}(\overline{X}^j)\|\}$, this would contradict (3.11).

We observe from (3.12) that the sequence $\{\overline{E}(\overline{X}^j) \geq 0\}$ is monotonically nonincreasing, and thus is convergent. Hence,

$$\lim_{j\to\infty} \Big[\overline{E}(\overline{X}^j) - \overline{E}(\overline{X}^{j+1})\Big] = 0. \tag{4.5}$$

By (3.11), (3.12), and (4.4), we have

$$\overline{E}(\overline{X}^j) - \overline{E}(\overline{X}^{j+1}) \geq -\sigma\beta^{l_j}\big\langle \operatorname{grad} \overline{E}(\overline{X}^j), \Delta \overline{X}^j\big\rangle \geq \sigma(1-\eta)^2\beta^{l_j}\eta_j \frac{\|\operatorname{grad} \overline{E}(\overline{X}^j)\|^2}{\|\mathrm{Hess}\, E(X^j)\|^2} \geq 0,$$

which, together with (4.5), implies

$$\lim_{j\to\infty} \beta^{l_j}\eta_j \|\operatorname{grad} \overline{E}(\overline{X}^j)\|^2 = 0.$$

This implies that $\liminf \beta^{l_j} = 0$. Otherwise, if $\liminf \beta^{l_j} > 0$, then, by the definition of η_j, we have $\operatorname{grad} \overline{E}(\overline{X}_*) = 0$, a contradiction. Therefore, we may assume that $\lim \beta^{l_j} = 0$, taking a subsequence if necessary. Then we get by (3.12),

$$\overline{E}\Big(\overline{R}_{\overline{X}^j}\Big(\frac{\beta^{l_j}\|\Delta \overline{X}^j\|}{\beta}\frac{\Delta \overline{X}^j}{\|\Delta \overline{X}^j\|}\Big)\Big) - \overline{E}(\overline{X}^j) > \sigma\frac{\beta^{l_j}\|\Delta \overline{X}^j\|}{\beta}\Big\langle \operatorname{grad} \overline{E}(\overline{X}^j), \frac{\Delta \overline{X}^j}{\|\Delta \overline{X}^j\|}\Big\rangle,$$

it follows that

$$\frac{\widehat{\overline{E}}_{\overline{X}_j}\Big(\frac{\beta^{l_j}\|\Delta \overline{X}^j\|}{\beta}\frac{\Delta \overline{X}^j}{\|\Delta \overline{X}^j\|}\Big) - \widehat{\overline{E}}_{\overline{X}_j}(0_{\overline{X}_j})}{\frac{\beta^{l_j}\|\Delta \overline{X}^j\|}{\beta}} > \sigma\Big\langle \operatorname{grad} \overline{E}(\overline{X}^j), \frac{\Delta \overline{X}^j}{\|\Delta \overline{X}^j\|}\Big\rangle, \tag{4.6}$$

where $\widehat{\overline{E}} = \overline{E} \circ \overline{R}$ means the pullback of $\overline{E}$ through the retraction $\overline{R}$ on $\mathrm{St}(k, n)$. Since $\Delta \overline{X}^j/\|\Delta \overline{X}^j\|$ has unit norm, we may assume that $\{\Delta \overline{X}^j/\|\Delta \overline{X}^j\|\}$ converges to $\overline{\xi}_*$ with $\|\overline{\xi}_*\| = 1$, taking a subsequence if necessary. By continuity of the Riemannian metric $\langle\cdot,\cdot\rangle$ and (4.6), we obtain

$$\big\langle \operatorname{grad} \overline{E}(\overline{X}_*), \overline{\xi}_*\big\rangle \geq \sigma\big\langle \operatorname{grad} \overline{E}(\overline{X}_*), \overline{\xi}_*\big\rangle,$$

and then

$$\big\langle \operatorname{grad} \overline{E}(\overline{X}_*), \overline{\xi}_*\big\rangle \geq 0, \tag{4.7}$$

since $0 < \sigma < 1$. By (3.11) and (4.4), we have

$$\Big\langle \operatorname{grad} \overline{E}(\overline{X}^j), \frac{\Delta \overline{X}^j}{\|\Delta \overline{X}^j\|} \Big\rangle \le -\eta_j \|\Delta \overline{X}^j\|. \tag{4.8}$$

Note that the sequence $\{\|\Delta \overline{X}^j\|\}$ is bounded below and $\operatorname{grad} \overline{E}(\overline{X}_*) \neq 0$ by assumption. Hence, we may assume that $\{\|\Delta \overline{X}^j\|\} \to \|\Delta \overline{X}_*\|$, taking a subsequence if necessary. Then, in (4.8) as $j \to \infty$, we get

$$\big\langle \operatorname{grad} \overline{E}(\overline{X}_*), \bar{\xi}_* \big\rangle \le -\min\{\eta, \|\operatorname{grad} \overline{E}(\overline{X}_*)\|\} \|\Delta \overline{X}_*\| < 0,$$

which contradicts (4.7). Therefore, $\|\operatorname{grad} \overline{E}(\overline{X}_*)\| = 0$. The proof is complete. □

4.2. Quadratic convergence. We establish the quadratic convergence of Algorithm 3.2. To do so, we need the following positive definiteness assumption on the Riemannian Hessian of E.

Assumption 4.2. *The Riemannian Hessian operator* $\operatorname{Hess} E([\overline{X}_*]) : T_{[\overline{X}_*]}\mathcal{Q} \to T_{[\overline{X}_*]}\mathcal{Q}$ *is positive definite, where* $\overline{X}_*$ *is an accumulation point of the sequence* $\{\overline{X}^j\}$ *generated by Algorithm* 3.2.

Assumption 4.2 guarantees that a stationary point $X_* := [\overline{X}_*]$ of E is an isolated local minimum point of E. In section 5 we provide a sufficient condition such that Assumption 4.2 is satisfied.

To establish the quadratic convergence of Algorithm 3.2, we need the following result.

LEMMA 4.3. *Let* $\overline{X}_*$ *be an accumulation point of the sequence* $\{\overline{X}^j\}$ *generated by Algorithm* 3.2*, i.e.,* $\overline{X}_*$ *is a limit point of a subsequence* $\{\overline{X}^j\}_{\mathcal{K}}$*. Suppose that Assumption* 4.2 *is satisfied. Then there exist two constants* $d_1, d_2 > 0$ *such that for all* $j \in \mathcal{K}$ *sufficiently large, it holds*

$$d_1 \|\operatorname{grad} E([\overline{X}^j])\| \le \|\Delta X^j\| \le d_2 \|\operatorname{grad} E([\overline{X}^j])\|.$$

Proof. Let $X_* := [\overline{X}_*]$ and $X^j := [\overline{X}^j]$ for all j. As $\{\overline{X}^j\}_{\mathcal{K}} \to \overline{X}_*$, we get $\{X^j\}_{\mathcal{K}} \to X_*$. By Assumption 4.2, there exist two scalars $\kappa_0, \kappa_1 > 0$ such that for all $j \in \mathcal{K}$ sufficiently large, $\operatorname{Hess} E(X^j)$ is nonsingular, and

$$\|\operatorname{Hess} E(X^j)\| \le \kappa_0, \qquad \|[\operatorname{Hess} E(X^j)]^{-1}\| \le \kappa_1. \tag{4.9}$$

By (4.9), we have for all $j \in \mathcal{K}$ sufficiently large,

$$\begin{aligned}
\|\Delta X^j\| &= \|[\operatorname{Hess} E(X^j)]^{-1}\big(\operatorname{Hess} E(X^j)[\Delta X^j] + \operatorname{grad} E(X^j) - \operatorname{grad} E(X^j)\big)\| \\
&\le \|[\operatorname{Hess} E(X^j)]^{-1}\|\big(\|\operatorname{Hess} E(X^j)[\Delta X^j] + \operatorname{grad} E(X^j)\| + \|\operatorname{grad} E(X^j)\|\big) \\
&\le \kappa_1(1+\eta_j)\|\operatorname{grad} E(X^j)\| \le \kappa_1(1+\eta)\|\operatorname{grad} E(X^j)\| \equiv d_2\|\operatorname{grad} E(X^j)\|,
\end{aligned}$$

and

$$\begin{aligned}
\|\operatorname{grad} E(X^j)\| &= \|\operatorname{Hess} E(X^j)[\Delta X^j] + \operatorname{grad} E(X^j) - \operatorname{Hess} E(X^j)[\Delta X^j]\| \\
&\le \|\operatorname{Hess} E(X^j)[\Delta X^j] + \operatorname{grad} E(X^j)\| + \|\operatorname{Hess} E(X^j)[\Delta X^j]\| \\
&\le \eta_j\|\operatorname{grad} E(X^j)\| + \|\operatorname{Hess} E(X^j)\| \cdot \|\Delta X^j\| \\
&\le \eta\|\operatorname{grad} E(X^j)\| + \kappa_0\|\Delta X^j\|.
\end{aligned}$$

Thus for all $j \in \mathcal{K}$ sufficiently large,

$$\|\Delta X^j\| \geq \frac{1-\eta}{\kappa_0}\|\operatorname{grad} E(X^j)\| \equiv d_1\|\operatorname{grad} E(X^j)\|.$$

The proof is complete. □

On the local convergence of Algorithm 3.2 related to the nondegenerate local minima, we have the following result. The proof follows that of Theorem 11(b) in [15].

LEMMA 4.4. *Let $\overline{X}_*$ be an accumulation point of the sequence $\{\overline{X}^j\}$ generated by Algorithm* 3.2. *Suppose that Assumption* 4.2 *holds. Then $\{[\overline{X}^j]\}$ converges to $[\overline{X}_*]$ on $\mathcal{Q}$ defined by* (3.1).

Proof. By Theorem 4.1, we have $\operatorname{grad} E([\overline{X}_*]) = 0$. Also, the Riemannian Hessian operator $\operatorname{Hess} E([\overline{X}_*])$ is positive definite by assumption. Then $X_* := [\overline{X}_*]$ is an isolated local minimum point of E. Let $\mathcal{S}$ be the set of limit points of the sequence $\{X^j := [\overline{X}^j]\}$, which is nonempty since $X_* \in \mathcal{S}$. Suppose that X_* is not the only limit points of the sequence $\{X^j\}$. Then

$$\varsigma := \begin{cases} \inf\limits_{X \in \mathcal{S}\backslash X_*} \big\{\operatorname{dist}(X, X_*)\big\}, & \text{if } \mathcal{S}\backslash X_* \neq \emptyset, \\ 1, & \text{otheriwse.} \end{cases}$$

Since X_* is an isolated local minimizer of E, it follows that $\varsigma > 0$. Define

$$\mathcal{S}_1 := \big\{Y \in \mathcal{Q} \mid \operatorname{dist}(Y|\mathcal{S}) \leq \varsigma/4\big\}, \qquad \mathcal{S}_2 := \big\{Y \in \mathcal{Q} \mid \operatorname{dist}(Y, X_*) \geq \varsigma\big\},$$

where $\operatorname{dist}(Y|\mathcal{S}) := \inf_{X\in\mathcal{S}} \operatorname{dist}(Y, X)$. Then for all j sufficiently large, X^j belongs to at least one of the sets $\mathcal{S}_1$ and $\mathcal{S}_2$. Next, let $\{X^j\}_{\mathcal{K}}$ be a subsequence of $\{X^j\}$ such that $\operatorname{dist}(X^j, X_*) \leq \varsigma/4$ for all $j \in \mathcal{K}$ sufficiently large. Thus, every limit point of $\{X^j\}_{j\in\mathcal{K}}$ lies in the compact set $B_{\frac{\varsigma}{4}}(X_*)$, which is also an accumulation point of the sequence $\{X^j\}$. Hence, $\{X^j\}_{j\in\mathcal{K}}$ converges to X_*, which is the unique accumulation point of $\{X^j\}$ in $B_{\frac{\varsigma}{4}}(X_*)$. By Theorem 4.1 again, $\{\|\operatorname{grad} E(X^j)\|\}_{\mathcal{K}} \to 0$. This, together with Lemma 4.3, yields that $\{\|\Delta X^j\|\}_{\mathcal{K}} \to 0$. Since $\mathcal{Q}$ is a compact manifold, for the retraction R on $\mathcal{Q}$, there exist two scalars $\varrho > 0$ and $\delta_\varrho > 0$ such that [2, p.149],

$$\|\Delta X\| \geq \varrho \operatorname{dist}(X, R_X\Delta X), \quad \text{for all } X \in \mathcal{Q}, \text{ for all } \Delta X \in T_X\mathcal{Q}, \ \|\Delta X\| \leq \delta_\varrho. \tag{4.10}$$

Notice that $\|\Delta X^j\| \leq \min\{\varsigma/4, \delta_\varrho, (\varrho\varsigma)/4\}$ for all $j \in \mathcal{K}$ sufficiently large. Let $\hat{j} \in \mathcal{K}$ be sufficiently large. Then, by using $X^{\hat{j}+1} := [\overline{X}^{\hat{j}+1}] = [\overline{R}_{\overline{X}^{\hat{j}}}(\beta^{l_{\hat{j}}}\Delta\overline{X}^{\hat{j}})Q^{\hat{j}}] = R_{X^{\hat{j}}}(\beta^{l_{\hat{j}}}\Delta X^{\hat{j}})$ and (4.10), we obtain

$$\begin{aligned} \operatorname{dist}(X^{\hat{j}+1}|\mathcal{S}\backslash X_*) &\geq \inf_{Y\in\mathcal{S}\backslash X_*}\{\operatorname{dist}(Y, X_*)\} - \operatorname{dist}(X^{\hat{j}+1}, X^{\hat{j}}) - \operatorname{dist}(X^{\hat{j}}, X_*) \\ &= \inf_{Y\in\mathcal{S}\backslash X_*}\{\operatorname{dist}(Y, X_*)\} - \operatorname{dist}(R_{X^{\hat{j}}}(\beta^{l_{\hat{j}}}\Delta X^{\hat{j}}), X^{\hat{j}}) - \operatorname{dist}(X^{\hat{j}}, X_*) \\ &\geq \inf_{Y\in\mathcal{S}\backslash X_*}\{\operatorname{dist}(Y, X_*)\} - \frac{1}{\varrho}\|\Delta X^{\hat{j}}\| - \operatorname{dist}(X^{\hat{j}}, X_*) \\ &\geq \varsigma - \varsigma/4 - \varsigma/4 = \varsigma/2, \end{aligned}$$

which shows $X^{\hat{j}+1} \notin \mathcal{S}_1\backslash B_{\frac{\varsigma}{4}}(X_*)$.

By using $\|\Delta X^{\hat{j}}\| \le \min\{\varsigma/4, \delta_\varrho, (\varrho\varsigma)/4\}$, $X^{\hat{j}+1} := R_{X^{\hat{j}}}(\beta^{l_{\hat{j}}}\Delta X^{\hat{j}})$, and (4.10) again, we get

$$\begin{aligned}\operatorname{dist}(X^{\hat{j}+1}, X_*) &\le \operatorname{dist}(X^{\hat{j}+1}, X^{\hat{j}}) + \operatorname{dist}(X^{\hat{j}}, X_*)\\ &\le \operatorname{dist}(R_{X^{\hat{j}}}(\beta^{l_{\hat{j}}}\Delta X^{\hat{j}}), X^{\hat{j}}) + \operatorname{dist}(X^{\hat{j}}, X_*)\\ &\le \frac{1}{\varrho}\|\Delta X^{\hat{j}}\| + \operatorname{dist}(X^{\hat{j}}, X_*)\\ &\le \varsigma/4 + \varsigma/4 = \varsigma/2,\end{aligned}$$

which implies $X^{\hat{j}+1} \notin \mathcal{S}_2$. Hence, $X^{\hat{j}+1} \in B_{\frac{\varsigma}{4}}(X_*)$. By definition, we derive that $\hat{j}+1 \in \mathcal{K}$. Therefore, by induction, we conclude that $j \in \mathcal{K}$ for all j sufficiently large and then the whole sequence $\{X^j\}$ converges to X_*. □

On the stepsize β^{l_j} in (3.12), we have the following result similar to Proposition 5 in [36].

LEMMA 4.5. *Let $\overline{X}_*$ be an accumulation point of the sequence $\{\overline{X}^j\}$ generated by Algorithm* 3.2. *Suppose that Assumption* 4.2 *holds, then for j sufficiently large, $l_j = 0$ satisfies* (3.12).

Proof. Let $X_* := [\overline{X}_*]$ and $X^j := [\overline{X}^j]$ for all j. Let $\Delta\overline{X}^j_N$ be the exact solution of the Newton equation (3.9). Then we have

$$\overline{\operatorname{Hess} E(X^j)[\Delta X^j - \Delta X^j_N]}_{\overline{X}^j} = \operatorname{grad}\overline{E}(\overline{X}^j) + \overline{\operatorname{Hess} E(X^j)[\Delta X^j]}_{\overline{X}^j},$$

and thus

$$\operatorname{Hess} E(X^j)[\Delta X^j - \Delta X^j_N] = \operatorname{grad} E(X^j) + \operatorname{Hess} E(X^j)[\Delta X^j]. \tag{4.11}$$

According to (4.1) and (4.2), we have

$$\operatorname{grad}\widehat{E}_{X^j}(0_{X^j}) + \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})[\Delta X^j_N] = \operatorname{grad} E(X^j) + \operatorname{Hess} E(X^j)[\Delta X^j_N] = 0_{X^j}. \tag{4.12}$$

By Lemma 4.4, we have $X^j \to X_*$. Thus, by Lemma 4.3, (3.10), (4.9), and (4.11), we have for all j sufficiently large,

$$\begin{aligned}\|\Delta X^j - \Delta X^j_N\| &= \|[\operatorname{Hess} E(X^j)]^{-1}(\operatorname{grad} E(X^j) + \operatorname{Hess} E(X^j)[\Delta X^j])\|\\ &\le \|[\operatorname{Hess} E(X^j)]^{-1}\| \cdot \|\operatorname{grad} E(X^j) + \operatorname{Hess} E(X^j)[\Delta X^j]\|\\ &\le \kappa_1\eta_j\|\operatorname{grad} E(X^j)\| \le \kappa_1\|\operatorname{grad} E(X^j)\|^2 \le \frac{\kappa_1}{d_1^2}\|\Delta X^j\|^2.\end{aligned} \tag{4.13}$$

In addition, Hess $\widehat{E}_X$ is Lipschitz-continuous at 0_X uniformly in a neighborhood of X_*, i.e., there exist scalars $\kappa_2 > 0$, $\delta_1 > 0$, and $\delta_2 > 0$, such that for all $X \in B_{\delta_1}(X_*)$ and all $\xi \in B_{\delta_2}(0_X)$, it holds

$$\|\operatorname{Hess}\widehat{E}_X(\xi) - \operatorname{Hess}\widehat{E}_X(0_X)\| \le \kappa_2\|\xi\|. \tag{4.14}$$

By Taylor's theorem, there exists some constant $\theta \in [0,1]$ such that

$$\widehat{E}_{X^j}(\Delta X^j) = \widehat{E}_{X^j}(0_{X^j}) + \big\langle \operatorname{grad}\widehat{E}_{X^j}(0_{X^j}), \Delta X^j\big\rangle + \frac{1}{2}\big\langle \operatorname{Hess}\widehat{E}_{X^j}(\theta\Delta X^j)[\Delta X^j], \Delta X^j\big\rangle.$$

By using (3.10), (4.9), (4.11), (4.12), (4.13), and (4.14), for all j sufficiently large,

$$\begin{aligned}
&\overline{E}\big(\overline{R}_{\overline{X}^j}(\Delta\overline{X}^j)\big) - \overline{E}(\overline{X}^j) - \frac{1}{2}\langle \operatorname{grad}\overline{E}(\overline{X}^j), \Delta\overline{X}^j\rangle \\
&= E(R_{X^j}(\Delta X^j)) - E(X^j) - \frac{1}{2}\big\langle \operatorname{grad} E(X^j), \Delta X^j\big\rangle \\
&= \widehat{E}_{X^j}(\Delta X^j) - \widehat{E}_{X^j}(0_{X^j}) - \frac{1}{2}\big\langle \operatorname{grad}\widehat{E}_{X^j}(0_{X^j}), \Delta X^j\big\rangle \\
&= \frac{1}{2}\big\langle \operatorname{grad}\widehat{E}_{X^j}(0_{X^j}) + \operatorname{Hess}\widehat{E}_{X^j}(\theta\Delta X^j)[\Delta X^j], \Delta X^j\big\rangle \\
&= \frac{1}{2}\big\langle \operatorname{grad}\widehat{E}_{X^j}(0_{X^j}) + \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})[\Delta X_N^j], \Delta X^j\big\rangle \\
&\quad + \frac{1}{2}\big\langle \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})[\Delta X^j - \Delta X_N^j], \Delta X^j\big\rangle \\
&\quad + \frac{1}{2}\big\langle \operatorname{Hess}\widehat{E}_{X^j}(\theta\Delta X^j)[\Delta X^j] - \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})[\Delta X^j], \Delta X^j\big\rangle \\
&\le 0 + \frac{1}{2}\|\operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})[\Delta X^j - \Delta X_N^j]\| \cdot \|\Delta X^j\| \\
&\quad + \frac{1}{2}\|\operatorname{Hess}\widehat{E}_{X^j}(\theta\Delta X^j) - \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})\| \cdot \|\Delta X^j\|^2 \\
&\le \frac{1}{2}\|\operatorname{Hess} E(X^j)\| \cdot \|\Delta X^j - \Delta X_N^j\| \cdot \|\Delta X^j\| + \frac{1}{2}\kappa_2\|\Delta X^j\| \cdot \|\Delta X^j\|^2 \\
&\le \frac{1}{2}\Big(\frac{\kappa_1\kappa_0}{d_1^2} + \kappa_2\Big)\|\Delta X^j\|^3 = \frac{1}{2}\Big(\frac{\kappa_1\kappa_0}{d_1^2} + \kappa_2\Big)\|\Delta\overline{X}^j\|^3 = c\|\Delta\overline{X}^j\|^3,
\end{aligned}$$

where c is a constant. This shows that (3.12) holds with $l_j = 0$ for all j sufficiently large. The proof is complete. □

We now establish the quadratic convergence of Algorithm 3.2.

THEOREM 4.6. *Let $\overline{X}_*$ be an accumulation point of the sequence $\{\overline{X}^j\}$ generated by Algorithm* 3.2. *Suppose that Assumption* 4.2 *holds. Then the whole sequence $\{[\overline{X}^j]\}$ converges to $[\overline{X}_*]$ quadratically.*

Proof. Let $X_* := [\overline{X}_*]$ and $X^j := [\overline{X}^j]$ for all j. By Lemma 4.4, we have $X^j \to X_*$. By Lemma 4.5, we obtain $X^{j+1} := [\overline{X}^{j+1}] = [\overline{R}_{\overline{X}^j}(\Delta\overline{X}^j)Q^j] = R_{X^j}(\Delta X^j)$ for all j sufficiently large. Then, by Lemmas 2.2–2.3, there exist three scalars $\tau_0, \tau_1, \tau_2 > 0$ such that for all j sufficiently large,

(4.15)
$$\begin{cases}
\tau_0 \operatorname{dist}(X^j, X_*) & \le \|\operatorname{grad} E(X^j)\| = \|\operatorname{grad}\overline{E}(\overline{X}^j)\| \le \tau_1 \operatorname{dist}(X^j, X_*), \\
\tau_0 \operatorname{dist}(X^{j+1}, X_*) & \le \|\operatorname{grad} E(X^{j+1})\| = \|\operatorname{grad} E(R_{X^j}(\Delta X^j))\| \\
& \le \tau_2\|\operatorname{grad}\widehat{E}_{X^j}(\Delta X^j)\|.
\end{cases}$$

We get by Taylor's formula for all j sufficiently large,

$$\begin{aligned}
\operatorname{grad}\widehat{E}_{X^j}(\Delta X^j) &= \operatorname{grad}\widehat{E}_{X^j}(0_{X^j}) + \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})[\Delta X^j] \\
&\quad + \int_0^1 \big(\operatorname{Hess}\widehat{E}_{X^j}(t\Delta X^j) - \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})\big)[\Delta X^j]dt.
\end{aligned} \tag{4.16}$$

From Lemma 4.3, (3.10), (4.14), (4.15), and (4.16), it follows that

$$\begin{aligned}
&\frac{\tau_0}{\tau_2}\operatorname{dist}(X^{j+1},X_*) \le \|\operatorname{grad}\widehat{E}_{X^j}(\Delta X^j)\| \\
&\le \|\operatorname{grad}\widehat{E}_{X^j}(0_{X^j}) + \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j})[\Delta X^j]\| \\
&\quad +\|\int_0^1 (\operatorname{Hess}\widehat{E}_{X^j}(t\Delta X^j) - \operatorname{Hess}\widehat{E}_{X^j}(0_{X^j}))[\Delta X^j]dt\| \\
&\le \|\operatorname{grad}E(X^j) + \operatorname{Hess}E(X^j)[\Delta X^j]\| + \kappa_2\|\Delta X^j\|^2 \\
&\le \eta_j\|\operatorname{grad}E(X^j)\| + \kappa_2\|\Delta X^j\|^2 \\
&\le \|\operatorname{grad}E(X^j)\|^2 + \kappa_2 d_2^2\|\operatorname{grad}E(X^j)\|^2 \\
&\le \tau_1^2(1+\kappa_2 d_2^2)\operatorname{dist}(X^j,X_*)^2. \qquad (4.17)
\end{aligned}$$

Thus

$$\operatorname{dist}(X^{j+1},X_*) \le \frac{\tau_2\tau_1^2}{\tau_0}\left(1+\kappa_2 d_2^2\right)\operatorname{dist}(X^j,X_*)^2 = c_0\operatorname{dist}(X^j,X_*)^2,$$

where c_0 is a constant. This completes the proof. □

As a direct consequence of (4.15) and (4.17), we can derive the following result.

COROLLARY 4.7. *Let $\overline{X}_*$ be an accumulation point of the sequence $\{\overline{X}^j\}$ generated by Algorithm* 3.2. *Suppose that Assumption* 4.2 *holds. Then there exists a constant $c>0$ such that for all j sufficiently large,*

$$\|\operatorname{grad}E([\overline{X}^{j+1}])\| \le c\|\operatorname{grad}E([\overline{X}^j])\|^2.$$

5. Positive Definiteness Condition. In this section, we give the positive definiteness condition of $\operatorname{Hess}E([\overline{X}_*])$ defined in Assumption 4.2, where $\overline{X}_*$ is an accumulation point of the sequence $\{\overline{X}^j\}$ generated by Algorithm 3.2. By using the Riemannian gradient and Riemannian Hessian of E given in (3.7) and (3.8), we can establish the following result on the positive definiteness of $\operatorname{Hess}E([\overline{X}_*])$.

THEOREM 5.1. *Let $\overline{X}_*$ be an accumulation point of the sequence $\{\overline{X}^j\}$ generated by Algorithm* 3.2, *and $\overline{X}_*$ is a global minimizer of problem* (1.1). *Let $\bar{\lambda}_1 \le \bar{\lambda}_2 \le \cdots \le \bar{\lambda}_n$ be eigenvalues of $H(\overline{X}_*)$. If $\bar{\lambda}_{k+1} > \bar{\lambda}_k$ and $\alpha \ge 0$, then $\operatorname{Hess}E([\overline{X}_*])$ is positive definite, i.e.,*

$$\left\langle \operatorname{Hess}E([\overline{X}_*])[Z_{[\overline{X}_*]}], Z_{[\overline{X}_*]}\right\rangle > 0, \quad \forall Z_{[\overline{X}_*]} \in T_{[\overline{X}_*]}\mathcal{Q}\backslash\{0_{[\overline{X}_*]}\}.$$

Proof. Let $X_* := [\overline{X}_*]$. Since

$$\left\langle \operatorname{Hess}E(X_*)[Z_{X_*}], Z_{X_*}\right\rangle = \left\langle \overline{\operatorname{Hess}E(X_*)[Z_{X_*}]}_{\overline{X}_*}, \overline{Z}_{\overline{X}_*}\right\rangle, \quad \forall Z_{X_*} \in T_{X_*}\mathcal{Q}\backslash\{0_{X_*}\},$$

$\operatorname{Hess}E(X_*)$ is positive definite if and only if

$$\left\langle \overline{\operatorname{Hess}E(X_*)[Z_{X_*}]}_{\overline{X}_*}, \overline{Z}_{\overline{X}_*}\right\rangle > 0, \quad \forall \overline{Z}_{\overline{X}_*} \in \mathcal{H}_{\overline{X}_*}\backslash\{0_{\overline{X}_*}\}. \qquad (5.1)$$

By (3.8), we have

$$\overline{\operatorname{Hess}E(X_*)[Z_{X_*}]}_{\overline{X}_*} = T_1(\overline{Z}_{\overline{X}_*}) + T_2(\overline{Z}_{\overline{X}_*}) + T_3(\overline{Z}_{\overline{X}_*}), \quad \forall \overline{Z}_{\overline{X}_*} \in \mathcal{H}_{\overline{X}_*},$$

where for each $\overline{Z}_{\overline{X}_*} \in \mathcal{H}_{\overline{X}_*}$,

$$\begin{cases} T_1(\overline{Z}_{\overline{X}_*}) &= -(I_n - \overline{X}_*\overline{X}_*^T)\overline{Z}_{\overline{X}_*}\overline{X}_*^T H(\overline{X}_*)\overline{X}_*, \\ T_2(\overline{Z}_{\overline{X}_*}) &= 2\alpha(I_n - \overline{X}_*\overline{X}_*^T)\mathrm{Diag}\big(L^{-1}\mathrm{diag}(\overline{Z}_{\overline{X}_*}\overline{X}_*^T)\big)\overline{X}_*, \\ T_3(\overline{Z}_{\overline{X}_*}) &= (I_n - \overline{X}_*\overline{X}_*^T)H(\overline{X}_*)\overline{Z}_{\overline{X}_*}. \end{cases}$$

To verify (5.1), we only need to show that

$$\big\langle T_1(\overline{Z}_{\overline{X}_*}), \overline{Z}_{\overline{X}_*}\big\rangle + \big\langle T_2(\overline{Z}_{\overline{X}_*}), \overline{Z}_{\overline{X}_*}\big\rangle + \big\langle T_3(\overline{Z}_{\overline{X}_*}), \overline{Z}_{\overline{X}_*}\big\rangle > 0, \tag{5.2}$$

for all $\overline{Z}_{\overline{X}_*} \in \mathcal{H}_{\overline{X}_*}\backslash\{0_{\overline{X}_*}\}$. Since $\overline{X}_*$ is a global minimizer of problem (1.1), $\overline{X}_*$ is the matrix consisting of the k eigenvectors corresponding to the k smallest eigenvalues of $H(\overline{X}_*)$. Let

$$\overline{\Lambda}_1 = \mathrm{Diag}(\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_k), \qquad \overline{\Lambda}_2 = \mathrm{Diag}(\overline{\lambda}_{k+1}, \overline{\lambda}_{k+2}, \ldots, \overline{\lambda}_n).$$

It follows that

$$H\big(\overline{X}_*\big)\overline{X}_* = \overline{X}_*\overline{\Lambda}_1, \qquad H\big(\overline{X}_*\big)\overline{X}_{*\perp} = \overline{X}_{*\perp}\overline{\Lambda}_2, \tag{5.3}$$

where $\overline{X}_{*\perp}$ denotes the orthogonal complement matrix of $\overline{X}_*$ with ${\overline{X}_{*\perp}}^T\overline{X}_{*\perp} = I_{n-k}$. In addition, for any $Z_{X_*} \in T_{X_*}\mathcal{Q}$, we have $\overline{Z}_{\overline{X}_*} \in \mathcal{H}_{\overline{X}_*}$, i.e.,

$$\overline{Z}_{\overline{X}_*} = \overline{X}_{*\perp}K, \quad K \in \mathbb{R}^{(n-k)\times k}. \tag{5.4}$$

Thus T_i are the functions of the variable $K \in \mathbb{R}^{(n-k)\times k}$ for $i = 1, 2, 3$, and $\overline{Z}_{\overline{X}_*} = 0$ if and only if $0 = K \in \mathbb{R}^{(n-k)\times k}$, where 0 is a zero matrix. It follows from (5.3) and (5.4) that for each $K \in \mathbb{R}^{(n-k)\times k}$,

$$\begin{cases} T_1(K) &= -(I_n - \overline{X}_*\overline{X}_*^T)\overline{X}_{*\perp}K\overline{\Lambda}_1 = -\overline{X}_{*\perp}K\overline{\Lambda}_1, \\ T_2(K) &= 2\alpha(I_n - \overline{X}_*\overline{X}_*^T)\mathrm{Diag}\big(L^{-1}\mathrm{diag}(\overline{X}_{*\perp}K\overline{X}_*^T)\big)\overline{X}_*, \\ T_3(K) &= (I_n - \overline{X}_*\overline{X}_*^T)H(\overline{X}_*)\overline{X}_{*\perp}K = (I_n - \overline{X}_*\overline{X}_*^T)\overline{X}_{*\perp}\overline{\Lambda}_2K = \overline{X}_{*\perp}\overline{\Lambda}_2K. \end{cases}$$

We recall that the meaning of the notation $\mathrm{diag}(\cdot)$ is different from that of the notation $\mathrm{Diag}(\cdot)$ (see section 1). Hence, (5.2) holds if and only if

$$\big\langle T_1(K), \overline{X}_{*\perp}K\big\rangle + \big\langle T_2(K), \overline{X}_{*\perp}K\big\rangle + \big\langle T_3(K), \overline{X}_{*\perp}K\big\rangle > 0, \tag{5.5}$$

for all $K \in \mathbb{R}^{(n-k)\times k}\backslash\{0\}$. By simple calculation, we have for each $K \in \mathbb{R}^{(n-k)\times k}$,

$$\big\langle T_1(K), \overline{X}_{*\perp}K\big\rangle = -\big\langle \overline{X}_{*\perp}K\overline{\Lambda}_1, \overline{X}_{*\perp}K\big\rangle = -\big\langle K, K\overline{\Lambda}_1\big\rangle, \tag{5.6}$$

$$\begin{aligned} \big\langle T_2(K), \overline{X}_{*\perp}K\big\rangle &= 2\alpha\big\langle (I_n - \overline{X}_*\overline{X}_*^T)\mathrm{Diag}\big(L^{-1}\mathrm{diag}(\overline{X}_{*\perp}K\overline{X}_*^T)\big)\overline{X}_*, \overline{X}_{*\perp}K\big\rangle \\ &= 2\alpha\big\langle \mathrm{Diag}\big(L^{-1}\mathrm{diag}(\overline{X}_{*\perp}K\overline{X}_*^T)\big), \overline{X}_{*\perp}K\overline{X}_*^T\big\rangle \\ &= 2\alpha\big\langle \mathrm{diag}(\overline{X}_{*\perp}K\overline{X}_*^T), L^{-1}\mathrm{diag}(\overline{X}_{*\perp}K\overline{X}_*^T)\big\rangle, \end{aligned} \tag{5.7}$$

$$\big\langle T_3(K), \overline{X}_{*\perp}K\big\rangle = \big\langle \overline{X}_{*\perp}\overline{\Lambda}_2K, \overline{X}_{*\perp}K\big\rangle = \big\langle K, \overline{\Lambda}_2K\big\rangle. \tag{5.8}$$

Since L is the Laplacian operator and $\alpha \geq 0$, we have by (5.7),

$$\big\langle T_2(K), \overline{X}_{*\perp} K \big\rangle = 2\alpha \big\langle \mathrm{diag}(\overline{X}_{*\perp} K \overline{X}_*^T), L^{-1}\mathrm{diag}(\overline{X}_{*\perp} K \overline{X}_*^T) \big\rangle \geq 0, \tag{5.9}$$

for all $K \in \mathbb{R}^{(n-k)\times k}\backslash\{0\}$. For any given $A_1 \in \mathbb{R}^{m_1\times m_2}$, $A_2 \in \mathbb{R}^{m_2\times m_3}$, and $A_3 \in \mathbb{R}^{m_3\times m_1}$, we have [8, Fact 7.4.8],

$$\mathrm{tr}(A_1A_2A_3) = \mathrm{vec}(A_1)^T(A_2 \otimes I_{m_1})\mathrm{vec}(A_3^T),$$

where the vec operator creates a column vector from a matrix by stacking its column vectors below one another, and $A \otimes B = [a_{ij}B] \in \mathbb{R}^{mp\times nq}$ for $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$. Also, the vectorized transpose matrix $P \in \mathbb{R}^{(n-k)k\times(n-k)k}$ is such that [22, Theorem 4.3.8],

$$\mathrm{vec}(K^T) = P\,\mathrm{vec}(K), \quad \forall K \in \mathbb{R}^{(n-k)\times k}.$$

This, together with (5.6), (5.8), and Corollary 4.3.10 in [22], gives rise to

$$\big\langle T_1(K), \overline{X}_{*\perp} K \big\rangle = -\big\langle K, K\overline{\Lambda}_1 \big\rangle = -\mathrm{tr}(K\overline{\Lambda}_1 K^T) = -\mathrm{vec}(K)^T\big(\overline{\Lambda}_1 \otimes I_{n-k}\big)\mathrm{vec}(K),$$

and

$$\begin{aligned} &\big\langle T_3(K), \overline{X}_{*\perp} K \big\rangle = \big\langle K, \overline{\Lambda}_2 K \big\rangle = \mathrm{tr}(K^T\overline{\Lambda}_2 K) = \mathrm{vec}(K^T)^T(\overline{\Lambda}_2 \otimes I_k)\mathrm{vec}(K^T) \\ = \;& \mathrm{vec}(K)^T\big(P^T(\overline{\Lambda}_2 \otimes I_k)P\big)\mathrm{vec}(K) = \mathrm{vec}(K)^T\big(I_k \otimes \overline{\Lambda}_2\big)\mathrm{vec}(K), \end{aligned}$$

for all $K \in \mathbb{R}^{(n-k)\times k}$. Thus

$$\big\langle T_1(K), \overline{X}_{*\perp} K \big\rangle + \big\langle T_3(K), \overline{X}_{*\perp} K \big\rangle = \mathrm{vec}(K)^T\big(I_k \otimes \overline{\Lambda}_2 - \overline{\Lambda}_1 \otimes I_{n-k}\big)\mathrm{vec}(K), \tag{5.10}$$

for all $K \in \mathbb{R}^{(n-k)\times k}$. The eigenvalues of $I_k \otimes \overline{\Lambda}_2 - \overline{\Lambda}_1 \otimes I_{n-k}$ are [22, Theorem 4.4.5]:

$$\overline{\lambda}_i - \overline{\lambda}_j, \quad i = k+1, \ldots, n,\ j = 1, \ldots, k.$$

By assumption,

$$\bar{\lambda}_1 \leq \bar{\lambda}_2 \leq \cdots \leq \bar{\lambda}_k < \bar{\lambda}_{k+1} \leq \cdots \leq \bar{\lambda}_n.$$

Then we get by (5.10),

$$\big\langle T_1(K), \overline{X}_{*\perp} K \big\rangle + \big\langle T_3(K), \overline{X}_{*\perp} K \big\rangle \geq (\bar{\lambda}_{k+1} - \bar{\lambda}_k)\|K\|^2 > 0,\ \forall K \in \mathbb{R}^{(n-k)\times k}\backslash\{0\}.$$

This, together with (5.2), (5.5), and (5.9), yields (5.1). The proof is complete. □

We point out that in our numerical tests, the positive definiteness condition in Theorem 5.1 holds for most numerical examples and the quadratic convergence is observed, where the starting points are chosen appropriately.

6. Numerical Experiments. In this section, we present numerical performance of Algorithm 3.2 for solving the nonlinear eigenvalue problem (1.2) by the solution of the total energy minimization problem (1.1). To illustrate the efficiency of our method, we compare the proposed algorithm with the trust region SCF (TRSCF) iteration [45], where the trust region parameter is chosen as suggested in [45], which should be based on the gap between the kth and $(k+1)$th eigenvalues of $H(\overline{X}^j)$ at

the current iterate $\overline{X}^j$. For Algorithm 3.2 and the TRSCF iteration, we randomly generate the starting points by the built-in functions `randn`, `svd` and `eigs`:

$$\overline{X} = \texttt{randn}\,(n,k),\ \big[U,S,V\big] = \texttt{svd}\,(\overline{X},0),\ \widehat{X} = UV^T,\ \overline{X}^0 = \texttt{eigs}\,(H(\widehat{X}),k,'\,\mathrm{sm}').$$

The stopping criterion is set to be

$$\|\mathrm{grad}\ \overline{E}(\overline{X}^j)\| < 10^{-6}.$$

In our numerical tests, we set $\eta = 0.1$, $\sigma = 10^{-4}$, $\beta = 1/2$, $Q_j = I_k$ for all j, and L is an n-by-n real symmetric tridiagonal matrix with 2 on its diagonal and -1 on its sub- and super- diagonals. In addition, the largest number of outer iterations in Algorithm 3.2 and the TRSCF iteration is set to be 3000, and the largest number of iterations in the CG method is set to be nk.

For Examples 6.1–6.6 below, we repeat our experiments over 10 different starting points. All the numerical tests are carried out by using `MATLAB` 7.1 on a Dell computer Intel(R) Core(TM)i7-2600 Duo of 3.40 GHz CPU and 16GB RAM. In what follows, '`cputime`', '`IT.`', '`NF.`', '`NCG.`', and '`Res.`' mean the averaged total CPU time in seconds, the averaged number of outer iterations, the averaged number of function evaluations, the averaged number of inner iterations of the CG method, and the averaged residual $\|\mathrm{grad}\ \overline{E}(\overline{X}^j)\|$ at the final iterate of the algorithms, respectively.

It was shown in [46] that the SCF iteration converged for Example 6.1 and failed to converge for Example 6.2. To show the effectiveness of our method over the TRSCF algorithm, we report the numerical results for Examples 6.1–6.3 with different choices of n, k, and α.

Example 6.1. [46] *We consider the nonlinear eigenvalue problem for different choices of* n, k, α*:* (a) $n = 2, k = 1, \alpha = 3$*;* (b) $n = 10, k = 2, \alpha = 0.6$*;* (c) $n = 100, k = 10, \alpha = 0.005$*;* (d) $n = 100, k = 4, \alpha = 0.001$.

Example 6.2. [46] *We consider the nonlinear eigenvalue problem for different choices of* n, k, α*:* (a) $n = 2, k = 1, \alpha = 9$*;* (b) $n = 10, k = 2, \alpha = 3$*;* (c) $n = 100, k = 10, \alpha = 1$*;* (d) $n = 100, k = 4, \alpha = 2$.

Tables 1–2 include numerical results for Examples 6.1–6.2. We observe from Tables 1–2 that both methods converge while our method performs much better than the TRSCF iteration in terms of the CPU time.

Table 1: Numerical results for Example 6.1.

Alg.	(n,k,α)	cputime	`IT.`	`NF.`	`NCG.`	`Res.`
Alg. 3.2	(a)	0.0031 s	2.6	3.6	1.0000	9.0589×10^{-8}
	(b)	0.0078 s	3.2	4.2	6.9063	5.5031×10^{-8}
	(c)	0.0203 s	5.0	6.0	20.6600	9.0158×10^{-8}
	(d)	0.0359 s	4.1	5.6	62.4390	1.7196×10^{-7}
TRSCF	(a)	0.0218 s	18.8	19.8		6.3051×10^{-7}
	(b)	0.1778 s	39.2	40.2		8.5019×10^{-7}
	(c)	2.8798 s	110.0	111.0		9.2728×10^{-7}
	(d)	1.2886 s	77.5	78.5		9.5057×10^{-7}

Example 6.3. *We consider the nonlinear eigenvalue problem for* $k = 10, \alpha = 1$*, and varying* $n = 200, 400, 800, 1000$.

Table 3 gives numerical results for Example 6.3. We see from Table 3 that our method is more efficient than the TRSCF iteration in terms of CPU time.

Table 2: Numerical results for Example 6.2.

Alg.	(n, k, α)	cputime	IT.	NF.	NCG.	Res.
Alg. 3.2	(a)	0.0047 s	4.1	6.7	1.0000	4.0109×10^{-8}
	(b)	0.0062 s	5.0	6.7	4.5200	7.4409×10^{-9}
	(c)	0.0296 s	8.0	10.3	16.9625	3.5290×10^{-10}
	(d)	0.0172 s	7.0	10.0	10.6571	2.6830×10^{-11}
TRSCF	(a)	0.0328 s	30.2	31.2		7.8182×10^{-7}
	(b)	0.1654 s	33.0	34.0		9.2415×10^{-7}
	(c)	6.7330 s	83.6	84.6		9.3917×10^{-7}
	(d)	3.7035 s	53.0	54.0		8.9907×10^{-7}

Table 3: Numerical results for Example 6.3.

Alg.	n	cputime	IT.	NF.	NCG.	Res.
Alg. 3.2	200	0.0343 s	8.0	10.0	16.7875	9.6512×10^{-11}
	400	0.1934 s	7.6	9.6	15.3947	3.8372×10^{-7}
	800	0.3744 s	8.0	10.0	17.4500	9.9671×10^{-12}
	1000	0.4181 s	8.0	10.0	17.4000	1.0358×10^{-11}
TRSCF	200	22.2083 s	85.0	86.0		9.6446×10^{-7}
	400	88.6444 s	86.0	87.0		9.6992×10^{-7}
	800	309.0770 s	87.0	88.0		9.0885×10^{-7}
	1000	322.8768 s	87.0	88.0		9.1770×10^{-7}

In Figure 6.1, we give the convergence history of Algorithms 3.2 for two tests with $(n, k, \alpha) = (100, 10, 0.005)$ and $(n, k, \alpha) = (1000, 10, 1)$. This figure depicts the logarithm of the residual versus the number of iterations for solving the nonlinear eigenvalue problem. We can see from Figure 6.1 that the TRSCF iteration converges slow while our method converges quadratically. This confirms our theoretical results.

Example 6.4. *We consider the nonlinear eigenvalue problem for* $n = 100$ *and* $k = 20$ *and varying* α.

Table 4: Numerical results for Example 6.4.

α	cputime	IT.	NF.	NCG.	Res.
0.0001	0.0437 s	1.0	2.0	37.7000	4.7089×10^{-8}
0.001	0.0811 s	2.0	3.0	36.3000	1.0550×10^{-9}
0.01	0.1482 s	9.2	10.3	14.9891	2.0247×10^{-7}
0.1	0.2371 s	12.3	14.3	16.9512	5.2409×10^{-8}
1	0.1810 s	7.0	9.0	23.3429	1.0884×10^{-9}
20	0.2153 s	6.0	8.0	32.0167	3.4250×10^{-9}
40	0.1560 s	5.0	7.0	29.5000	2.9122×10^{-8}
80	0.1872 s	6.0	8.0	25.9167	1.0831×10^{-9}
100	0.1622 s	6.0	8.0	24.8833	2.2898×10^{-9}

Table 4 lists numerical results for Example 6.4. We observe from Table 4 and other tests that when n and k are fixed, the smaller the parameter α, the smaller the averaged number of outer iterations. This shows that the magnitude of α, which measures the amount of non-linearity, has some effect on the convergence of the proposed method (perhaps in terms of Lipschitz bound constant). On the other hand, the larger the parameter α is, the more nonlinear the problem becomes but there is no obvious increase in the averaged numbers of inner and outer iterations for most numerical tests.

To further illustrate the efficiency of our method, we compare our method with

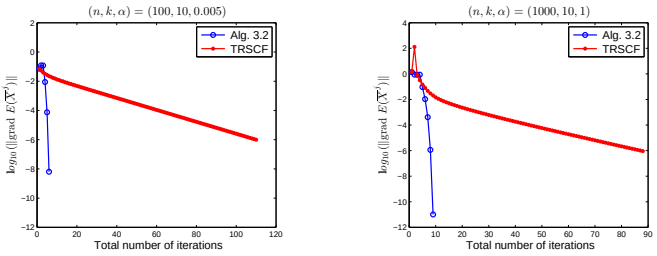

Fig. 6.1: Convergence history of two tests

the classical interior point method (IPM) [12]. For simplicity, as in [19], we call the MATLAB-provided function `fmincon` by

```
[X,RESNRM] = fmincon(@myfun, X0, [], [], [], [], [], [], ...
                     @nonlcon, options, L, k, alpha);
```

Here, the MATLAB function `myfun.m` evaluates the total energy in (1.1) with its gradient. The function `nonlcon.m` computes the nonlinear equality constraints $X^T X - I_k$ with its gradient. `L, k, alpha` are, respectively, matrices and scalar parameters used for function and gradient calculations. The parameter `options` provides other parameters used by the `fmincon` function. Here, `HessX` is a function to compute the product of the Hessian of the Lagrangian and an arbitrary vector.

```
options = optimset('Algorithm','interior-point','MaxIter',5000,...
          'MaxSQPIter',100, 'MaxFunEvals',6000,'GradObj','on',...
          'LargeScale','on','GradConstr','on','Hessian','on',...
          'SubproblemAlgorithm','cg','HessMult',@HessX,...
          'TolX',1.e-8,'TolCon',1.e-8,'TolFun',1.e-8);
```

In our numerical tests, the stopping criterion for Algorithm 3.2 is the same as the IPM. The starting points for Algorithm 3.2 and the interior point method are set as above and the other parameters in Algorithm 3.2 are set as above. For Example 6.5 below, we repeat our experiments over 10 different starting points.

Example 6.5. *We consider the nonlinear eigenvalue problem for varying* n *and* k. *We report numerical results for* (a) $k = 10$, $\alpha = 1$, *and* $n = 200$, 400, 800, 1000; *and* (b) $n = 1000, \alpha = 1$, *and* $k = 10, 20, 30, 40$.

Table 5 lists numerical results for Example 6.5. We see from Table 5 that our method is much more effective than the interior point method.

Finally, we consider the following large-scale nonlinear eigenvalue problem.

Example 6.6. *We consider the nonlinear eigenvalue problem for varying* n *and* k. *We report numerical results for* (a) $k = 20$, $\alpha = 1$, *and* $n = 20000$, 40000, 80000, 100000, 200000, 300000; *and* (b) $n = 10000, \alpha = 1$, *and* $k = 20, 40, 60, 80, 100, 120$.

Table 6 lists numerical results for Example 6.6. Table 6 shows that the proposed algorithm is very efficient for solving large-scale nonlinear eigenvalue problems (the largest numerical examples that we have tested are: (1) $k = 20$ and $n = 300,000$ and (2) $n = 10,000$ and $k = 120$. For case (1), there are $k(n-k) = 5999,600$ unknowns in problem (3.2) while for case (2), the number of unknowns is $k(n-k) = 1185,600$).

Table 5: Numerical results for Example 6.5.

Alg.	n	cputime	IT.	NF.	NCG.	Res.
Alg. 3.2	200	0.0359 s	8.0	10.0	16.9500	7.6187×10^{-11}
	400	0.2387 s	8.0	10.0	17.3000	1.8736×10^{-11}
	800	0.3619 s	8.0	10.0	17.4250	1.0506×10^{-11}
	1000	0.4118 s	8.0	10.0	17.4625	1.0108×10^{-11}
IPM	200	1.0358 s	16.0	260.4	10.5875	8.9404×10^{-9}
	400	9.2306 s	19.0	228.5	6.1316	1.0640×10^{-8}
	800	16.0821 s	16.0	253.7	10.0375	2.2648×10^{-8}
	1000	28.9007 s	26.0	301.8	4.6462	1.5872×10^{-8}
Alg.	k	cputime	IT.	NF.	NCG.	Res.
Alg. 3.2	10	0.4118 s	8.0	10.0	17.4625	1.0108×10^{-11}
	20	0.7145 s	7.0	9.0	20.4286	2.2575×10^{-8}
	30	1.5335 s	7.0	8.0	27.4000	6.8743×10^{-8}
	40	2.7425 s	7.8	8.0	37.6571	4.1741×10^{-8}
IPM	10	28.9007 s	26.0	301.8	4.6462	1.5872×10^{-8}
	20	226.1640 s	60.0	560.6	3.6767	4.4018×10^{-8}
	30	800.1073 s	98.0	850.2	3.3388	5.1536×10^{-8}
	40	2346.3120 s	170.1	1408.7	3.0500	2.4373×10^{-7}

Table 6: Numerical results for Example 6.6.

$k = 20$, $\alpha = 1$					
n	cputime	IT.	NF.	NCG.	Res.
20000	17.7202 s	7.0	8.0	18.3286	1.6802×10^{-7}
40000	36.0019 s	7.0	8.0	18.3000	1.7313×10^{-7}
80000	69.4969 s	7.0	8.0	18.2571	1.7518×10^{-7}
100000	85.0049 s	7.0	8.0	18.2571	1.8369×10^{-7}
200000	152.7234 s	7.0	8.0	18.3286	1.7869×10^{-7}
300000	203.7217 s	7.0	8.0	18.3143	1.8667×10^{-7}
$n = 10000$, $\alpha = 1$					
k	cputime	IT.	NF.	NCG.	Res.
20	8.7017 s	7.0	8.0	18.3714	1.6110×10^{-7}
40	34.4918 s	7.0	8.0	36.9143	5.3578×10^{-8}
60	76.9600 s	7.0	8.0	53.0429	1.0157×10^{-7}
80	131.6056 s	7.0	8.0	65.7286	2.9916×10^{-7}
100	215.4530 s	7.0	8.0	85.6714	2.0304×10^{-7}
120	312.6525 s	7.0	8.0	101.1429	2.9930×10^{-7}

7. Conclusions. This paper is concerned with the solution of nonlinear eigenvalue problems. We propose a Riemannian Newton algorithm for solving the corresponding total energy minimization problem subject to the orthogonality constraint. The Riemannian gradient and Riemannian Hessian of the total energy function are derived, and a matrix-form Riemannian Newton algorithm is presented. Under some mild conditions, we establish the global and quadratic convergence of the proposed method. Numerical results demonstrate our method is very efficient for large-scale problems. Our numerical experiments show that the linear equation (3.9) may be ill-conditioned when the parameter k is large. In this case, the prconditioned CG method with a good preconditioner may reduce much computing time and thus improve the efficiency [14, 23]. This needs further study. Another interesting topic is to apply our method to the total energy minimization in electronic structure [32, 34, 40, 41].

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, *Trust-region methods on Riemannian manifolds*, Found. Comput. Math., 7 (2007), pp. 303–330.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, 2008.

[3] P.-A. Absil and J. Malick, *Projection-like retractions on matrix manifolds*, SIAM J. Optim., 22 (2012), pp. 135–158.

[4] 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.

[5] T. A. Arias, M. C. Payne, and J. D. Joannopoulos, *Ab initio molecular dynamics: Analytically continued energy functionals and insights into iterative solutions*, Phys. Rev. Lett., 69 (1992), pp. 1077–1080.

[6] Z. J. Bai, S. Serra-Capizzano, and Z. Zhao, *Nonnegative inverse eigenvalue problems with partial eigendata*, Numer. Math., 120 (2012), pp. 387–431.

[7] P. Bendt and A. Zunger, *New approach for solving the density-functional self-consistent-field problem*, Phys. Rev. B, 26 (1982), pp. 3114–3137.

[8] D. Bernstein, *Matrix Mathematics – Theory, Facts, and Formulas*, 2nd edition, Princeton University Press, Princeton, 2009.

[9] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, 1999.

[10] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Revised 2nd edition, Academic Press, 2007.

[11] C. Le Bris, *Computational chemistry from the perspective of numerical analysis*, Acta Numer., 14 (2005), pp. 363–444.

[12] R. H. Bryd, M. E. Hbribar, and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, SIAM J. Optim., 9 (1999), pp. 877–900.

[13] E. Cancès and C. L. Bris, *On the convergence of SCF algorithms for the Hartree-Fock equations*, Math. Model. Numer. Anal., 34 (2000), pp. 749–774.

[14] R. Chan and X. Jin, *An Introduction to Iterative Toeplitz Solvers*, SIAM, Philadelpha, 2007.

[15] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Program., 75 (1996), pp. 407–439.

[16] 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.

[17] J. B. Francisco, J. M. Martínez, and L. Martínez, *Globally convergent trust-region methods for self-consistent field electronic structure calculations*, J. Chem. Phys., 121 (2004), pp. 10863–10878.

[18] J. B. Francisco, J. M. Martínez, and L. Martínez, *Density-based globally convergent trust-region methods for self-consistent field electronic structure calculations*, J. Math. Chem., 40 (2006), pp. 349–377.

[19] W. Gao, C. Yang, and J. Meza, *Solving a class of nonlinear eigenvalue problems by Newton's method*, Technical report, Lawrence Berkeley National Laboratory, Berkeley, CA, 2009.

[20] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th edition, Johns Hopkins University Press, Baltimore, 2013.

[21] M. W. Hirsch, *Differential Topology*, GTM 33, Springer-Verlag, New York, 1976.

[22] R. Horn and C. Johnson, *Topics on Matrix Analysis*, Cambridge University Press, Cambridge, 1994.

[23] X. Jin, *Preconditioning Techniques for Toeplitz Systems*, Higher Education Press, Beijing, 2010.

[24] G. P. Kerker, *Efficient iteration scheme for self-consistent pseudopotential calculations*, Phys. Rev. B, 23 (1981), pp. 3082–3084.

[25] G. Kresse and J. Furthmüller, *Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set*, Comput. Mater. Sci., 6 (1996), pp. 15–50.

[26] X. P. Li, R. W. Nunes, and D. Vanderbilt, *Density-matrix electronic-structure method with linear system-size scaling*, Phys. Rev. B, 47 (1993), pp. 10891–10894.

[27] X. Liu, X. Wang, Z. W. Wen, and Y. X. Yuan, *On the convergence of the self-consistent field iteration in Kohn-Sham density functional theory*, arXiv:1302.6022 [physics.comp-ph].

[28] J. H. Manton, *Optimization algorithms exploiting unitary constraints*, IEEE Trans. Signal Processing, 50 (2002), pp. 635–650.

[29] J. H. Manton, R. Mahony, and Y. Hua, *The geometry of weighted low rank approximations*, IEEE Trans. Signal Processing, 51 (2003), pp. 500–514.

[30] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge Univer-

sity Press, Cambridge, 2004.
[31] G. Meyer, *Geometric optimization algorithms for linear regression on fixed-rank matrices*, Ph.D. thesis, Department of electrical engineering and computer science, University of Liège, 2011, http://orbi.ulg.ac.be/handle/2268/97713.
[32] J. M. Millam and G. E. Scuseria, *Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations*, J. Chem. Phys., 106 (1997), pp. 5569–5577.
[33] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
[34] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients*, Rev. Mod. Phys., 64 (1992), pp. 1045–1097.
[35] B. G. Pfrommer, J. Demmel, and H. Simon, *Unconstrained energy functionals for electronic structure calculations*, J. Comput. Phys., 150 (1999), pp. 287–298.
[36] W. Ring and B. Wirth, *Optimization methods on Riemannian manifolds and their application to shape space*, SIAM J. Optim., 22 (2012), pp. 596–627.
[37] M. Shub, *Some remarks on dynamical systems and numerical analysis*, in: Dynamical Systems and Partial Differential Equations, Proc. VII ELAM (L. Lara-Carrero and J. Lewowicz, eds.), pp. 69–92, Equinoccio, U. Simón Bolívar, Caracas, 1986.
[38] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: An Introduction to Advanced Electronic Structure Theory*, Dover, New York, 1996.
[39] L. Thogersen, J. Olsen, D. Yeager, P. Jorgensen, P. Salek, and T. Helgaker, *The trust-region self-consistent field method: Towards a black-box optimization in Hartree-Fock and Kohn-Sham theories*, J. Chem. Phys., 121 (2004), pp. 16–27.
[40] J. VandeVondele and J. Hutter, *An efficient orbital transformation method for electronic structure calculations*, J. Chem. Phys., 118 (2003), pp. 4365–4369.
[41] T. Van Voorhis and M. Head-Gordon, *A geometric approach to direct minimization*, Mol. Phys., 100 (2002), pp. 1713–1721.
[42] Z. Wen and W. Yin, *A feasible method for optimization with orthogonality constraints*, Math. Program., 142 (2013), pp. 397–434.
[43] N. Yamashita and M. Fukushima, *Modified Newton methods for solving semismooth reformulations of monotone complementarity problems*, Math. Program., 76 (1997), pp. 469–491.
[44] C. Yang, J. C. Meza, and L. W. Wang, *A constrained optimization algorithm for total energy minimization in electronic structure calculations*, J. Comput. Phys., 217 (2006), pp. 709–721.
[45] C. Yang, J. C. Meza, and L. W. Wang, *A trust region direct constrained minimization algorithm for the Kohn-Sham equation*, SIAM J. Sci. Comput., 29 (2007), pp. 1854–1875.
[46] C. Yang, W. Gao, and J. C. Meza, *On the convergence of the self-consistent field iteration for a class of nonlinear eigenvalue problems*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1773–1788.