Riemannian Optimal Identification Method for Linear Systems With Symmetric Positive-Definite Matrix

Kazuhiro Sato ⬤, *Member, IEEE*, Hiroyuki Sato ⬤, *Member, IEEE*, and Tobias Damm ⬤

Abstract—**This article develops identification methods for linear continuous-time symmetric systems, such as electrical network systems, multiagent network systems, and temperature dynamics in buildings. To this end, we formulate three system identification problems for the corresponding discrete-time systems. The first is a least-squares problem in which we wish to minimize the sum of squared errors between the true and model outputs on the product manifold of the manifold of symmetric positive-definite matrices and two Euclidean spaces. In the second problem, to reduce the search dimensions, the product manifold is replaced with the quotient set under a specified group action by the orthogonal group. In the third problem, the manifold of symmetric positive-definite matrices in the first problem is replaced by the manifold of matrices with only positive diagonal elements. In particular, we examine the quotient geometry in the second problem. We propose Riemannian conjugate gradient methods for the three problems, and select initial points using a popular subspace method. The effectiveness of our proposed methods is demonstrated through numerical simulations and comparisons with the Gauss–Newton method, which is one of the most popular approach for solving least-squares problems.**

Index Terms—**Riemannian optimization, symmetry, system identification.**

I. INTRODUCTION

MANY important systems involved in electrical networks [1]–[3], multiagent networks [4], [5], and tempe-

K. Sato is with the Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: kazuhiro@mist.i.u-tokyo.ac.jp).

H. Sato is with the Department of Applied Mathematics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (e-mail: hsato@amp.i.kyoto-u.ac.jp).

T. Damm is with the Department of Mathematics, University of Kaiserslautern, Kaiserslautern 67663, Germany (e-mail: damm@mathematik.uni-kl.de).

rature dynamics in buildings [6], [7] can be modeled as

$$\begin{cases} \dot{\hat{x}}(t) = F\hat{x}(t) + G\hat{u}(t) \\ \hat{y}(t) = H\hat{x}(t) \end{cases} \tag{1}$$

where $\hat{x}(t) \in \mathbf{R}^n$, $\hat{u}(t) \in \mathbf{R}^m$, and $\hat{y}(t) \in \mathbf{R}^p$ are the state, input, and output of the system, respectively, and $F \in \mathrm{Sym}(n)$, $G \in \mathbf{R}^{n \times m}$, and $H \in \mathbf{R}^{p \times n}$ are constant matrices. Because the matrix F is symmetric, we call (1) a linear continuous-time symmetric system. In controlling a system described by (1), it is important to have an accurate identification of (F, G, H). Therefore, many identification techniques have been developed, such as prediction error methods [8]–[13] and subspace identification methods [14]–[20] for discrete-time systems, as well as for continuous-time systems [21]–[26].

However, it is difficult to identify a symmetric matrix F from the $K + 1$ input/output data $\{(u_0, y_0), (u_1, y_1), \ldots, (u_K, y_K)\}$ over the sampling interval h by using an indirect method that estimates a corresponding discrete-time system and, then, transforms to a continuous-time system. Here, $y_0, y_1, \ldots, y_K$ are noisy data observed from the true system, which is different from (1). This is because no system identification method has been derived for the corresponding discrete-time system

$$\begin{cases} \hat{x}_{k+1} = A\hat{x}_k + B\hat{u}_k \\ \hat{y}_k = C\hat{x}_k \end{cases} \tag{2}$$

where $\hat{x}_k := \hat{x}(kh)$, $\hat{u}_k := \hat{u}(kh)$, $\hat{y}_k := \hat{y}(kh)$, and

$$A := \exp(Fh) \in \mathrm{Sym}_+(n) \tag{3}$$

$$B := \left(\int_0^h \exp(Ft)dt \right) G \tag{4}$$

$$C := H. \tag{5}$$

That is, the existing methods in [8]–[20] for identifying the triplet (A, B, C) do not provide a symmetric positive-definite matrix A.

For this reason, we present novel prediction error methods for identifying

$$\Theta := (A, B, C) \in M := \mathrm{Sym}_+(n) \times \mathbf{R}^{n \times m} \times \mathbf{R}^{p \times n}$$

using the input/output data $\{(u_0, y_0), (u_1, y_1), \ldots, (u_K, y_K)\}$ under the assumption that the matrix A is stable. That is, we develop an identification method for the matrix A to be symmetric positive definite. If this is achieved

1) we can also obtain the matrices F, G, and H by

$$F = \log A / h \tag{6}$$

$$G = \left(\int_0^h \exp(Ft) dt \right)^{-1} B \tag{7}$$

$$H = C. \tag{8}$$

In particular, the matrix F is symmetric and is uniquely determined, because the map $\exp : \mathrm{Sym}(n) \to \mathrm{Sym}_+(n)$ is bijective [27].

2) we can directly identify discrete-time system (2) with a symmetric positive-definite matrix A. That is, our method works well even if the target system to be identified is not continuous-time system (1) but the discrete-time system.

To develop the prediction error methods, we formalize three different problems by introducing the Riemannian metric

$$\langle (\xi_1, \eta_1, \zeta_1), (\xi_2, \eta_2, \zeta_2) \rangle_\Theta$$
$$:= \mathrm{tr}(A^{-1} \xi_1 A^{-1} \xi_2) + \mathrm{tr}(\eta_1^\top \eta_2) + \mathrm{tr}(\zeta_1^\top \zeta_2) \tag{9}$$

for $(\xi_1, \eta_1, \zeta_1), (\xi_2, \eta_2, \zeta_2) \in T_\Theta M$, where the metric has also been used for a model reduction problem [28]. The first problem is the least-squares problem of minimizing the sum of squared errors on the Riemannian manifold M. In the second problem, to reduce the search dimension of the first problem, the manifold M is replaced by a quotient set. In the third problem, we replace the $\mathrm{Sym}_+(n)$ component of M with $\mathrm{Diag}_+(n)$.

The contributions of this article are summarized as follows.

1) In Section II-B, we show that the quotient set $N/O(n)$ in the second problem is indeed a manifold, where

$$N := M \cap S_{\mathrm{con}} \cap S_{\mathrm{ob}}. \tag{10}$$

Here, $S_{\mathrm{con}} := \{ (A, B, C) \in \mathbf{R}^{n \times n} \times \mathbf{R}^{n \times m} \times \mathbf{R}^{p \times n} \, | \, (A, B, C) \text{ is controllable} \}$ and $S_{\mathrm{ob}} := \{ (A, B, C) \in \mathbf{R}^{n \times n} \times \mathbf{R}^{n \times m} \times \mathbf{R}^{p \times n} \, | \, (A, B, C) \text{ is observable} \}$, where we say that (A, B, C) is controllable (resp. observable) if the corresponding discrete-time system described by (2) is controllable (resp. observable). Moreover, in Section II-B, we prove that Riemannian metric (9) on M induces a Riemannian metric into $N/O(n)$ by using a general theorem, as shown in Appendix C. That is, the quotient set $N/O(n)$ is shown to be a Riemannian manifold.

2) In Section IV, we propose Riemannian conjugate gradient (CG) methods for solving the aforementioned three problems. In developing the CG method for the first problem, we derive the Riemannian gradient of the objective function in terms of Riemannian metric (9), and use the concept of parallel transport. For the modified second problem on the quotient manifold $N/O(n)$, the parallel transport in the first problem is replaced by the projection onto the horizontal space that is a subspace of a tangent space of the manifold N, although the Riemannian gradient is the same. In Section III-C1, it is shown that the projection is obtained using the skew-symmetric solution to a linear matrix equation. In Appendix E, we prove that there exists a unique skew-symmetric solution to the equation under a mild assumption. Moreover, for the third problem, we derive another Riemannian gradient different from that in the first and second problems. Furthermore, in Section IV-D, we propose a technique for choosing initial points in the proposed algorithms for solving the three problems based on a subspace method such as numerical algorithms for subspace state space system identification

(N4SID) [17], multivariable output-error state space method (MOESP) [19], canonical variate analysis method (CVA) [15], orthogonal decomposition subspace method (ORT) [14], or nuclear norm subspace identification (N2SID) [20].

3) We demonstrate the effectiveness of our proposed methods for single-input single-output (SISO) and multi-input multi-output (MIMO) cases.

a) Our proposed methods for solving the aforementioned three problems can produce $A \in \mathrm{Sym}_+(n)$, unlike the Gauss–Newton (GN) method, which has been widely used for solving least-squares problems. In other words, we illustrate that the usual GN method as explained in Section V is not adequate for identifying system (1).

b) Our proposed methods significantly improve the results produced by a modified MOESP method in terms of various indices.

c) In MIMO cases, the rate of instability in the estimated matrix A_{est} produced by our method when solving the third problem is much higher than that for solving the first and second problems. In other words, the proposed methods for solving the first and second problems have a high degree of stability.

d) A hybrid approach combining the CG methods for solving the first and second problems may be more efficient than the individual CG methods.

The remainder of this article is organized as follows. In Section II, we formulate the aforementioned three problems mathematically. In particular, in Section II-B, we show that the quotient set $N/O(n)$ is a manifold. Moreover, we prove that Riemannian metric (9) on M induces a Riemannian metric on $N/O(n)$. In Section III, we discuss Riemannian geometries of our problems. In Section IV, we propose optimization algorithms for solving the three problems. In addition, we propose a technique for choosing an initial point in the algorithms. In Section V, we summarize the GN method. In Section VI, we demonstrate the effectiveness of our proposed methods. Finally, the conclusions of this article are presented in Section VII.

Notation: The sets of real and complex numbers are denoted by $\mathbf{R}$ and $\mathbf{C}$, respectively. The symbols $\mathrm{Sym}(n)$ and $\mathrm{Skew}(n)$ denote the vector spaces of symmetric matrices and skew-symmetric matrices in $\mathbf{R}^{n \times n}$, respectively. The symbol $\mathrm{Diag}(n)$ is the vector space of diagonal matrices in $\mathbf{R}^{n \times n}$. The manifold of symmetric positive definite matrices in $\mathrm{Sym}(n)$ is denoted by $\mathrm{Sym}_+(n)$. The manifold of matrices with positive diagonal elements in $\mathrm{Diag}(n)$ is denoted by $\mathrm{Diag}_+(n)$. The symbol $O(n)$ denotes the orthogonal group in $\mathbf{R}^{n \times n}$. The tangent space at p on a manifold $\mathcal{M}$ is denoted by $T_p \mathcal{M}$. The identity matrix of size n is denoted by I_n. Given vectors $v = (v_i), w = (w_i) \in \mathbf{R}^n$, (v, w) denotes the Euclidean inner product, i.e., $(v, w) = \sum_{i=1}^n v_i w_i$, and $\|v\|_2$ denotes the Euclidean norm, i.e., $\|v\|_2 := \sqrt{(v,v)} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$. Given a matrix $A \in \mathbf{R}^{n \times n}$, $\|A\|_F$ denotes the Frobenius norm, i.e., $\|A\|_F := \sqrt{\mathrm{tr}(A^\top A)}$, where the superscript $\top$ denotes the transpose and $\mathrm{tr}(A)$ denotes the trace of A, i.e., the sum of the diagonal elements of A. The symbol $\lambda(A)$ denotes the set of eigenvalues of A, and $\mathrm{sym}(A)$ and $\mathrm{sk}(A)$ denote the symmetric and skew-symmetric parts of A, respectively, i.e., $\mathrm{sym}(A) = \frac{A + A^\top}{2}$ and $\mathrm{sk}(A) = \frac{A - A^\top}{2}$. Given a smooth function f between finite dimensional manifolds $\mathcal{M}$ and $\mathcal{N}$, the differential of f at x is denoted by $\mathrm{D}f(x)$.

II. PROBLEM SETTINGS

This section presents the formulation of the three problems.

A. Problem 1

As described earlier, the aim of this article is to develop a novel prediction error method for identifying $\Theta \in M$ using the input/output data. To this end, we consider the following problem.

Problem 1: Suppose that the input/output data $\{(u_0, y_0), (u_1, y_1), \ldots, (u_K, y_K)\}$ and state dimension n are given. Then, find the minimizer of

$$\text{minimize} \quad f_1(\Theta) := \|e(\Theta)\|_2^2$$

$$\text{subject to} \quad \Theta \in M.$$

Here

$$e(\Theta) := \begin{pmatrix} y_1 - \hat{y}_1(\Theta) \\ y_2 - \hat{y}_2(\Theta) \\ \vdots \\ y_K - \hat{y}_K(\Theta) \end{pmatrix} \in \mathbf{R}^{pK} \qquad (11)$$

and $\hat{y}_k(\Theta)$ is $\hat{y}_k$ obtained by substituting the input data u_k into $\hat{u}_k$ of (2). The initial state $\hat{x}_0 \in \mathbf{R}^n$ in (2) is arbitrary. Note that $\hat{y}_k(\Theta)$ is different from the output data y_k, which is obtained by observing the output of the true system. That is, (2) is a mathematical model but is not the true system.

In this article, as mentioned in Section I, we endow M with Riemannian metric (9). Thus, Problem 1 is a Riemannian optimization problem.

B. Problem 2

It is possible to reduce the dimension of the problem of minimizing $\|e(\Theta)\|_2^2$ under the assumption that the initial state $\hat{x}_0$ is equal to zero. This is because Θ and

$$U \circ \Theta := (U^\top A U, U^\top B, C U)$$

realize input/output equivalent systems for any $U \in O(n)$, where $\circ$ denotes a group action of $O(n)$ on M. That is, they attain the same value of the prediction error, i.e., $\|e(\Theta)\|_2 = \|e(U \circ \Theta)\|_2$. Moreover, if $\Theta \in M$, we have $U \circ \Theta \in M$ for any $U \in O(n)$. This leads to the idea of equating Θ with $U \circ \Theta$ to reduce the dimension of the problem of minimizing $\|e(\Theta)\|_2^2$.

To this end, we endow M with an equivalence relation $\sim$, where $\Theta_1 \sim \Theta_2$ if and only if there exists some $U \in O(n)$ such that $\Theta_2 = U \circ \Theta_1$. Defining the equivalence class $[\Theta]$ by $[\Theta] := \{\Theta_1 \in M \mid \Theta_1 \sim \Theta\}$, we can equate Θ with any Θ_1 that is equivalent to Θ. Thus, instead of Problem 1, we can consider a minimization problem on the quotient set $M/O(n) := \{[\Theta] \mid \Theta \in M\}$.

However, it is an open problem whether the quotient set $M/O(n)$ is a manifold, although this set is a Hausdorff space from [29, Prop. 21.4]. In fact, although there are topological studies on control systems [30]–[34], there is no existing work on the quotient set $M/O(n)$. Thus, it is difficult to guarantee that $\pi_M^{-1}([\Theta])$ is a submanifold of M for all $\Theta \in M$, because we cannot use well-known general results such as [35, Prop. 3.4.4].

Here, the map $\pi_M : M \to M/O(n)$ denotes the canonical projection, i.e., $\pi_M(\Theta) = [\Theta]$ for any $\Theta \in M$. If $\pi_M^{-1}([\Theta])$ is not a manifold for some $\Theta \in M$, then $T_\Theta \pi_M^{-1}([\Theta])$ cannot be defined. That is, in this case, we cannot consider the vertical space in $T_\Theta M$. As a result, it may be impossible to define the horizontal space that is the orthogonal complement of the vertical space with respect to metric (9). This makes it difficult to develop an optimization method for solving the problem.

To resolve this issue, we consider the set N defined by (10) instead of M. The set N is an open submanifold of $\mathbf{R}^{n \times n} \times \mathbf{R}^{n \times m} \times \mathbf{R}^{p \times n}$, because, in addition to M, S_{con} and S_{ob} are open sets in $\mathbf{R}^{n \times n} \times \mathbf{R}^{n \times m} \times \mathbf{R}^{p \times n}$, as shown in [36, Prop. 3.3.12]. A group action of $O(n)$ on N, as in M, is given by

$$U \circ \Theta := (U^\top A U, U^\top B, C U) \qquad (12)$$

where $\Theta \in N$. Then, $U \circ \Theta \in N$ for any $U \in O(n)$. By introducing the equivalence class $[\Theta] := \{\Theta_1 \in N \mid \Theta_1 \sim \Theta\}$, we can define the quotient set $N/O(n) := \{[\Theta] \mid \Theta \in N\}$.

Unlike $M/O(n)$, we can guarantee that $N/O(n)$ is a manifold using the quotient manifold theorem [29], which is explained in Appendix B. To see this, we must confirm that action (12) is free and proper. Action (12) is proper because the Lie group $O(n)$ is compact (for a more detailed explanation, see [29, Cor. 21.6]). Thus, we show that action (12) is free. Suppose that the general linear group $GL(n)$ acts on N as

$$T \diamondsuit \Theta := (T^{-1} A T, T^{-1} B, C T), \quad T \in GL(n), \ \Theta \in N.$$

This action is free, as explained in [36, Rem. 6.5.10], i.e.,

$$\{T \in GL(n) \mid T \diamondsuit \Theta = \Theta\} = \{I_n\} \qquad (13)$$

for any $\Theta \in N$. Moreover, we have that

$$\{I_n\} \subset \{U \in O(n) \mid U \circ \Theta = \Theta\} \subset \{T \in GL(n) \mid T \diamondsuit \Theta = \Theta\} \qquad (14)$$

for any $\Theta \in N$. From (13) and (14), action (12) is free.

Thus, the following problem is an optimization problem on a manifold.

Problem 2: Suppose that the input/output data $\{(u_0, y_0), (u_1, y_1), \ldots, (u_K, y_K)\}$ and state dimension n are given. Then, find the minimizer of

$$\text{minimize} \quad f_2([\Theta]) := \|e(\Theta)\|_2^2$$

$$\text{subject to} \quad [\Theta] \in N/O(n).$$

That is, we also develop a prediction error method on the quotient manifold $N/O(n)$. Note that this development is different from that in [12], which considered a group action of the general linear group $GL(n)$ on a manifold instead of that of $O(n)$. It is not adequate to use the action in [12] for our problem, because this action does not, in general, preserve the symmetric positive-definiteness of the matrix A. For this reason, we consider the group action of $O(n)$ on the manifold N.

To introduce a Riemannian metric into $N/O(n)$, we define Riemannian metric (9) on N.

Because $N/O(n)$ is a quotient manifold, [35, Prop. 3.4.4] implies that $\pi^{-1}([\Theta])$ is an embedded submanifold of N for any $\Theta \in N$, where the map $\pi : N \to N/O(n)$ denotes the canonical projection, i.e., $\pi(\Theta) = [\Theta]$ for any $\Theta \in N$. Thus, we can define the vertical space $\mathcal{V}_\Theta := T_\Theta \pi^{-1}([\Theta])$ in $T_\Theta N$ for any $\Theta \in N$.
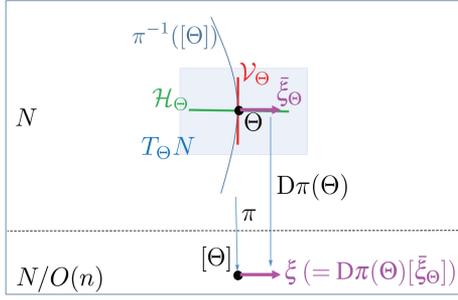
Fig. 1.   Conceptual diagram of vertical space $\mathcal{V}_\Theta$, horizontal space $\mathcal{H}_\Theta$, and horizontal lift $\bar{\xi}_\Theta$.

Moreover, from [29, Prop. 3.9], $T_\Theta M = T_\Theta N$ for any $\Theta \in N$, because N is an open set in M. Hence, we can consider $\mathcal{V}_\Theta$ in $T_\Theta M$ for any $\Theta \in N \subset M$. Additionally, the horizontal space $\mathcal{H}_\Theta$ can be defined as the orthogonal complement of the vertical space $\mathcal{V}_\Theta$ in $T_\Theta N$ with respect to metric (9). Furthermore, the horizontal lift $\bar{\xi}_\Theta \in \mathcal{H}_\Theta$ of $\xi \in T_{[\Theta]}(N/O(n))$ is defined as the unique element of the horizontal space $\mathcal{H}_\Theta$ satisfying $D\pi(\Theta)[\bar{\xi}_\Theta] = \xi$. Fig. 1 presents a diagram of these concepts.

In the following, we show that a Riemannian metric on $N/O(n)$ can be defined by

$$\langle \xi, \zeta \rangle_{[\Theta]} := \langle \bar{\xi}_\Theta, \bar{\zeta}_\Theta \rangle_\Theta \qquad (15)$$

where $\xi, \zeta \in T_{[\Theta]}(N/O(n))$, $\Theta \in \pi^{-1}([\Theta])$, and $\bar{\xi}_\Theta$ and $\bar{\zeta}_\Theta$ are the horizontal lifts of ξ and ζ at $\Theta \in N$, respectively. Note that $\langle \cdot, \cdot \rangle_\Theta$ of the right-hand side of (15) is Riemannian metric (9).

To this end, we must prove that

$$\langle \bar{\xi}_{\Theta_1}, \bar{\zeta}_{\Theta_1} \rangle_{\Theta_1} = \langle \bar{\xi}_{\Theta_2}, \bar{\zeta}_{\Theta_2} \rangle_{\Theta_2} \qquad (16)$$

for any $\Theta_1, \Theta_2 \in \pi^{-1}([\Theta])$. To prove this, we first note that (9) yields

$$\langle D\phi_U(\Theta)[\xi_1], D\phi_U(\Theta)[\xi_2] \rangle_{\phi_U(\Theta)} = \langle \xi_1, \xi_2 \rangle_\Theta \qquad (17)$$

for any $\xi_1, \xi_2 \in T_\Theta N$ and any $U \in O(n)$, where $\phi_U(\Theta) := U \circ \Theta$. That is, $D\phi_U(\Theta)$ is an isometry in terms of Riemannian metric (9). Equation (17) implies the following theorem.

Theorem 1: For any $U \in O(n)$

$$\bar{\xi}_{\phi_U(\Theta)} = D\phi_U(\Theta)[\bar{\xi}_\Theta]. \qquad (18)$$

We provide the proof of Theorem 1 in Appendix C.

Using Theorem 1 and (17), we can prove (16) as follows: For any $\Theta_1, \Theta_2 \in \pi^{-1}([\Theta])$, there exists some $U \in O(n)$ such that $\Theta_2 = \phi_U(\Theta_1)$. Thus

$$\langle \bar{\xi}_{\Theta_2}, \bar{\zeta}_{\Theta_2} \rangle_{\Theta_2} = \langle \bar{\xi}_{\phi_U(\Theta_1)}, \bar{\zeta}_{\phi_U(\Theta_1)} \rangle_{\phi_U(\Theta_1)}$$
$$= \langle D\phi_U(\Theta_1)[\bar{\xi}_{\Theta_1}], D\phi_U(\Theta_1)[\bar{\zeta}_{\Theta_1}] \rangle_{\phi_U(\Theta_1)}$$
$$= \langle \bar{\xi}_{\Theta_1}, \bar{\zeta}_{\Theta_1} \rangle_{\Theta_1}$$

where the second and third equalities follow from (18) and (17), respectively. Note that (16) is based on isometric condition (17) in terms of Riemannian metric (9). In other words, if (17) is not satisfied, we cannot conclude that (16) holds.

Based on the above discussion, $N/O(n)$ endowed with (15) is a Riemannian quotient manifold of N, and the natural projection $\pi : N \to N/O(n)$ becomes a Riemannian submersion. That is, the projection π is a smooth submersion, and for any $\Theta \in N$, the differential $D\pi_\Theta$ is an isometry between the horizontal space $\mathcal{H}_\Theta$ and $T_{\pi(\Theta)}(N/O(n))$. Moreover, (15) is a unique

Riemannian metric such that $\pi : N \to N/O(n)$ is a Riemannian submersion. This is because, as shown in previously, $O(n)$ is a Lie group of isometries of the manifold N endowed with (9) that acts smoothly, freely, and properly on N (for a more general description, see [37, Prop. 2.28]). This means that if we introduce Riemannian metric (9) into N, the geometry of $N/O(n)$ is uniquely determined.

To summarize, Problem 2 is a Riemannian optimization problem, and most of the geometry of $N/O(n)$ can be studied by lifting from $N/O(n)$ to N.

C. Problem 3

Moreover, we can consider a simpler problem than Problems 1 and 2. This is because, for any $\Theta \in M$, there is a unique $\tilde{U} \in O(n)$ such that

$$\tilde{U} \circ \Theta = (\Lambda, \tilde{U}^\top B, C\tilde{U})$$

where $\Lambda \in \text{Diag}_+(n)$. That is, the above Θ and $\tilde{U} \circ \Theta$ realize input/output equivalent systems, i.e., $\|e(\Theta)\|_2 = \|e(\tilde{U} \circ \Theta)\|_2$, under the assumption that the initial state $\hat{x}_0$ is equal to zero. The simpler problem is formulated as follows.

Problem 3: Suppose that the input/output data $\{(u_0, y_0), (u_1, y_1), \dots, (u_K, y_K)\}$ and state dimension n are given. Then, find the minimizer of

$$\text{minimize} \quad f_3(\Theta) := \|e(\Theta)\|_2^2$$

$$\text{subject to} \quad \Theta \in \tilde{M}.$$

Here, $\tilde{M} := \text{Diag}_+(n) \times \mathbf{R}^{n \times m} \times \mathbf{R}^{p \times n}$. However, we demonstrate in Section VI that, if the output data $y_0, y_1, \dots, y_K$ are noisy, the results provided by our algorithm for solving Problem 3 are more noise-sensitive than those produced by our algorithms for solving Problems 1 and 2.

Similar to Riemannian metric (9) on M, we define the Riemannian metric on $\tilde{M}$ as

$$\langle (\xi_1, \eta_1, \zeta_1), (\xi_2, \eta_2, \zeta_2) \rangle_\Theta$$
$$:= \text{tr}(A^{-1} \xi_1 A^{-1} \xi_2) + \text{tr}(\eta_1^\top \eta_2) + \text{tr}(\zeta_1^\top \zeta_2)$$
$$= \text{tr}((A^{-1})^2 \xi_1 \xi_2) + \text{tr}(\eta_1^\top \eta_2) + \text{tr}(\zeta_1^\top \zeta_2) \qquad (19)$$

for $(\xi_1, \eta_1, \zeta_1), (\xi_2, \eta_2, \zeta_2) \in T_\Theta \tilde{M}$. Here, the second equality follows from the fact that A^{-1}, ξ_1, and ξ_2 are diagonal matrices. Thus, Problem 3 is a Riemannian optimization problem.

D. Another Riemannian Metric on M, $N/O(n)$, and $\tilde{M}$

Instead of Riemannian metric (9), we can introduce the Riemannian metric

$$\langle (\xi_1, \eta_1, \zeta_1), (\xi_2, \eta_2, \zeta_2) \rangle_\Theta$$
$$:= \text{tr}(\xi_1 \xi_2) + \text{tr}(\eta_1^\top \eta_2) + \text{tr}(\zeta_1^\top \zeta_2) \qquad (20)$$

into the manifolds M and $\tilde{M}$. Moreover, even if we define Riemannian metric (20) into N, a Riemannian metric on $N/O(n)$ can be defined by (15). This is because (16) holds under metric (20). That is, in addition to Riemannian metric (9), Riemannian metric (20) also implies that the natural projection $\pi : N \to N/O(n)$ is a Riemannian submersion. However, this
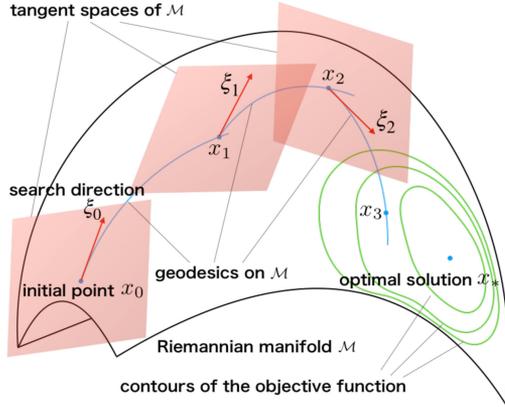
Fig. 2. Optimization process on a Riemannian manifold $\mathcal{M}$.

simple metric is not adequate for solving Problems 1, 2, and 3, as explained in Section IV-A.

Remark 1: In this article, we assume that the state dimension n is given. In practice, the dimension n must be determined before solving Problems 1, 2, and 3. For example, we can determine n by using Akaike's information criterion [38] or calculating the singular value decomposition of a matrix related to the input and output matrices [39].

Remark 2: As mentioned in Section I, we can identify (F, G, H) in (1) using (6), (7), and (8) after the identification of (A, B, C) in (2). In addition to Problems 1, 2, and 3, we consider the following problem:

$$\text{minimize} \quad \|\exp(Fh) - A\|_F^2$$

$$\text{subject to} \quad F \in \text{Sym}(n).$$

One may think that, by solving the above problem, $F \in \text{Sym}(n)$ for (1) can be obtained even if $A \notin \text{Sym}_+(n)$. However, this is not true. For example, if $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, then there is no solution F to the above problem. In fact, the infimum of the objective function is 0, whereas this value cannot be obtained with any $F \in \text{Sym}(n)$.

Remark 3: Note that system (1) does not correspond to a symmetric continuous-time system discussed in [3] and [40]. Here, system (1) is said to be symmetric in the sense of the definition in [3] and [40] if there exists some $T \in GL(n) \cap \text{Sym}(n)$ such that $F^\top T = TF$ and $H^\top = TG$.

III. GEOMETRIES OF PROBLEMS 1, 2, AND 3

A. Riemannian Optimization

In preparation for subsequent sections, we introduce the concepts of the exponential mapping and the Riemannian gradient for Riemannian optimization [35], [41], [42], and provide a brief description of an optimization algorithm. In this section, we consider a general Riemannian optimization problem of minimizing an objective function f defined on a Riemannian manifold $\mathcal{M}$. That is, $\mathcal{M}$ is equipped with a Riemannian metric $\langle \cdot, \cdot \rangle$ that endows the tangent space $T_x\mathcal{M}$ at each point $x \in \mathcal{M}$ with an inner product.

Fig. 2 illustrates an optimization process on $\mathcal{M}$. As shown in this figure, the next point is determined by using geodesics and search direction vectors. The following explains the details.

1) Exponential Mapping: For the purpose of optimization on a Riemannian manifold $\mathcal{M}$, the update formula $x + \xi$ does not make sense for $x \in \mathcal{M}$ and $\xi \in T_x\mathcal{M}$. This is in contrast to the case of optimization on a Euclidean space $\mathcal{E}$. That is, on $\mathcal{E}$, we can compute a point $x_+ \in \mathcal{E}$ from the current point $x \in \mathcal{E}$ and search direction $d \in \mathcal{E}$ as $x_+ = x + d$. Thus, we seek the next point x_+ on a curve called a geodesic on $\mathcal{M}$ emanating from x in the direction of ξ. For any $x, y \in \mathcal{M}$, on a geodesic between two points x and y that are sufficiently close, the path along the geodesic is the shortest among all curves connecting x and y. It is known that, for any $\xi \in T_x\mathcal{M}$, there exists an interval $I \subset \mathbf{R}$ around 0 and a unique geodesic $\Gamma_{(x,\xi)} : I \to \mathcal{M}$ such that $\Gamma_{(x,\xi)}(0) = x$ and $\dot{\Gamma}_{(x,\xi)}(0) = \xi$. The exponential mapping Exp at $x \in \mathcal{M}$ can be defined through the geodesic as

$$\text{Exp}_x(\xi) := \Gamma_{(x,\xi)}(1)$$

because the geodesic $\Gamma_{(x,\xi)}$ has the homogeneity property $\Gamma_{(x,a\xi)}(t) = \Gamma_{(x,\xi)}(at)$ for any $a \in \mathbf{R}$ satisfying $at \in I$.

2) Riemannian Gradient: In addition to the exponential mapping, we need a Riemannian gradient to solve our problems. The Riemannian gradient $\text{grad } f(x)$ of f at $x \in \mathcal{M}$ is defined as a tangent vector at x that satisfies

$$\text{D}f(x)[\xi] = \langle \text{grad } f(x), \xi \rangle_x$$

for any $\xi \in T_x\mathcal{M}$, where $\text{D}f(x)[\xi]$ is defined as

$$\text{D}f(x)[\xi] := \xi f.$$

Note that a tangent vector can be identified with a derivative.

3) Algorithm: The update formula of a gradient algorithm for minimizing the objective function f on $\mathcal{M}$ is given by

$$x_{k+1} = \text{Exp}_{x_k}(\xi_k)$$

with an initial point $x_0 \in \mathcal{M}$, where $\xi_k \in T_{x_k}\mathcal{M}$ is a search direction defined by using the Riemannian gradients.

B. Geometry of Problem 1

In first-order optimization algorithms such as the steepest descent (SD) and CG methods on the manifold M equipped with Riemannian metric (9), we need the Riemannian gradient of the objective function f_1.

Let $\bar{f}_1$ denote the extension of the objective function f_1 to the ambient Euclidean space $\mathbf{R}^{n \times n} \times \mathbf{R}^{n \times m} \times \mathbf{R}^{p \times n}$. Then, the directional derivative of $\bar{f}_1$ at $\Theta \in M$ along $\xi = (\xi_A, \xi_B, \xi_C) \in T_\Theta M$ is given by

$$\text{D}\bar{f}_1(\Theta)[\xi] = 2(\text{D}e(\Theta)[\xi], e(\theta)) \tag{21}$$

where

$$\text{D}e(\Theta)[\xi] = \begin{pmatrix} -\text{D}\hat{y}_1(\Theta)[\xi] \\ -\text{D}\hat{y}_2(\Theta)[\xi] \\ \vdots \\ -\text{D}\hat{y}_K(\Theta)[\xi] \end{pmatrix}. \tag{22}$$

Equation (2) implies that

$$\text{D}\hat{y}_k(\Theta)[\xi] = C \sum_{i=0}^{k-1} A^{k-i-1}(\xi_A\hat{x}_i + \xi_B u_i) + \xi_C\hat{x}_k. \tag{23}$$

It follows from (21), (22), and (23) that

$$\mathrm{D}\bar{f}_1(\Theta)[\xi] = \mathrm{tr}(\xi_A \operatorname{sym}(G_A)) + \mathrm{tr}(\xi_B^\top G_B) + \mathrm{tr}(\xi_C^\top G_C) \tag{24}$$

where

$$G_A := -2\sum_{k=1}^{K}\sum_{i=0}^{k-1} A^{k-i-1}C^\top(y_k - \hat{y}_k(\Theta))\hat{x}_i^\top \tag{25}$$

$$G_B := -2\sum_{k=1}^{K}\sum_{i=0}^{k-1} A^{k-i-1}C^\top(y_k - \hat{y}_k(\Theta))u_i^\top \tag{26}$$

$$G_C := -2\sum_{k=1}^{K}(y_k - \hat{y}_k(\Theta))\hat{x}_k^\top. \tag{27}$$

Here, we used the property $\xi_A = \xi_A^\top$. We can observe that the Euclidean gradient of $\bar{f}_1$ at Θ is given by

$$\nabla \bar{f}_1(\Theta) = (G_A, G_B, G_C). \tag{28}$$

In [12], we can find a similar derivation for a more complicated system. Because we introduced Riemannian metric (9), the Euclidean gradient in (28) yields the Riemannian gradient

$$\operatorname{grad} f_1(\Theta) = (A \operatorname{sym}(G_A)A, G_B, G_C). \tag{29}$$

For a detailed explanation, see Appendix A.

C. Geometry of Problem 2

Because the natural projection $\pi : N \to N/O(n)$ is a Riemannian submersion, most of the geometry of $N/O(n)$ can be studied by lifting from $N/O(n)$ to N, as described in Section II-B.

1) Orthogonal Projection Onto the Horizontal Space $\mathcal{H}_\Theta$: In Section IV-B, we need the concept of vector transport (which is a generalized concept of parallel transport [35]) on the manifold $N/O(n)$ equipped with Riemannian metric (15) to develop a Riemannian CG method. Here, note that we have introduced Riemannian metric (9) into N. To this end, for any $\Theta \in N$, we use the orthogonal projection P_Θ onto the horizontal space $\mathcal{H}_\Theta$.

To derive P_Θ, we need to explicitly describe the vertical space $\mathcal{V}_\Theta$ and the horizontal space $\mathcal{H}_\Theta$. First, we specify $\mathcal{V}_\Theta$. Consider any curve $\Theta(t)$ on $\pi^{-1}([\Theta]) \subset N$ with $\Theta(0) = \Theta$ that is expressed as

$$\Theta(t) = (U^\top(t)AU(t), U^\top(t)B, CU(t))$$

where $U(t)$ denotes a curve on $O(n)$ with $U(0) = I_n$. Differentiating both sides with respect to t, we obtain

$$\dot{\Theta}(0) = (\dot{U}^\top(0)A + A\dot{U}(0), \dot{U}^\top(0)B, C\dot{U}(0))$$

where $\dot{U}(0) \in T_{I_n}O(n) \cong \operatorname{Skew}(n)$. Thus, we have that

$$\mathcal{V}_\Theta = \{(-U'A + AU', -U'B, CU') \,|\, U' \in \operatorname{Skew}(n)\}.$$

Next, we characterize the horizontal space $\mathcal{H}_\Theta$. Let $(A,' B,' C') \in \mathcal{H}_\Theta$, i.e.,

$$\langle(-U'A + AU', -U'B, CU'), (A', B', C')\rangle_\Theta = 0 \tag{30}$$

for all $U' \in \operatorname{Skew}(n)$. This means that

$$\mathrm{tr}(U'(2A'A^{-1} + BB'^\top + C^\top C')) = 0.$$

Because $U' \in \operatorname{Skew}(n)$ is arbitrary, we conclude that $2A'A^{-1} + BB'^\top + C^\top C' \in \operatorname{Sym}(n)$. That is, $\operatorname{sk}(2A'A^{-1} + BB'^\top + C^\top C') = 0$. Thus

$$\mathcal{H}_\Theta \subset \{(A', B', C') \,|\, \operatorname{sk}(2A'A^{-1} + BB'^\top + C^\top C') = 0\}.$$

Conversely, if $(A', B', C') \in \{(A', B', C') \,|\, \operatorname{sk}(2A'A^{-1} + BB'^\top + C^\top C') = 0\}$, we have that $(A', B', C') \in \mathcal{H}_\Theta$, because (30) holds. Hence, we obtain

$$\mathcal{H}_\Theta = \{(A', B', C') \,|\, \operatorname{sk}(2A'A^{-1} + BB'^\top + C^\top C') = 0\}. \tag{31}$$

We are in a position to describe the orthogonal projection P_Θ onto the horizontal space $\mathcal{H}_\Theta$.

Theorem 2: The orthogonal projection P_Θ onto $\mathcal{H}_\Theta$ is given by

$$P_\Theta(\eta) = \eta + (XA - AX, XB, -CX) \tag{32}$$

where $\eta = (a, b, c) \in T_\Theta N$, and X is the skew-symmetric solution to the linear matrix equation

$$\mathcal{L}_1(X) + 2\mathcal{L}_0(X) + \beta = 0 \tag{33}$$

where the linear matrix mappings $\mathcal{L}_0, \mathcal{L}_1 : \mathbf{R}^{n\times n} \to \mathbf{R}^{n\times n}$ are defined by $\mathcal{L}_0(X) := AXA^{-1} + A^{-1}XA - 2X$, $\mathcal{L}_1(X) := (BB^\top + C^\top C)X + X(BB^\top + C^\top C)$, and $\beta := 2\operatorname{sk}(2A^{-1}a + bB^\top + c^\top C)$.

We provide the proof in Appendix D.

We can guarantee that there exists a unique solution $X \in \operatorname{Skew}(n)$ to (33) under the assumption

$$\dim(\operatorname{Ker}(\lambda I_n - A) \cap \operatorname{Ker} B^\top \cap \operatorname{Ker} C) \le 1 \text{ for any } \lambda \in \mathbf{R}. \tag{34}$$

Assumption (34) holds if matrix A has only simple eigenvalues, because then $\dim(\operatorname{Ker}(\lambda I_n - A)) \le 1$ for all $\lambda \in \mathbf{R}$. Furthermore, if (A, C) is observable, i.e.,

$$\operatorname{rank}\begin{pmatrix}\lambda I_n - A \\ C\end{pmatrix} = n \Leftrightarrow \operatorname{Ker}(\lambda I_n - A) \cap \operatorname{Ker} C = \{0\}$$

for all $\lambda \in \mathbf{C}$, then (34) holds. Analogously, the controllability of (A, B), i.e.,

$$\operatorname{rank}\begin{pmatrix}\lambda I_n - A & B\end{pmatrix} = \operatorname{rank}\begin{pmatrix}\lambda I_n - A \\ B^\top\end{pmatrix} = n$$

also implies (34).

Theorem 3: Assume that (34) holds, and let $\mathcal{L} := \mathcal{L}_1 + 2\mathcal{L}_0$. Then, $\operatorname{Ker}\mathcal{L} = \operatorname{Ker}\mathcal{L}_1 \cap \operatorname{Ker}\mathcal{L}_0 \subset \operatorname{Ker}\mathcal{L}_0 \subset \operatorname{Sym}(n)$. In particular, $\mathcal{L} : \operatorname{Skew}(n) \to \operatorname{Skew}(n)$ is an automorphism. That is, for any $Y \in \operatorname{Skew}(n)$, there exists a unique $X \in \operatorname{Skew}(n)$ with $\mathcal{L}(X) = Y$.

The proof is given in Appendix E.

2) Riemannian Gradient: In numerical computations, we can use the horizontal lift $\overline{\operatorname{grad} f_2}_\Theta$ as the Riemannian gradient at $[\Theta] \in N/O(n)$. The horizontal lift belongs to the horizontal space $\mathcal{H}_\Theta$, and we have that

$$\overline{\operatorname{grad} f_2}_\Theta = \operatorname{grad} f_1(\Theta) \tag{35}$$

as shown in [35, Sec. 3.6.2]. Thus, as the Riemannian gradient at $[\Theta] \in N/O(n)$, we can use $\operatorname{grad} f_1(\Theta)$, i.e., (29).

D. Geometry of Problem 3

We have introduced Riemannian metric (19) into the manifold $\tilde{M}$. Let $\bar{f}_3$ denote the extension of the objective function f_3 to the ambient Euclidean space $\mathbf{R}^{n \times n} \times \mathbf{R}^{n \times m} \times \mathbf{R}^{p \times n}$. Then, the directional derivative of $\bar{f}_3$ at $\Theta \in \tilde{M}$ along $\xi = (\xi_A, \xi_B, \xi_C) \in T_\Theta \tilde{M}$ is given by

$$
\begin{aligned}
\mathrm{D}\bar{f}_3(\Theta)[\xi] &= \mathrm{tr}(\xi_A^\top G_A) + \mathrm{tr}(\xi_B^\top G_B) + \mathrm{tr}(\xi_C^\top G_C) \\
&= \mathrm{tr}(\xi_A \,\mathrm{diag}(G_A)) + \mathrm{tr}(\xi_B^\top G_B) + \mathrm{tr}(\xi_C^\top G_C)
\end{aligned}
\tag{36}
$$

where G_A, G_B, and G_C are defined by (25), (26), and (27), respectively. Here, we used the property that $\xi_A \in T_A \mathrm{Diag}_+(n) \cong \mathrm{Diag}(n)$. Moreover, it follows from (19) and $\mathrm{D}f_3(\Theta)[\xi] = \langle \mathrm{grad}\, f_3(\Theta), \xi \rangle_\Theta$ that

$$
\begin{aligned}
\mathrm{D}f_3(\Theta)[\xi] &= \mathrm{tr}((A^{-1})^2 (\mathrm{grad}\, f_3(\Theta))_A \xi_A) \\
&\quad + \mathrm{tr}(\xi_B^\top (\mathrm{grad}\, f_3(\Theta))_B) + \mathrm{tr}(\xi_C^\top (\mathrm{grad}\, f_3(\Theta))_C).
\end{aligned}
\tag{37}
$$

Because $\mathrm{D}f_3(\Theta)[\xi] = \mathrm{D}\bar{f}_3(\Theta)[\xi]$, (36) and (37) yield

$$
\mathrm{grad}\, f_3(\Theta) = (A^2 \mathrm{diag}(G_A), G_B, G_C).
$$

IV. OPTIMIZATION ALGORITHMS FOR SOLVING PROBLEMS 1, 2, AND 3

This section describes optimization algorithms for solving Problems 1, 2, and 3, and introduces a technique for choosing initial points in the algorithms.

A. Optimization Algorithm for Solving Problem 1

Algorithm 1 describes a Riemannian CG method for solving Problem 1. Because the Riemannian metric on the manifold M is defined by (9), the exponential map Exp on M is given by

$$
\begin{aligned}
&\mathrm{Exp}_\Theta(A', B', C') \\
&= (A^{1/2} \exp(A^{-1/2} A' A^{-1/2}) A^{1/2}, B + B', C + C') \\
&= (A \exp(A^{-1} A'), B + B', C + C')
\end{aligned}
\tag{38}
$$

and the parallel transport $\mathcal{P}$ is given by

$$
\begin{aligned}
&\mathcal{P}_{\Theta_1, \Theta_2}(A', B', C') \\
&= ((A_2 A_1^{-1})^{1/2} A' ((A_2 A_1^{-1})^{1/2})^\top, B', C')
\end{aligned}
\tag{39}
$$

where $\Theta_i = (A_i, B_i, C_i) \in M$ $(i = 1, 2)$, as shown in [43]. We choose t_k in step 4 as the Armijo step size [35]. The parameter β_{k+1} in step 5 is called the Dai–Yuan type parameter [44].

Note the if we introduce Riemannian metric (20) instead of (9), exponential mapping (38) is replaced with

$$
\mathrm{Exp}_\Theta(A', B', C') = (A + A', B + B', C + C').
$$

Thus, $\mathrm{Exp}_\Theta(A', B', C') \notin M$ for some $(A', B', C') \in T_\Theta M$ because $A + A'$ is not always positive-definite. As a result, we have to carefully choose $(A', B', C') \in T_\Theta M$ unlike for Riemannian metric (9).

The computational complexity of calculating the gradient $\mathrm{grad}\, f_1(\Theta)$ is higher than that of the other steps in Algorithm 1. To estimate the complexity, we examine the complexities of

G_A, G_B, and G_C. To this end, we note that G_A in (25) can be rewritten as

$$
G_A = -2 \sum_{i=0}^{K-1} \sum_{k=i+1}^{K} A^{k-i-1} C^\top (y_k - \hat{y}_k(\Theta)) \hat{x}_i^\top.
$$

Thus, we can recursively calculate G_A as

$$
G_A(i+1) = G_A(i) - 2\gamma(i) \hat{x}_{K-(i+1)}^\top
\tag{40}
$$

where

$$
\begin{aligned}
G_A(0) &= 0 \\
\gamma(i) &= C^\top (y_{K-i} - \hat{y}_{K-i}(\Theta)) + A\gamma(i-1) \\
\gamma(0) &= C^\top (y_K - \hat{y}_K(\Theta)).
\end{aligned}
$$

In fact, $G_A(K) = G_A$. If $p < n$, i.e., the number of outputs is less than that of states, the computational complexity of $\gamma(i) \hat{x}_{K-(i+1)}^\top$ for each $i \in \{0, 1, \ldots, K-1\}$ in (40) is $\mathcal{O}(n^2)$, because that of $\gamma(i)$ for each $i \in \{0, 1, \ldots, K-1\}$ is $\mathcal{O}(n^2)$. Thus, the complexity of G_A is $\mathcal{O}(Kn^2)$. Similarly, if $m < n$ and $p < n$, then the complexity of G_B is $\mathcal{O}(Kn^2)$. Moreover, if $p < n$, (27) implies that the complexity of G_C is $\mathcal{O}(Kn^2)$. Hence, if $p, m < n < K$, (29) implies that the complexity of $\mathrm{grad}\, f_1(\Theta)$ is $\mathcal{O}(Kn^2)$.

B. Optimization Algorithm for Solving Problem 2

In numerically solving Problem 2, we regard the manifold M as N, because [36, Prop. 3.3.12] implies that the manifold N is a dense set in the manifold M. The proposed CG-based method for solving Problem 2 is obtained by replacing the parallel transport $\mathcal{P}_{\Theta_k, \Theta_{k+1}}$ in Algorithm 1 with the orthogonal projection $P_{\Theta_{k+1}}$ given by (32) onto the horizontal space $\mathcal{H}_{\Theta_{k+1}}$. Note that the orthogonal projection $P_{\Theta_{k+1}}$ defines a vector transport on the quotient manifold $N/O(n)$ [35].

C. Optimization Algorithm for Solving Problem 3

The Riemannian CG method for solving Problem 3 is the same as Algorithm 1, except for the following:
1) replace M with $\tilde{M}$;
2) replace $\mathrm{grad}\, f_1(\Theta_k)$ with $\mathrm{grad}\, f_3(\Theta_k)$.

However, the computational complexity is lower than in the case of Problem 1. This is because the matrices A_k ($k = 0, 1, \ldots$) in Algorithm 1 are diagonal when solving Problem 3, unlike for Problem 1.

D. Initial Points in Algorithm 1

To select an initial point Θ_0 in Algorithm 1 for solving Problems 1 and 2, we propose Algorithm 2 in which rand denotes a single uniformly distributed random number in the interval $(0, 1)$. In step 1, we obtain a triplet (A, B, C) using an existing subspace method such as N4SID [17], MOESP [19], CVA [15], ORT [14], and N2SID [20]. However, at this stage, A is not necessarily contained in $\mathrm{Sym}(n)$ or in $\mathrm{Sym}_+(n)$. Thus, in step 2, we replace A with the symmetric part of A. That is, at this stage, $A \in \mathrm{Sym}(n)$, but $A \notin \mathrm{Sym}_+(n)$ in general. In fact, if there is $i \in \{1, 2, \ldots, n\}$ such that $\lambda_i \leq 0$ in step 3, $A \notin \mathrm{Sym}_+(n)$. In steps 4, 5, 6, 7, and 8, any negative eigenvalue of A is replaced with a random value in $(0, 0.01)$. That is, we consider negative eigenvalues of A to be perturbed small positive

---

Algorithm 1: Conjugate Gradient Method for Solving Problem 1.

1: Set input/output data $\{(u_0, y_0), (u_1, y_1), \ldots, (u_K, y_K)\}$, the state dimension n, and an initial point $\Theta_0 := (A_0, B_0, C_0) \in M$.

2: Set $\eta_0 = -\mathrm{grad}\, f_1(\Theta_0)$ using (29).

3: **for** $k = 0, 1, 2, \ldots$ **do**

4:     Compute a step size $t_k > 0$, and set

$$\Theta_{k+1} = \mathrm{Exp}_{\Theta_k}(t_k \eta_k). \tag{41}$$

5:     Set

$$\beta_{k+1} = \frac{\|g_{k+1}\|_{k+1}^2}{\langle g_{k+1}, \mathcal{P}_{\Theta_k, \Theta_{k+1}}(\eta_k)\rangle_{k+1} - \langle g_k, \eta_k\rangle_k},$$

    where $g_k := \mathrm{grad}\, f_1(\Theta_k)$, and $\|\cdot\|_k$ and $\langle\cdot,\cdot\rangle_k$ denote the norm and the inner product in the tangent space $T_{\Theta_k}M$, respectively.

6:     Set

$$\eta_{k+1} = -g_{k+1} + \beta_{k+1}\mathcal{P}_{\Theta_k, \Theta_{k+1}}(\eta_k). \tag{42}$$

7: **end for**

---

eigenvalues. Thus, steps 9 and 10 produce $A \in \mathrm{Sym}_+(n)$ and $\Theta_0 \in M$, respectively.

For Problem 3, we replace step 9 in Algorithm 2 with

$$A \leftarrow \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),\ B \leftarrow V^\top B,\ C \leftarrow CV$$

where $V := (v_1\ \ v_2\ \ \cdots\ \ v_n)$. Then, Θ_0 in step 10 is contained in $\tilde{M}$.

Remark 4: A Riemannian SD method for solving Problems 1 and 2 can be derived by replacing steps 5 and 6 in Algorithm 1 with $\eta_{k+1} = -\mathrm{grad}\, f_1(\Theta_{k+1})$. That is, in contrast to the case of the CG methods, the SD method for Problem 2 is the same as that for Problem 1, because (35) holds. However, the SD method is not more efficient than CG methods [45]. We demonstrate this fact in Section VI-A.

V. GN METHOD FOR SOLVING THE PROPOSED PROBLEMS

The GN method has been widely used for solving least-squares problems. Before comparing our proposed methods with the GN method, this section summarizes the GN method.

In the GN method [10], [13], we often use the vector parameter

$$\theta := \begin{pmatrix} \mathrm{vec}_{\mathrm{sym}}(A) \\ \mathrm{vec}(B) \\ \mathrm{vec}(C) \end{pmatrix} \in \mathbf{R}^{n_\theta} \tag{43}$$

with $n_\theta := n(\frac{n+1}{2} + m + p)$, where vec denotes the usual vec-operator, i.e., $\mathrm{vec}(A) \in \mathbf{R}^{n^2}$ is obtained by stacking the columns of $A \in \mathbf{R}^{n \times n}$, and for a symmetric matrix $A \in \mathrm{Sym}(n)$, $\mathrm{vec}_{\mathrm{sym}}(A)$ denotes the $\frac{1}{2}n(n+1)$-vector that is obtained from $\mathrm{vec}(A)$ by eliminating the redundant elements. For example, if $A = (a_{ij}) \in \mathrm{Sym}(3)$

$\mathrm{vec}(A)$

$$= \begin{pmatrix} a_{11} & a_{21} & a_{31} & a_{12} & a_{22} & a_{32} & a_{13} & a_{23} & a_{33} \end{pmatrix}^\top$$

Algorithm 2: Constructing an Initial Point $\Theta_0 \in M$.

1: Set (A, B, C) using a subspace method.

2: $A \leftarrow \mathrm{sym}(A)$.

3: Let $A = \sum_{i=1}^n \lambda_i v_i v_i^\top$ be the eigenvalue decomposition. That is, λ_i is an eigenvalue of A, and v_i is the associated eigenvector.

4: **for** $i = 1, 2, \ldots, n$ **do**

5:     **if** $\lambda_i \leq 0$ **then**

6:         $\lambda_i = 0.01 \times \mathrm{rand}$.

7:     **end if**

8: **end for**

9: $A \leftarrow \sum_{i=1}^n \lambda_i v_i v_i^\top$.

10: $\Theta_0 := (A, B, C)$.

$\mathrm{vec}_{\mathrm{sym}}(A)$

$$= \begin{pmatrix} a_{11} & a_{21} & a_{31} & a_{22} & a_{32} & a_{33} \end{pmatrix}^\top.$$

The parameter θ defined by (43) is a global coordinate system for the manifold M, and, thus, we regard Θ on M as θ. Hence, we write the prediction error vector $e(\Theta)$ defined by (11) as $e(\theta)$ and the objective function $f_1(\Theta)$ as $V(\theta) := \|e(\theta)\|_2^2$. The aim of the GN method is to minimize $V(\theta)$.

The update formula of the GN method is given by

$$\theta_{k+1} = \theta_k + t_k \Delta\theta_k \tag{44}$$

where $t_k > 0$ is a step size, and $\Delta\theta_k$ satisfies

$$J(\theta_k)^\top J(\theta_k)\Delta\theta_k = -J(\theta_k)^\top e(\theta_k) \tag{45}$$

with

$$J(\theta) := \frac{\partial e}{\partial \theta}(\theta) \in \mathbf{R}^{pK \times n_\theta}. \tag{46}$$

Here

$$\frac{\partial e}{\partial \theta_i}(\theta) = -\left(\left(\frac{\partial \hat{y}_1}{\partial \theta_i}(\theta)\right)^\top\ \left(\frac{\partial \hat{y}_2}{\partial \theta_i}(\theta)\right)^\top\ \cdots\ \left(\frac{\partial \hat{y}_K}{\partial \theta_i}(\theta)\right)^\top\right)^\top$$

and

$$\begin{cases} \dfrac{\partial \hat{y}_j}{\partial \theta_i}(\theta) = \dfrac{\partial C}{\partial \theta_i}\hat{x}_j + C\dfrac{\partial \hat{x}_j}{\partial \theta_i} \\ \dfrac{\partial \hat{x}_j}{\partial \theta_i} = \dfrac{\partial A}{\partial \theta_i}\hat{x}_{j-1} + A\dfrac{\partial \hat{x}_{j-1}}{\partial \theta_i} + \dfrac{\partial B}{\partial \theta_i}u_{j-1} \end{cases}$$

with $\frac{\partial \hat{x}_0}{\partial \theta_i} = 0$. Note that if $\theta_k \in N \subset M$ and the step size t_k is sufficiently small, then $\Delta\theta_k$ can be regarded as an element of $T_{\theta_k}N$. In this case

$$e(\theta_k + t_k\Delta\theta_k) \approx e(\theta_k) + t_k J(\theta_k)\Delta\theta_k \tag{47}$$

and

$$J(\theta_k) : T_{\theta_k}N \to \mathbf{R}^{pK}. \tag{48}$$

Note that (43) is an overparameterization. This means that different θ may have the equivalent input–output properties. In fact, from Section III-C, each element (A, B, C) of $\pi^{-1}([\Theta]) \subset N$, which can be regarded as different θ, has the same input–output properties, where the dimension of $\pi^{-1}([\Theta])$ is

$$\dim \pi^{-1}([\Theta]) = \dim N - \dim N/O(n) = \frac{n(n-1)}{2}. \tag{49}$$

Equation (49) follows from Proposition 1 in Appendix B and [35, Prop. 3.4.4]. Hence, if $\Delta\theta_k \in T_{\theta_k}\pi^{-1}([\Theta_k]) \subset T_{\theta_k}N$ under the identification of Θ_k and θ_k, it follows from (47) that $J(\theta_k)\Delta\theta_k = 0$, that is, $\Delta\theta_k \in \text{Ker } J(\theta_k)$. Therefore

$$T_{\theta_k}\pi^{-1}([\Theta_k]) \subset \text{Ker } J(\theta_k)$$

and (49) yields

$$\dim \text{Ker } J(\theta_k) \geq \frac{n(n-1)}{2}. \tag{50}$$

It follows from (50) that the matrix $J(\theta_k)$ is rank-deficient, and, thus, there are infinitely many solutions to (45). In the data-driven local coordinates introduced in [10], $\Delta\theta_k$ is chosen as

$$\Delta\theta_k = -V_1 S_1^{-1} U_1^\top e(\theta_k) \tag{51}$$

as shown in [13], where $U_1 S_1 V_1^\top$ is the truncated singular value decomposition of $J(\theta_k)$, and $S_1 \in \mathbf{R}^{n_\theta \times n_\theta}$ is a diagonal matrix.

Note that update formula (44) preserves the symmetry of A, but does not, in general, preserve the positive-definiteness. More precisely, if θ_k is contained in the manifold M or N, θ_{k+1} given by (44) is also contained in M or N by choosing sufficiently small $t_k > 0$. This is because M and N are open sets. However, if $t_k > 0$ is too small, the value of the objective function V does not change very much. Thus, we need to choose sufficiently large $t_k > 0$, but then θ_{k+1} may not be contained in M nor in N, as demonstrated in Section VI. That is, it is difficult to determine an appropriate step size t_k for some examples.

Instead of using the full parameterization $\theta \in \mathbf{R}^{n_\theta}$ defined by (43), we can use canonical forms of linear systems, which reduce the number of free parameters. However, canonical forms may lead to numerically ill-conditioned problems due to noise, as pointed out in [10, Sec. 1 and 4]. Moreover, the results of Section VI-B1 justify this point, because Problem 3 can be regarded as a case of using a canonical form. Thus, in practice, the use of canonical forms for system identification may not be adequate.

To resolve the numerically ill-conditioned problem, the use of overlapping parameterization has been proposed in [46], and Hanzon and Ober [33] introduced block-balanced input normal forms, which are overlapping parameterizations. However, this approach requires monitoring the condition of the parametrization and switching to a new structure if the current structure is bad. That is, this needs a number of extra calculations, which are not necessary in the case of Algorithm 1 based on Riemannian optimization.

Remark 5: Although the Jacobian $J(\theta)$ is rank-deficient in our case, if $J(\theta)$ is of full-rank, update formula (44) for the GN method can be regarded as a Riemannian SD method for the specific choice of the Riemannian metric

$$g_\theta^{\text{GN}}(\dot\theta_1, \dot\theta_2) := \dot\theta_1^\top R(\theta)\dot\theta_2 \tag{52}$$

with $R(\theta) := 2J(\theta)^\top J(\theta)$ into $\mathbf{R}^{n_\theta}$, as stated in [8] and [11]. The Riemannian gradient of the objective function V at $\theta \in \mathbf{R}^{n_\theta}$ is given by $R(\theta)^{-1}\frac{\partial V}{\partial\theta}(\theta)$, where the gradient is called the natural gradient in the Riemannian manifold $\mathbf{R}^{n_\theta}$ endowed with the Riemannian metric (52) [47]. Using the natural gradient, update formula (44) can be expressed as

$$\theta_{k+1} = \theta_k - R(\theta_k)^{-1}\frac{\partial V}{\partial\theta}(\theta_k). \tag{53}$$

TABLE I
PARAMETERS OF SYSTEM (54)

$V(t) \in \mathbf{R}^n$	node-voltage vector
$u(t) \in \mathbf{R}^m$	voltage-source vector
$y(t) \in \mathbf{R}^p$	measurement output vector
$C_{\text{cap}} \in \text{Sym}_+(n)$	capacitance matrix
$G_{\text{con}} \in \text{Sym}_+(n)$	conductance matrix
$\mathcal{L}_{\text{res}} \in \text{Sym}(n)$	Laplacian matrix associated with $\mathcal{G}$
$\tilde{G} \in \mathbf{R}^{n \times m}$ and $\tilde{H} \in \mathbf{R}^{p \times n}$	constant matrices

VI. NUMERICAL SIMULATIONS

In this section, we demonstrate the effectiveness of the proposed method. To this end, we evaluate the identified systems using various indices, in addition to the value of the objective function in Problems 1, 2, and 3, to prevent overfitting to noisy data. Note that, in the simulations, we used MOESP [19] as the subspace method for step 1 in Algorithm 2. This was implemented in the *system identification toolbox* of MATLAB. Hence, we can easily implement Algorithm 2.

We consider identification problems of the RC electrical network system [1], [2] represented as the undirected graph $\mathcal{G} = \{\{1, 2, \ldots, n\}, \mathcal{E}\}$, which is composed of n nodes and the set $\mathcal{E}$ of k undirected edges. A mathematical model of the system is described by

$$\begin{cases} C_{\text{cap}}\dot{V}(t) = -(\mathcal{L}_{\text{res}} + G_{\text{con}})V(t) + \tilde{G}u(t) \\ y(t) = \tilde{H}V(t). \end{cases} \tag{54}$$

Table I explains the parameters of system (54). Here, $\mathcal{L}_{\text{res}} := \mathcal{B}R_{\text{res}}^{-1}\mathcal{B}^\top$ is a symmetric positive semidefinite matrix, $\mathcal{B} \in \mathbf{R}^{n \times k}$ is the incidence matrix of $\mathcal{G}$, and $R_{\text{res}} \in \text{Diag}_+(n)$ is the resistance matrix. The incidence matrix $\mathcal{B} = (\mathcal{B}_{ij}) \in \mathbf{R}^{n \times k}$ is defined by

$$\mathcal{B}_{ij} := \begin{cases} 1, & \text{if } i \text{ is the source node of edge } j \\ -1, & \text{if } i \text{ is the sink node of edge } j \\ 0, & \text{otherwise.} \end{cases}$$

System (54) can be transformed into (1) by defining $x(t) := C_{\text{cap}}^{1/2}V(t)$;

$$\begin{cases} \dot{x}(t) = Fx(t) + Gu(t) \\ y(t) = Hx(t) \end{cases} \tag{55}$$

with

$$F := -C_{\text{cap}}^{-1/2}(\mathcal{L}_{\text{res}} + G_{\text{con}})C_{\text{cap}}^{-1/2} \in \mathbf{R}^{n \times n}$$

$$G := C_{\text{cap}}^{-1/2}\tilde{G} \in \mathbf{R}^{n \times m}$$

$$H := \tilde{H}C_{\text{cap}}^{1/2} \in \mathbf{R}^{p \times n}.$$

Note that the matrix $-F$ is contained in $\text{Sym}_+(n)$, and, thus, F is stable. That is, all the eigenvalues of F are negative.

Although we consider mathematical model (55) to be noise-free, measurement noise is inevitable in practice, as explained in [9, Sec. 4.3]. Thus, we assume that the true system is given by

$$\begin{cases} x_{k+1} = Ax_k + Bu_k \\ y_k = Cx_k + v_k \end{cases} \tag{56}$$

where A, B, and C are defined by (3), (4), and (5), respectively, and $v_k \in \mathbf{R}^p$ is measurement noise. That is, the input/output data (u_k, y_k) is generated by (56). Because F is stable, the matrix A is also stable. That is, all eigenvalues of A are in the interval (0, 1). The signal-to-noise ratio of system (56) is defined as

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\sum_{k=0}^{K} \|y_k - v_k\|_2^2}{\sum_{k=0}^{K} \|v_k\|_2^2} \right). \tag{57}$$

In the following, we present the results of numerical simulations for SISO and MIMO cases. For SISO cases, we illustrate a frequency response using the Bode plots. For MIMO cases, the values of various indices are given, because Bode plots of MIMO cases do not clarify the distance between the true and estimated systems.

To this end, we set $n = 20$, and generated the undirected graph $\mathcal{G}$ using the Watts and Strogatz model [48] with 20 nodes of mean degree 10 and rewiring probability 0.4. Additionally, C_{cap}, R_{res}, and G_{con} were given by

$$\begin{cases} C_{\mathrm{cap}} = 10 \times \mathrm{diag}(\mathrm{rand}, \mathrm{rand}, \ldots, \mathrm{rand}) \\ R_{\mathrm{res}} = 0.1 \times \mathrm{diag}(1, 1, \ldots, 1) \\ G_{\mathrm{con}} = \mathrm{diag}(\mathrm{rand}, \mathrm{rand}, \ldots, \mathrm{rand}) \end{cases} \tag{58}$$

where each rand denotes a uniformly distributed random number in the interval (0, 1). Moreover, we generated each component of u_k from the Gaussian random distribution with mean 0 and variance 100, and the components of v_k from the Gaussian random distribution with mean 0 and variance σ^2. The sampling interval h was 0.1.

We denote the results given by Algorithm 1 for solving Problems 1, 2, and 3 as CG_1, CG_2, and CG_3, respectively. Moreover, we write SD to denote the Riemannian SD method, as briefly explained in Remark 4.

A. SISO Case

First, we considered SISO cases with $m = p = 1$. The parameters $\tilde{G}$ and $\tilde{H}$ were given by

$$\tilde{G} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \tilde{H} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}.$$

1) Identification by the GN Method: Fig. 3 illustrates the eigenvalues of the true matrix A corresponding to F of system (55), the estimated matrix A produced by Algorithm 2, and the estimated matrix A provided by the prediction error method using the GN method with the update formula (44) and (51), as explained in Section V, after ten iterations. Here, we used the result (A, B, C) obtained by Algorithm 2 as the initial point of the GN method, and the step sizes t_k in (44) were $t_k = 10^{-9}$ for all $k \in \{1, 2, \ldots, 10\}$. According to Fig. 3, the prediction error method using the GN method did not provide $\Theta \in M$. In fact, some eigenvalues of A produced by the GN method took negative values, whereas all eigenvalues of the true matrix A are positive. Moreover, we confirmed the following results.

1) When $t_k = 10^{-9}$, the positive-definite property of matrix A produced by the GN method was lost after only a few iterations.
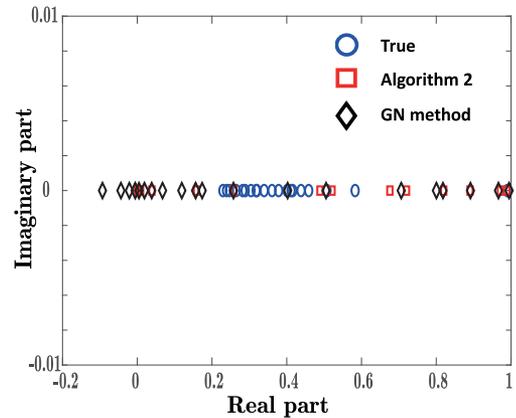
Fig. 3. Eigenvalues of A in the original system, estimated system produced by Algorithm 2, and estimated system provided by the GN method after ten iterations.
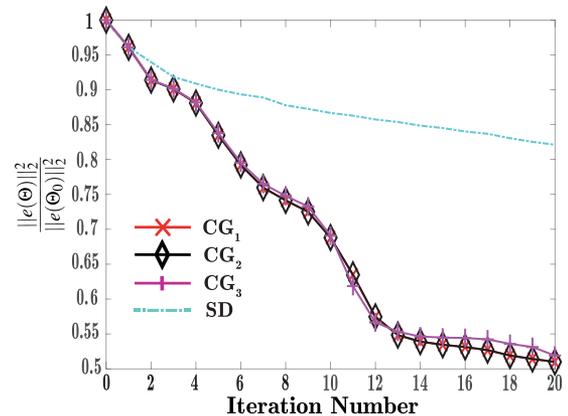
Fig. 4. Relative objective values obtained by CG_1, CG_2, CG_3, and SD.

2) If we set $t_k > 10^{-9}$, the symmetric matrices A produced by the GN method, in many cases, were unstable after ten iterations.

3) Even if we set $t_k < 10^{-9}$, some eigenvalues of the symmetric matrices A produced by the GN method were negative after ten iterations.

Thus, the GN method described in Section V is not adequate for solving our problem. Hence, we hereafter compare CG_1, CG_2, CG_3, SD, and Algorithm 2.

2) Comparison of CG_1, CG_2, CG_3, and SD: Figs. 4 and 5 illustrate a comparison of CG_1, CG_2, CG_3, and SD with $K = 600$, $\sigma^2 = 0.1$, and SNR $= 12.803$. Here, Θ_0 in Fig. 4 was obtained using Algorithm 2. According to these figures, the results for CG_1 and CG_2 completely overlap, and Fig. 4 demonstrates that the convergence speeds of CG_1, CG_2, and CG_3 are superior to that of SD. Moreover, Fig. 5 shows that CG_1, CG_2, CG_3, and SD improve the frequency response of Algorithm 2. In particular, the Bode plots of the estimated systems obtained by CG_1 and CG_2 are almost the same as that of the true system, unlike CG_3, SD, and Algorithm 2. Note that no destabilization occurred for CG_1, CG_2, or CG_3.
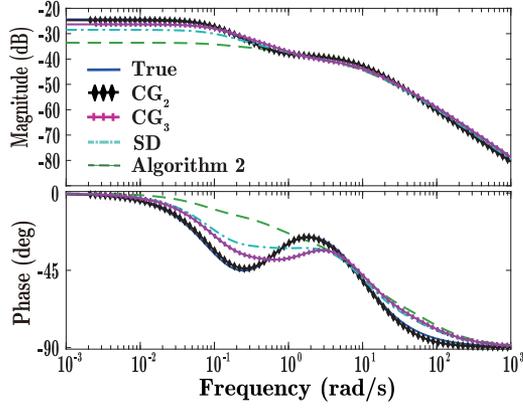
Fig. 5. Bode plots of true and estimated systems obtained by CG_1, CG_2, CG_3, SD, and Algorithm 2. Because the Bode plot of the estimated system obtained by CG_1 completely overlapped with that obtained by CG_2, the illustration of CG_1 was omitted.

TABLE II
NUMBER OF UNSTABLE CASES OVER 20 ITERATIONS WHEN $K = 400$

Algorithm	$\sigma^2 = 0.05$	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$
CG_1	3	2	3
CG_2	1	1	2
CG_3	15	17	26

B. MIMO Case

Next, we considered the MIMO case with $m = p = 2$. The parameters $\tilde{G}$ and $\tilde{H}$ were given by

$$\tilde{G} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \quad \tilde{H} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{pmatrix}.$$

As with the SISO case, the conventional GN method did not produce $A_{est} \in \mathrm{Sym}_+(n)$ and the convergence speeds of CG_1, CG_2, and CG_3 were faster than that of SD. Thus, we present the results of comparisons among CG_1, CG_2, CG_3, and Algorithm 2.

1) Stability of the Estimated Matrices $A \in \mathrm{Sym}_+(n)$ *Produced by* CG_1, CG_2, *and* CG_3: Because all of the matrices A_{est} [estimates of A in true system (56)] produced by CG_1, CG_2, and CG_3 are contained in $\mathrm{Sym}_+(n)$, the eigenvalues of A_{est} are positive real numbers, unlike the eigenvalues given by the conventional GN method. However, even if A is stable, A_{est} may be unstable.

Thus, we compared the stability of the estimated matrices A_{est} provided by CG_1, CG_2, and CG_3. We performed numerical simulations 100 times with $\sigma^2 = 0.05$, $\sigma^2 = 0.1$, and $\sigma^2 = 0.5$. Table II presents the number of unstable cases over 20 iterations when $K = 400$. According to Table II, the rate of instability in A_{est} produced by CG_3 is far higher than when using CG_1 or CG_2. Because we used different C_{cap}, R_{res}, and G_{con} for each simulation, the SNR defined by (57) was also different. Table III describes the relation between σ^2 and SNR. Here, SNR_{ave} and SNR_{dev} are the average and standard deviation over 10 000

TABLE III
SNR_{ave} AND SNR_{dev} WHEN $K = 400$

	$\sigma^2 = 0.05$	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$
SNR_{ave}	26.261	20.174	6.212
SNR_{dev}	6.686	6.660	6.686

simulations, defined by

$$\mathrm{SNR}_{ave} := \frac{\sum_{i=1}^{10\,000} \mathrm{SNR}_i}{10\,000}$$

$$\mathrm{SNR}_{dev} := \sqrt{\frac{\sum_{i=1}^{10\,000} (\mathrm{SNR}_i - \mathrm{SNR}_{ave})^2}{10\,000}}$$

where SNR_i denotes SNR in the ith simulation. According to Table III, SNR_{ave} decreases as σ^2 increases, although the SNR_{dev} values are similar. We also obtained similar results to those described in Tables II and III for different values of K. Hence, we conclude that the rate of instability in A_{est} produced by CG_3 is far higher than those when using CG_1 and CG_2. This is in contrast to the SISO case. In addition, the instability rate for CG_1 and CG_2 is independent of SNR, unlike that for CG_3.

The reason for the high instability rate produced by CG_3 is that the noise component of the output directly influences the diagonal matrix A_{est}, i.e., eigenvalues of A_{est}. This is essentially the same phenomenon observed in system identification problems, whereby canonical forms lead to numerically ill-conditioned problems [10]. In contrast, the noise component does not have a significant effect on the eigenvalues of the estimated matrices A_{est} produced by CG_1 and CG_2, because the matrices are not diagonal.

2) Evaluation of Proposed Methods: We evaluated the results with respect to the cost function $\|e(\Theta_{est})\|_2^2$, the relative H^2 and H^∞ norms, and the maximum eigenvalues $\lambda_{\max}(F_{est})$ of the estimated matrix F_{est} of F. Here, $\lambda_{\max}(F)$ was -0.086 in all cases. Note that the maximum eigenvalue $\lambda_{\max}(F_{est})$ is important, because the transient state $\hat{x}(t)$ in system (1) is dominated by $\lambda_{\max}(F)$ under $\hat{u}(t) = 0$. That is, if $\lambda_{\max}(F)$ and $\lambda_{\max}(F_{est})$ are closer, we can expect the true and estimated transient states to be more similar. When we used our proposed methods CG_1, CG_2, and CG_3, the number of iterations was set to 20. Increasing the number of iterations would decrease the value of the objective function $\|e(\Theta_{est})\|_2^2$. However, other indices such as g_2, g_∞, and $\lambda_{\max}(F_{est})$ may become worse due to overfitting with noisy data.

To define the relative H^2 and H^∞ norms, we use T and T_{est} as the transfer functions from the input u to the output y of the true and estimated systems, respectively. That is

$$T(s) := C(sI_n - F)^{-1}G, \quad s \in \mathbf{C}$$

$$T_{est}(s) := C_{est}(sI_n - F_{est})^{-1}G_{est}$$

where G_{est} and C_{est} are the estimated matrices of G and C, respectively. Here, we estimate the matrices F_{est} and G_{est} using (6) and (7), respectively. Using T and T_{est}, we define the following relative H^2 and H^∞ norms

$$g_2 := \frac{\|T - T_{est}\|_{H^2}}{\|T\|_{H^2}} \quad \text{and} \quad g_\infty := \frac{\|T - T_{est}\|_{H^\infty}}{\|T\|_{H^\infty}}.$$

Tables IV, V, VI, VII, and VIII present values of $\|e(\Theta_{est})\|_2^2$, g_2, g_∞, and $\lambda_{\max}(F_{est})$ for different K, as given by estimating

TABLE IV
EVALUATION RESULTS WHEN $K = 200$ AND SNR $= 15.043$

$K = 200$	$||e(\Theta_{\text{est}})||_2^2$	g_2	g_∞	$\lambda_{\max}(F_{\text{est}})$
Algorithm 2	23.097	0.660	0.868	-0.138
CG_1	4.675	0.288	0.328	-0.075
CG_2	4.662	0.287	0.322	-0.075
Hybrid CG	4.759	0.241	0.258	-0.098

TABLE V
EVALUATION RESULTS WHEN $K = 400$ AND SNR $= 15.286$

$K = 400$	$||e(\Theta_{\text{est}})||_2^2$	g_2	g_∞	$\lambda_{\max}(F_{\text{est}})$
Algorithm 2	76.360	0.576	0.774	-0.117
CG_1	5.231	0.124	0.123	-0.064
CG_2	5.226	0.124	0.122	-0.064
Hybrid CG	4.900	0.111	0.087	-0.065

TABLE VI
EVALUATION RESULTS WHEN $K = 600$ AND SNR $= 15.669$

$K = 600$	$||e(\Theta_{\text{est}})||_2^2$	g_2	g_∞	$\lambda_{\max}(F_{\text{est}})$
Algorithm 2	75.850	0.624	0.835	-0.258
CG_1	7.028	0.144	0.218	-0.067
CG_2	7.025	0.144	0.217	-0.067
Hybrid CG	6.538	0.072	0.052	-0.100

TABLE VII
EVALUATION RESULTS WHEN $K = 800$ AND SNR $= 16.714$

$K = 800$	$||e(\Theta_{\text{est}})||_2^2$	g_2	g_∞	$\lambda_{\max}(F_{\text{est}})$
Algorithm 2	358.798	0.653	0.868	-0.387
CG_1	59.391	0.343	0.263	-0.115
CG_2	59.000	0.343	0.265	-0.115
Hybrid CG	20.623	0.191	0.211	-0.129

TABLE VIII
EVALUATION RESULTS WHEN $K = 1000$ AND SNR $= 16.380$

$K = 1000$	$||e(\Theta_{\text{est}})||_2^2$	g_2	g_∞	$\lambda_{\max}(F_{\text{est}})$
Algorithm 2	320.573	0.620	0.831	-0.237
CG_1	17.541	0.141	0.058	-0.082
CG_2	17.464	0.141	0.057	-0.082
Hybrid CG	12.865	0.094	0.055	-0.093

Θ_{est}, F_{est}, G_{est}, and C_{est} using Algorithm 2, CG_1, CG_2, and a combined CG approach called Hybrid CG. Hybrid CG is a combination of CG_1 and CG_2 obtained by applying CG_1 for the first 15 iterations and CG_2 for the next 5 iterations. Note that in terms of the various indices, we confirmed that Algorithm 2 provides a considerably better initial point Θ_0 in Algorithm 1 than randomly choosing $\Theta_0 \in M$.

According to Tables IV, V, VI, VII, and VIII, the results for $||e(\Theta_{\text{est}})||_2^2$, g_2, g_∞, and $\lambda_{\max}(F_{\text{est}})$ given by CG_1 and CG_2 are better than those given by Algorithm 2 for all K. The results from CG_1 and CG_2 are almost the same for all K. However, with the exception of $\lambda_{\max}(F_{\text{est}})$, the results from Hybrid CG are superior to those of CG_1 and CG_2. Even when the combination of iterations was changed, the results of Hybrid CG were better than those of CG_1 and CG_2 in many cases. Moreover, we should note that the evaluation results may be worse as the data length K increases.

Remark 6: In addition to our proposed methods, the MATLAB command pem can identify F, G, and H in (55), as explained in [49]. However, for the same initial systems produced by Algorithm 2, the results of our proposed methods

were considerably better than those of pem in MATLAB in some cases, whereas in certain other cases, pem yielded better results. A main reason may be the nonconvexity of the objective function, that is, because there are many local optimal solutions, different algorithms produce different solutions. Thus, the models produced by our proposed methods may be good in some cases where the model produced by pem in MATLAB is not satisfactory.

VII. CONCLUSION AND FUTURE WORK

We developed identification methods for linear continuous-time symmetric systems using Riemannian optimization. For this, we formulated three least-squares problems of minimizing the sum of squared errors on Riemannian manifolds, and described the geometry of each problem. In particular, we examined the quotient geometry in one problem in depth. We proposed Riemannian CG methods for the three problems, and selected initial points using the modified MOESP method. The results from a series of numerical simulations demonstrated the effectiveness of our proposed methods with comparisons to the traditional GN method.

The following problems should be addressed in future studies.

1) As mentioned in Remark 3, system (1) does not correspond to a symmetric continuous-time system discussed in [3] and [40]. To identify such systems, we need to develop a novel method that is fundamentally different from the methods proposed in this article.

2) In Section VI-B2, we confirmed that the results produced by Hybrid CG, a combination of CG_1 and CG_2, were better than those of CG_1 and CG_2 in many cases. It would be interesting to study how the combination of iterations of CG_1 and CG_2 should be determined.

3) It is shown in [31, Lem. 2] that the manifold of transfer functions of SISO systems, i.e., $m = p = 1$, is partitioned into multiple connected components. Thus, it is expected that $N/O(n)$ with $m = p = 1$ will have multiple connected components, because each element in $N/O(n)$ corresponds to a transfer function. If this is the case, different initial points on the different connected components will converge to different points, and, thus, initial points on $N/O(n)$ may considerably affect the system identification results. In fact, we have confirmed that Algorithm 2 provides a better initial point than a random choice. This provides a practical insight, and so it would be interesting to study whether or not the expectation is true.

4) In this article, we proposed methods for identifying a target system as (1) with no noise. As illustrated in Section VI, our proposed methods are effective for identifying (1), even if the output data were noisy. However, if we were to consider the effect of noise on our methods, we may be able to derive better algorithms. Thus, it is desirable to extend our proposed methods under the consideration of noise.

5) As mentioned in Section I, our proposed method can directly identify discrete-time system (2) with a symmetric-positive definite matrix A. Thus, it is desirable to find a practical example that can be expressed by the

discrete-time system without going through continuous-time system (1) with a symmetric matrix F.

APPENDIX

A. Geometry of the Manifold $\mathrm{Sym}_+(n)$

We review the geometry of $\mathrm{Sym}_+(n)$ to develop optimization algorithms for solving our problems (for a detailed explanation, see [28] and [50]).

For $\xi_1, \xi_2 \in T_S \mathrm{Sym}_+(n)$, we define Riemannian metric into $\mathrm{Sym}_+(n)$ as

$$\langle \xi_1, \xi_2 \rangle_S := \mathrm{tr}(S^{-1}\xi_1 S^{-1}\xi_2). \tag{59}$$

Let $f : \mathrm{Sym}_+(n) \to \mathbf{R}$ be a smooth function and $\bar{f}$ be the extension of f to Euclidean space $\mathbf{R}^{n \times n}$. Riemannian gradient $\mathrm{grad}\, f(S)$ with respect to Riemannian metric (59) is given by

$$\mathrm{grad}\, f(S) = S\mathrm{sym}(\nabla \bar{f}(S))S \tag{60}$$

where $\nabla \bar{f}(S)$ denotes the Euclidean gradient of $\bar{f}$ at $S \in \mathrm{Sym}_+(n)$. The exponential map on $\mathrm{Sym}_+(n)$ is given by

$$\mathrm{Exp}_S(\xi) = S^{\frac{1}{2}} \exp(S^{-\frac{1}{2}}\xi S^{-\frac{1}{2}}) S^{\frac{1}{2}}$$
$$= S \exp(S^{-1}\xi) \tag{61}$$

where $\exp$ is the matrix exponential function.

We note that Riemannian metric (59) is essentially the same with Fisher information metric

$$g_S^{\mathrm{FIM}} := \mathrm{E}(\mathrm{D}l_x(S) \otimes \mathrm{D}l_x(S))$$

where

$$l_x(S) := \log p(x|S^{-1}).$$

Here, E is the expectation operator with respect to $p(x|S^{-1})$, $\otimes$ is the tensor product, and $p(x|S^{-1})$ denotes the Gaussian distribution with zero mean vector and covariance $S^{-1} \in \mathrm{Sym}_+(n)$, i.e.,

$$p(x|S^{-1}) = \sqrt{\frac{\det S}{(2\pi)^n}} \exp\left(-\frac{1}{2}x^\top Sx\right).$$

Thus, $l_x(S)$ is the log-likelihood function of $p(x|S^{-1})$, and

$$l_x(S) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log \det S - \frac{1}{2}x^\top Sx. \tag{62}$$

To see the relation between (59) and g_S^{FIM}, we use

$$g_S^{\mathrm{FIM}} = -\mathrm{E}(\mathrm{D}^2 l_x(S)) \tag{63}$$

where $\mathrm{D}^2 l_x(S) : \mathrm{Sym}(n) \times \mathrm{Sym}(n) \to \mathbf{R}$ is the second derivative of l_x at S. Equation (63) can be found in [50, Th. 1]. The directional derivative of $l_x : \mathrm{Sym}(n) \to \mathbf{R}$ at $S \in \mathrm{Sym}_+(n)$ along $\xi \in T_S \mathrm{Sym}_+(n) \cong \mathrm{Sym}(n)$ is given by

$$\mathrm{D}l_x(S)[\xi] = \frac{1}{2}\mathrm{tr}(S^{-1}\xi) - \frac{1}{2}\mathrm{tr}(xx^\top \xi) \tag{64}$$

where the first term of the right-hand side is obtained by using Jacobi's formula $\mathrm{D}\det S[\xi] = \mathrm{tr}(\det(S)S^{-1}\xi)$. We define the inner product of $\mathrm{Sym}(n)$ as $\mathrm{tr}(\xi_1 \xi_2)$ for any $\xi_1, \xi_2 \in \mathrm{Sym}(n)$.

Then, from (64), the gradient of l_x at $S \in \mathrm{Sym}_+(n)$ is provided as

$$\nabla l_x(S) = \frac{1}{2}(S^{-1} - xx^\top).$$

Moreover, according to [35], the Hessian of l_x at S is given by

$$\mathrm{Hess}\, l_x(S)[\xi] = \mathrm{D}\nabla l_x(S)[\xi] = -\frac{1}{2}S^{-1}\xi S^{-1} \tag{65}$$

and

$$\mathrm{D}^2 l_x(S)[\xi_1, \xi_2] = \mathrm{tr}(\mathrm{Hess}\, l_x(S)[\xi_1]\xi_2). \tag{66}$$

Substituting (65) into (66), we obtain that

$$\mathrm{D}^2 l_x(S)[\xi_1, \xi_2] = -\frac{1}{2}\mathrm{tr}(S^{-1}\xi_1 S^{-1}\xi_2). \tag{67}$$

That is, $\mathrm{D}^2 l_x(S)[\xi_1, \xi_2]$ is independent of x. Hence, from (59), (63), and (67), we obtain that

$$g_S^{\mathrm{FIM}}(\xi_1, \xi_2) = -\mathrm{D}^2 l_x(S)[\xi_1, \xi_2] = \frac{1}{2}\langle \xi_1, \xi_2 \rangle_S.$$

Thus, Riemannian metric (59) is essentially the same with Fisher information metric g_S^{FIM}.

B. Quotient Manifold Theorem

This appendix explains how to use the following quotient manifold theorem as shown in [29, Th. 21.10] in our discussion of Section III-C.

Proposition 1: Suppose that $\mathcal{G}$ is a Lie group acting smoothly, freely, and properly on a smooth manifold $\mathcal{M}$. Then, the orbit space $\mathcal{M}/\mathcal{G}$ is a topological manifold of dimension equal to $\dim \mathcal{M} - \dim \mathcal{G}$, and has a unique smooth structure with the property that the quotient map $\pi : \mathcal{M} \to \mathcal{M}/\mathcal{G}$ is a smooth submersion.

Here, the action $\cdot$ of Lie group $\mathcal{G}$ on a smooth manifold $\mathcal{M}$ is called

1) free if $\{g \in \mathcal{G} \,|\, g \cdot p = p\} = \{e\}$ for each $p \in \mathcal{M}$, where e is the identity of $\mathcal{G}$;
2) proper if the map $f : \mathcal{G} \times \mathcal{M} \to \mathcal{M} \times \mathcal{M}$ defined by $(g, p) \mapsto (g \cdot p, p)$ is a proper map. That is, for every compact set $K \in \mathcal{M} \times \mathcal{M}$, the preimage $f^{-1}(K) \subset \mathcal{G} \times \mathcal{M}$ is compact.

We can apply the quotient manifold theorem in our case, if (12) is a free and proper action. This is because the orthogonal group $O(n)$ is a Lie group, and (12) is a smooth action on the smooth manifold N. We, thus, confirm that action (12) is free and proper in Section III-C.

C. Proof of Theorem 1

In this appendix, we provide a proof of Theorem 1 without deriving specific expressions of the vertical and horizontal spaces. More concretely, we prove a more general theorem, and point out that Theorem 1 is a corollary of the general theorem.

Let $\mathcal{M}$ be a Riemannian manifold with the Riemannian metric $\langle \cdot, \cdot \rangle$, and let $\mathcal{G}$ be a group that smoothly acts on $\mathcal{M}$. Here, we call $\phi_g : \mathcal{M} \to \mathcal{M}$ a smooth group action if ϕ_g is smooth and satisfies the following.

1) For any $g_1, g_2 \in \mathcal{G}$ and any $x \in \mathcal{M}$, $\phi_{g_1 g_2}(x) = \phi_{g_1}(\phi_{g_2}(x))$ holds.

2) For the identity element $1_\mathcal{G} \in \mathcal{G}$ and any $x \in \mathcal{M}$, $\phi_{1_\mathcal{G}}(x) = x$ holds.

We write the derivative map of ϕ_g at $x \in \mathcal{M}$ as $\mathrm{D}\phi_g(x) : T_x\mathcal{M} \to T_{\phi_g(x)}\mathcal{M}$. By definition

$$
\begin{aligned}
\mathrm{D}\phi_{1_\mathcal{G}}(x) &= \mathrm{D}(\phi_{g^{-1}} \circ \phi_g)(x) \\
&= \mathrm{D}\phi_{g^{-1}}(\phi_g(x)) \circ \mathrm{D}\phi_g(x)
\end{aligned}
$$

and thus

$$
\mathrm{D}\phi_{g^{-1}}(\phi_g(x)) = (\mathrm{D}\phi_g(x))^{-1}. \tag{68}
$$

Let $\mathcal{M}/\mathcal{G}$ be a quotient Riemannian manifold of $\mathcal{M}$ with the canonical projection $\pi : \mathcal{M} \to \mathcal{M}/\mathcal{G}$. That is, $\pi(x) = [x]$ for any $x \in \mathcal{M}$, where $[x] := \{x_1 \in \mathcal{M} \,|\, x = \phi_g(x_1) \text{ for some } g \in \mathcal{G}\}$. Let $\mathcal{V}_x := T_x\pi^{-1}([x])$ be the vertical space in $T_x\mathcal{M}$, and let $\mathcal{H}_x$ be the horizontal space that is the orthogonal complement of $\mathcal{V}_x$ with respect to the metric $\langle \cdot, \cdot \rangle$. Let V be a vector space in $T_x\mathcal{M}$, and let

$$
\mathrm{D}\phi_g(x)[V] := \{\mathrm{D}\phi_g(x)[\xi] \,|\, \xi \in V\}.
$$

Lemma 1: For any $g \in \mathcal{G}$ and $x \in \mathcal{M}$

$$
\mathcal{V}_{\phi_g(x)} = \mathrm{D}\phi_g(x)[\mathcal{V}_x]. \tag{69}
$$

Proof: Let $\xi \in \mathcal{V}_{\phi_g(x)} = T_{\phi_g(x)}\pi^{-1}([\phi_g(x)])$. Then, there exists a curve γ such that $\gamma(0) = \phi_g(x)$ and $\dot{\gamma}(0) = \xi$. Because $\mathcal{G}$ acts on $\mathcal{M}$, $\gamma_0(t) := \phi_{g^{-1}}(\gamma(t))$ is on $\pi^{-1}([x])$. We have that $\gamma_0(0) = \phi_{g^{-1}}(\phi_g(x)) = x$, and $\dot{\gamma}_0(t) = \mathrm{D}\phi_{g^{-1}}(\gamma(t))[\dot{\gamma}(t)]$. Hence, it follows from (68) that

$$
\begin{aligned}
\dot{\gamma}_0(0) &= \mathrm{D}\phi_{g^{-1}}(\phi_g(x))[\xi] \\
&= (\mathrm{D}\phi_g(x))^{-1}[\xi] \in T_x\pi^{-1}([x]) = \mathcal{V}_x
\end{aligned}
$$

and thus $\xi \in \mathrm{D}\phi_g(x)[\mathcal{V}_x]$. That is

$$
\mathcal{V}_{\phi_g(x)} \subset \mathrm{D}\phi_g(x)[\mathcal{V}_x].
$$

Considering the dimension of both sides, we obtain (69). ∎

Lemma 1 implies the following theorem.

Theorem 4: Suppose that the group action ϕ_g is an isometry in terms of Riemannian metric $\langle \cdot, \cdot \rangle$; i.e., for any $g \in \mathcal{G}$ and any $\xi_1, \xi_2 \in T_x\mathcal{M}$

$$
\langle \mathrm{D}\phi_g(x)[\xi_1], \mathrm{D}\phi_g(x)[\xi_2] \rangle_{\phi_g(x)} = \langle \xi_1, \xi_2 \rangle_x. \tag{70}
$$

Then

$$
\mathcal{H}_{\phi_g(x)} = \mathrm{D}\phi_g(x)[\mathcal{H}_x]. \tag{71}
$$

Proof: Taking the orthogonal complement of both sides of (69), we have that

$$
\mathcal{H}_{\phi_g(x)} = (\mathrm{D}\phi_g(x)[\mathcal{V}_x])^\perp. \tag{72}
$$

Because (70) holds, we obtain that $\langle \mathrm{D}\phi_g(x)[\xi_1], \mathrm{D}\phi_g(x)[\xi_2] \rangle_{\phi_g(x)} = \langle \xi_1, \xi_2 \rangle_x = 0$ for any $\xi_1 \in \mathcal{V}_x$ and $\xi_2 \in \mathcal{H}_x$. This means that $\mathrm{D}\phi_g(x)[\xi_2] \in (\mathrm{D}\phi_g(x)[\mathcal{V}_x])^\perp$, which yields $\mathrm{D}\phi_g(x)[\mathcal{H}_x] \subset (\mathrm{D}\phi_g(x)[\mathcal{V}_x])^\perp$. Considering the dimension of both sides, we have that

$$
\mathrm{D}\phi_g(x)[\mathcal{H}_x] = (\mathrm{D}\phi_g(x)[\mathcal{V}_x])^\perp. \tag{73}
$$

It follows from (72) and (73) that (71) holds. ∎

Theorem 4 yields the following corollary.

Corollary 1: Suppose that the group action ϕ_g is an isometry in terms of Riemannian metric $\langle \cdot, \cdot \rangle$; i.e., (70) holds for any $g \in \mathcal{G}$ and any $\xi_1, \xi_2 \in T_x\mathcal{M}$. Then

$$
\bar{\xi}_{\phi_g(x)} = \mathrm{D}\phi_g(x)[\bar{\xi}_x] \tag{74}
$$

where $\bar{\xi}_x$ and $\bar{\xi}_{\phi_g(x)}$ are the horizontal lifts of $\xi \in T_{[x]}(\mathcal{M}/\mathcal{G})$ at $x \in \mathcal{M}$ and $\phi_g(x) \in \mathcal{M}$, respectively.

Proof: Because $\pi \circ \phi_g = \pi$

$$
\mathrm{D}(\pi \circ \phi_g)(x)[\bar{\xi}_x] = \mathrm{D}\pi(x)[\bar{\xi}_x] = \xi \tag{75}
$$

where the second equality follows from the definition of the horizontal lift. Moreover, by the chain rule

$$
\mathrm{D}(\pi \circ \phi_g)(x)[\bar{\xi}_x] = \mathrm{D}\pi(\phi_g(x))[\mathrm{D}\phi_g(x)[\bar{\xi}_x]]. \tag{76}
$$

It follows from (75) and (76) that $\mathrm{D}\pi(\phi_g(x))[\mathrm{D}\phi_g(x)[\bar{\xi}_x]] = \xi$, and Theorem 4 yields $\mathrm{D}\phi_g(x)[\bar{\xi}_x] \in \mathcal{H}_{\phi_g(x)}$. By the definition of the horizontal lift, we obtain (74). ∎

Theorem 1 follows from Corollary 1. This is because the group action $\phi_U(\Theta) := U \circ \Theta$ is an isometry, as shown in (17).

D. Proof of Theorem 2

Because $T_\Theta N = \mathcal{V}_\Theta \oplus \mathcal{H}_\Theta$, η can be uniquely decomposed into

$$
\eta = \eta^v + \eta^h, \quad \eta^v \in \mathcal{V}_\Theta, \ \eta^h \in \mathcal{H}_\Theta.
$$

Since $\eta^v \in \mathcal{V}_\Theta$, there exists $X \in \mathrm{Skew}(n)$ such that

$$
\eta^v = (-XA + AX, -XB, CX).
$$

Thus, η^h can be described as

$$
\eta^h = (a + XA - AX, b + XB, c - CX).
$$

Because $\eta^h \in \mathcal{H}_\Theta$, we obtain that

$$
\begin{aligned}
\mathrm{sk}(2(a + XA - AX)A^{-1} + B(b + XB)^\top \\
+ C^\top(c - CX)) = 0.
\end{aligned}
$$

It follows from this equation that (33) holds, because $a^\top = a$ and $X^\top = -X$.

E. Proof of Theorem 3

Using the Kronecker product and vec-operator, the operators $\mathcal{L}_0$ and $\mathcal{L}_1$ have the matrix representations $K_0 = A^{-1} \otimes A + A \otimes A^{-1} - 2I_{n^2}$ and $K_1 = I_n \otimes (BB^\top + C^\top C) + (BB^\top + C^\top C) \otimes I_n$, respectively, where $\otimes$ denotes the Kronecker product. Both are symmetric, and K_1 is positive semidefinite [51]. Thus, $\mathcal{L}_1 \geq 0$. Note also that both summands of K_1 and, thus, of $\mathcal{L}_1$ are positive semidefinite, whence

$$
\mathcal{L}_1(X) = 0 \Rightarrow (BB^\top + C^\top C)X = 0. \tag{77}
$$

If $\lambda_j, \lambda_k \in \lambda(A)$ with corresponding orthonormal eigenvectors v_j, v_k, then

$$
\mathcal{L}_0(v_j v_k^\top) = \mu_{jk} v_j v_k^\top
$$

where $\mu_{jk} := \frac{(\lambda_j - \lambda_k)^2}{\lambda_j \lambda_k}$. From the n orthonormal eigenvectors v_j, $j = 1, 2, \ldots, n$, of the matrix A, we thus obtain n^2 orthonormal eigenvectors $v_j v_k^\top$, $j, k = 1, 2, \ldots, n$, of the linear matrix mapping. Because $\mu_{jk} \geq 0$ for all j, k, it follows that $\mathcal{L}_0 \geq 0$. Together with $\mathcal{L}_1 \geq 0$, this implies that (see [52, Fact 8.7.3])

$$\operatorname{Ker} \mathcal{L} = \operatorname{Ker} \mathcal{L}_1 \cap \operatorname{Ker} \mathcal{L}_0. \tag{78}$$

Moreover, the kernel of $\mathcal{L}_0$ is spanned by the matrices $v_j v_k^\top + v_k v_j^\top$ and $v_j v_k^\top - v_k v_j^\top$ with $\lambda_j = \lambda_k$, $j, k = 1, 2, \ldots, n$. That is

$$\operatorname{Ker} \mathcal{L}_0 \cap \operatorname{Skew}(n) = \operatorname{span}\{ v_j v_k^\top - v_k v_j^\top \mid \lambda_j = \lambda_k \}.$$

The matrix A can be expressed as

$$A = V \operatorname{diag}(\lambda_{n_1} I_{n_1}, \lambda_{n_2} I_{n_2}, \ldots, \lambda_{n_l} I_{n_l}) V^\top$$

where $n_1 + n_2 + \cdots + n_l = n$, $\lambda_1 = \cdots = \lambda_{n_1} < \lambda_{n_1+1} = \cdots = \lambda_{n_2} < \cdots < \lambda_{n_1+n_2+\cdots+n_{l-1}+1} = \cdots = \lambda_{n_l}$, and after suitable ordering and partitioning, $V = (V_1 \ \cdots \ V_l) = (v_1 \ \cdots \ v_n)$ is orthogonal to $\operatorname{Im} V_j = \operatorname{Ker}(\lambda_{n_j} I_n - A)$. We, thus, obtain that

$$\operatorname{Ker} \mathcal{L}_0 \cap \operatorname{Skew}(n)$$
$$= \{ V \operatorname{diag}(S_1, S_2, \ldots, S_l) V^\top \mid S_j \in \operatorname{Skew}(n_j) \}.$$

To see this, note that the right-hand side is the linear subspace of $\operatorname{Skew}(n)$, spanned by

$$v_j v_k^\top - v_k v_j^\top = V(e_j e_k^\top - e_k e_j^\top) V^\top$$

where $\lambda_j = \lambda_k$ and e_j is the jth unit vector in $\mathbf{R}^n$. Thus, it follows from (77) and (78) that $U \in \operatorname{Ker} \mathcal{L} \cap \operatorname{Skew}(n)$ implies $U = V \operatorname{diag}(S_1, S_2, \ldots, S_l) V^\top$ with $(BB^\top + C^\top C)U = 0$. In particular, we have that

$$\begin{cases} 0 = B^\top U V_j = B^\top V_j S_j \\ 0 = CUV_j = CV_j S_j \end{cases}$$

for $j = 1, 2, \ldots, l$, and thus

$$\operatorname{Ker}(\lambda_{n_j} I_n - A) \cap \operatorname{Ker} B^\top \cap \operatorname{Ker} C \supset \operatorname{Im}(V_j S_j).$$

Therefore

$$\dim(\operatorname{Ker}(\lambda_{n_j} I_n - A) \cap \operatorname{Ker} B^\top \cap \operatorname{Ker} C) \geq \operatorname{rank} S_j. \tag{79}$$

Because each $S_j \in \operatorname{Skew}(n_j)$ necessarily has even rank, assumption (34) and (79) yield that $\operatorname{rank} S_j = 0$ for $j = 1, 2, \ldots, l$, whence $U = 0$. This implies that

$$\operatorname{Ker} \mathcal{L} \cap \operatorname{Skew}(n) = \{0\} \tag{80}$$

or equivalently $\operatorname{Ker} \mathcal{L} \subset \operatorname{Sym}(n)$. Equation (80) implies that $\mathcal{L} : \operatorname{Skew}(n) \to \operatorname{Skew}(n)$ is an automorphism.

REFERENCES

[1] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, "Electrical networks and algebraic graph theory: Models, properties, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 977–1005, May 2018.

[2] L. Vandenberghe, S. Boyd, and A. El Gamal, "Optimal wire and transistor sizing for circuits with non-tree topology," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design*, 1997, pp. 252–259.

[3] J. C. Willems, "Realization of systems with internal passivity and symmetry constraints," *J. Franklin Inst.*, vol. 301, no. 6, pp. 605–621, 1976.

[4] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Princeton, NJ, USA: Princeton Univ. Press, 2010.

[5] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.

[6] T. Hatanaka, X. Zhang, W. Shi, M. Zhu, and N. Li, "Physics-integrated hierarchical/distributed HVAC optimization for multiple buildings with robustness against time delays," in *Proc. IEEE 56th Annu. Conf. Decision Control*, 2017, pp. 6573–6579.

[7] C. Lidström and A. Rantzer, "Optimal H^∞ state feedback for systems with symmetric and Hurwitz state matrix," in *Proc. Am. Control Conf.*, 2016, pp. 3366–3371.

[8] B. Hanzon and R. L. M. Peeters, "On the Riemannian interpretation of the Gauss–Newton algorithm," in *Mutual Impact of Computing Power and Control Theory*. New York, NY, USA: Springer, 1993, pp. 111–121.

[9] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.

[10] T. Mckelvey, A. Helmersson, and T. Ribarits, "Data driven local coordinates for multivariable linear systems and their application to system identification," *Automatica*, vol. 40, no. 9, pp. 1629–1635, 2004.

[11] R. Peeters, *System Identification Based on Riemannian Geometry: Theory and Algorithms* Tinbergen Institute Reserach Series, vol. 64. Amsterdam, Netherlands: Thesis Publishers, 1994.

[12] H. Sato and K. Sato, "Riemannian optimal system identification algorithm for linear MIMO systems," *IEEE Control Syst. Lett.*, vol. 1, no. 2, pp. 376–381, Oct. 2017.

[13] A. Wills and B. Ninness, "On gradient-based search for multivariable system estimates," *IEEE Trans. Autom. Control*, vol. 53, no. 1, pp. 298–306, Feb. 2008.

[14] T. Katayama, *Subspace Methods for System Identification*. New York, NY, USA: Springer, 2006.

[15] W. E. Larimore, "Canonical variate analysis in identification, filtering, and adaptive control," in *Proc. 29th IEEE Conf. Decision Control*, 1990, pp. 596–604.

[16] S. J. Qin, "An overview of subspace identification," *Comput. Chem. Eng.*, vol. 30, no. 10, pp. 1502–1513, 2006.

[17] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.

[18] P. Van Overschee and B. De Moor, "A unifying theorem for three subspace system identification algorithms," *Automatica*, vol. 31, no. 12, pp. 1853–1864, 1995.

[19] M. Verhaegen and P. Dewilde, "Subspace model identification part 1. The output-error state-space model identification class of algorithms," *Int. J. Control*, vol. 56, no. 5, pp. 1187–1210, 1992.

[20] M. Verhaegen and A. Hansson, "N2SID: Nuclear norm subspace identification of innovation models," *Automatica*, vol. 72, pp. 57–63, 2016.

[21] M. C. Campi, T. Sugie, and F. Sakai, "An iterative identification method for linear continuous-time systems," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1661–1669, Aug. 2008.

[22] H. Garnier, M. Gilson, T. Bastogne, and M. Mensler, "The contsid toolbox: A software support for data-based continuous-time modelling," in *Identification of Continuous-Time Models from Sampled Data*. New York, NY, USA: Springer, 2008, pp. 249–290.

[23] R. Johansson, M. Verhaegen, and C. T. Chou, "Stochastic theory of continuous-time state-space identification," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 41–51, Jan. 1999.

[24] I. Maruta and T. Sugie, "Projection-based identification algorithm for grey-box continuous-time models," *Syst. Control Lett.*, vol. 62, no. 11, pp. 1090–1097, 2013.

[25] A. Ohsumi, K. Kameyama, and K. Yamaguchi, "Subspace identification for continuous-time stochastic systems via distribution-based approach," *Automatica*, vol. 38, no. 1, pp. 63–79, 2002.

[26] P. Van Overschee and B. De Moor, "Continuous-time frequency domain subspace system identification," *Signal Process.*, vol. 52, no. 2, pp. 179–194, 1996.

[27] S. Lang, *Fundamentals of Differential Geometry*. New York, NY, USA: Springer, 2012.

[28] K. Sato and H. Sato, "Structure-preserving H^2 optimal model reduction based on Riemannian trust-region method," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 505–511, Feb. 2018.

[29] J. M. Lee, *Introduction to Smooth Manifolds*. New York, NY, USA: Springer, 2013.

[30] B. Afsari and R. Vidal, "Bundle reduction and the alignment distance on spaces of state-space LTI systems," *IEEE Trans. Autom. Control*, vol. 62, no. 8, pp. 3804–3819, Aug. 2017.

[31] R. Brockett, "Some geometric questions in the theory of linear systems," *IEEE Trans. Autom. Control*, vol. AC-21, no. 4, pp. 449–455, Aug. 1976.

[32] D. F. Delchamps, "Global structure of families of multivariable linear systems with an application to identification," *Math. Syst. Theory*, vol. 18, no. 1, pp. 329–380, 1985.

[33] B. Hanzon and R. J. Ober, "Overlapping block-balanced canonical forms for various classes of linear systems," *Linear Algebra Appl.*, vol. 281, no. 1-3, pp. 171–225, 1998.

[34] M. Hazewinkel, "Moduli and canonical forms for linear dynamical systems II: The topological case," *Math. Syst. Theory*, vol. 10, no. 1, pp. 363–385, 1977.

[35] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

[36] E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. New York, NY, USA: Springer, 1998.

[37] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*. New York, NY, USA: Springer, 1993.

[38] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[39] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61–74, 1994.

[40] K. Tan and K. M. Grigoriadis, "Stabilization and H^∞ control of symmetric systems: An explicit solution," *Syst. Control Lett.*, vol. 44, no. 1, pp. 57–72, 2001.

[41] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[42] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*. London, U.K.: Springer-Verlag, 1994.

[43] S. Sra and R. Hosseini, "Conic geometric optimization on the manifold of positive definite matrices," *SIAM J. Optim.*, vol. 25, no. 1, pp. 713–739, 2015.

[44] H. Sato, "A Dai–Yuan-type Riemannian conjugate gradient method with the weak Wolfe conditions," *Comput. Optim. Appl.*, vol. 64, no. 1, pp. 101–118, 2016.

[45] W. Ring and B. Wirth, "Optimization methods on Riemannian manifolds and their application to shape space," *SIAM J. Optim.*, vol. 22, no. 2, pp. 596–627, 2012.

[46] A. J. M. Van Overbeek and L. Ljung, "On-line structure selection for multivariable state-space models," *Automatica*, vol. 18, no. 5, pp. 529–543, 1982.

[47] S. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.

[48] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[49] L. Ljung, "Experiments with identification of continuous time models," *IFAC Proc. Volumes*, vol. 42, no. 10, pp. 1175–1180, 2009.

[50] S. T. Smith, "Covariance, subspace, and intrinsic Crameŕ-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, May 2005.

[51] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*. Philadelphia, PA, USA: SIAM, 2005.

[52] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

Kazuhiro Sato (M'17) received the B.S., M.S., and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 2009, 2011, and 2014, respectively.

He was a Postdoctoral Fellow with Kyoto University from 2014 to 2017. From 2017 to 2019, he was an Assistant Professor with Kitami Institute of Technology. He is currently a Lecturer with the Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan. His research interests include mathematical control theory and optimization.

Dr. Sato is a member of Society of Instrument and Control Engineers of Japan.

Hiroyuki Sato (M'13) received the B.S. degree in engineering from Kyoto University, Kyoto, Japan, in 2009 and the M.S. and Ph.D. degrees in informatics from Kyoto University in 2011 and 2013, respectively.

He is currently a Program-Specific Associate Professor with the Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University. His research interest includes Riemannian optimization theory including its applications to other fields such as control engineering.

Tobias Damm received the Diploma in mathematics from the University of Würzburg, Würzburg, Germany, in 1996 and the Doctoral degree from the University of Bremen, Bremen, Germany, in 2002.

From 2002 to 2006, he held Postdoctoral Positions with Technical University of München and Technical University of Braunschweig. He was appointed as a Junior Professor with the TU Kaiserslautern in 2006 and Associate Professor with the University of Bayreuth in 2011. Since 2012, he has been a Professor for systems and control theory with the Technical University of Kaiserslautern, Kaiserslautern, Germany. His research interests encompass stochastic control systems, matrix equations, and numerical methods.