

Stable Linear System Identification with Prior Knowledge by Riemannian Sequential Quadratic Optimization

Mitsuaki Obara, Kazuhiro Sato, *Member, IEEE*, Hiroki Sakamoto, Takayuki Okuno, and Akiko Takeda

Abstract—We consider an identification method for a linear continuous time-invariant autonomous system from noisy state observations. In particular, we focus on the identification to satisfy the asymptotic stability of the system with some prior knowledge. To this end, we propose to model this identification problem as a Riemannian nonlinear optimization (RNLO) problem, where the stability is ensured through a certain Riemannian manifold and the prior knowledge is expressed as nonlinear constraints defined on this manifold. To solve this RNLO, we apply the Riemannian sequential quadratic optimization (RSQO) that was proposed by Obara, Okuno, and Takeda (2022) most recently. RSQO performs quite well with theoretical guarantee to find a point satisfying the Karush-Kuhn-Tucker conditions of RNLO. In this paper, we demonstrate that the identification problem can be indeed solved by RSQO more effectively than competing algorithms.

Index Terms—Riemannian sequential quadratic optimization, Riemannian nonlinear optimization, Stable linear system, System identification.

I. INTRODUCTION

SYSTEM identification of a linear continuous time-invariant autonomous system

$$\dot{x}(t) = Ax(t) \quad (1)$$

with the state vector $x(t) \in \mathbb{R}^n$ is a task to estimate $A \in \mathbb{R}^{n \times n}$ from measured state data and is one of the most important

Submitted on December 28, 2021 and revised on September 15, 2023. This work was supported in part by the Japan Society for the Promotion of Science KAKENHI under Grant 23K03899, 22KJ0563, 21J20493, 20K14760, 20K19748 and 19H04069, and was part of the results of Value Exchange Engineering, a joint research project between Mercari, Inc. and the RIISE.

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

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. Sakamoto is with the Department of Mathematical Informatics, Graduate School of Information Science and Technology, the University of Tokyo, Tokyo, 113-8656 Japan (e-mail: soccer-books0329@g.ecc.u-tokyo.ac.jp)

T. Okuno is with the Department of Science and Technology, Faculty of Science and Technology, Seikei University, Tokyo, 180-8633, Japan (e-mail: takayuki-okuno@st.seikei.ac.jp)

A. Takeda is with the Department of Creative Informatics, Graduate School of Information Science and Technology, the University of Tokyo, Tokyo, 113-8656 Japan, and the Generic Technology Research Group, Center for Advanced Intelligence Project, RIKEN, Tokyo, 103-0037 Japan (e-mail: takeda@mist.i.u-tokyo.ac.jp, akiko.takeda@riken.jp)

topics to predict the future state of a real system. System identification has been investigated for many years in several settings. For example, prediction error methods [1, 2, 3, 4, 5] and subspace identification methods [6, 7, 8, 9, 10] are conventional identification methods for linear time-invariant state space systems. Dynamic mode decomposition (DMD) deals with the system of the particular form (1), which can be possibly large-scale one such as fluid flows [11, 12].

In real applications, it is often crucial to identify a system satisfying the asymptotic stability. Yet, it is completely non-trivial to ensure the stability in the identification methods such as those above; for example, DMD spectrum has been shown to be considerably sensitive to measurement noise and hence its resultant system may be unstable [13]. Several approaches have been investigated to overcome the difficulty: Subspace identification methods with guaranteed stability have been proposed in [14, 15, 16]. In [17], a constraint generation approach has been proposed. A Lagrange relaxation is used to ensure several types of stability including the asymptotic stability for linear time-invariant state space models [18] as well as a different stability called the global incremental ℓ^2 stability for nonlinear systems [19, 20]. It is worth mentioning that, among the above identification methods, those presented in [14, 15, 16, 17, 18, 19, 20] are based on convex optimization.

In addition to the stability, another important characteristic in the identification is prior knowledge. As described in [1, Chapter 16], identification should reflect the prior knowledge peculiar to the system, such as the nonnegativity of all or partial components of the system A . Nevertheless, the conventional identification methods are not able to generate a stable matrix A with prior knowledge information. Indeed, to impose the knowledge, it is necessary to deal with their nonconvexity, which requires additional techniques from non-convex optimization.

In this paper, we propose a new method for identifying a system, from noisy data, that satisfies both asymptotic stability and prior knowledge. This method first discretizes the system, and then applies the prediction error method, which is formulated as solving a least squares optimization problem. However, to identify the desired system, we must incorporate both stability and prior knowledge into this optimization problem. Motivated by the fact that an asymptotically stable linear system can be expressed in a port-Hamiltonian form [21, Proposition 1], we handle the stability of the system through a Riemannian manifold as in [22]. Meanwhile, we deal with the

prior knowledge information as nonlinear constraints defined on the manifold. After all, our identification method reduces to solving a constrained nonlinear optimization problem on a Riemannian manifold, called RNLO for short.

RNLO is one of the general optimization classes proposed in [23] and researches for algorithms to solve RNLO are in progress, for example, [24, 25]. Remarkably, the modeling enables us to identify a stable A even when we use noisy observations, which usually spoil the stability in system identification. The modeling also allows us to take various prior knowledge into account by virtue of the generality of RNLO. To solve RNLO, we propose to apply the Riemannian sequential quadratic optimization method (RSQO) that was presented recently in [25]. Strength of RSQO is the global convergence property to Karush-Kuhn-Tucker (KKT) points of RNLO and locally fast convergence speed.

Our contributions are summarized as follows:

- 1) We propose a prediction error method to identify a linear system satisfying the asymptotic stability with prior knowledge from noisy state observations. The key ingredients of this method are formulating the identification problem as RNLO and solving it with RSQO.
- 2) We conduct numerical experiments with comparisons to demonstrate that RNLO modeling and RSQO are very promising.

Organization of paper

The rest of this paper is organized as follows. In the following subsection, we introduce notations and terminologies. In Section III, we formulate the system identification problem as RNLO. In Section IV, we introduce RSQO of the form tailored to the obtained RNLO and moreover show its theoretical properties. We also exploit the geometry of the problem. In Section V, we demonstrate that the proposed formulation and RSQO are promising through experimental comparisons. Finally, in Section VI, we conclude this paper with some remarks and discussions about future work.

Notations and terminologies

The sets of real numbers and complex ones are denoted by \mathbb{R} and \mathbb{C} , respectively. Let \mathbb{N} be the set of all the natural numbers, that is, $\mathbb{N} := \{1, 2, \dots\}$. We denote the identity matrix of size n by $I \in \mathbb{R}^{n \times n}$. Given a matrix $Y \in \mathbb{R}^{n \times n}$, Y^\top and $\text{tr}(Y)$ represent the transpose of Y and the trace of Y , respectively. Let $\|\cdot\|_F$ denote the Frobenius norm of a matrix, that is, $\|Y\|_F := \sqrt{\text{tr}(Y^\top Y)}$ for any $Y \in \mathbb{R}^{n \times n}$. Let $\|\cdot\|$ denote the Euclidean norm of a vector, that is, $\|v\| := \sqrt{v_1^2 + \dots + v_n^2}$ for any $v \in \mathbb{R}^n$. Define the i -th standard basis of \mathbb{R}^n by $e_i := (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^n$, where the i -th element is one and the others are zeros. Given a sufficiently smooth function $w: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$, the Euclidean gradient of w at $Y \in \mathbb{R}^{n \times n}$ is denoted by $\nabla w(Y)$. Let \mathcal{V} be any vector space and $T_W \mathcal{V}$ be a tangent space at $W \in \mathcal{V}$. We canonically identify $T_W \mathcal{V}$ with \mathcal{V} and use the symbol \simeq to denote the canonical identification. Under $T_W \mathcal{V} \simeq \mathcal{V}$, given a sufficiently smooth function $l: \mathcal{V} \rightarrow \mathbb{R}$, we define the

directional derivative of l at $W \in \mathcal{V}$ along $\xi_W \in T_W \mathcal{V}$ by $Dl(W)[\xi_W] := \lim_{t \downarrow 0} \frac{l(W+t\xi_W) - l(W)}{t} \in \mathbb{R}$.

II. PRELIMINARIES

To formulate our problem as RNLO, we characterize the stability of the system in Section II-B. To this end, we first summarize the Riemannian geometries of the sets of skew-symmetric matrices and symmetric positive definite matrices.

A. Riemannian geometries of skew-symmetric matrices and symmetric positive definite matrices

We define the vector space of the skew-symmetric matrices in $\mathbb{R}^{n \times n}$ by $\text{Skew}(n) := \{J \in \mathbb{R}^{n \times n} | J^\top = -J\}$. $\text{Skew}(n)$ can be regarded as a linear manifold, whose tangent space at $J \in \text{Skew}(n)$, denoted by $T_J \text{Skew}(n)$, is canonically identified with $\text{Skew}(n)$ itself. Under $T_J \text{Skew}(n) \simeq \text{Skew}(n)$, the Riemannian metric at J is defined as $\langle \xi_J, \eta_J \rangle_J := \text{tr}(\xi_J^\top \eta_J)$ for any $\xi_J, \eta_J \in \text{Skew}(n)$. Define an operator $\text{skew}: \mathbb{R}^{n \times n} \rightarrow \text{Skew}(n)$ by $\text{skew}(Y) := \frac{Y - Y^\top}{2}$. Let $l: \text{Skew}(n) \rightarrow \mathbb{R}$ be a twice continuously differentiable function and \bar{l} be the smooth extension of l to the Euclidean space $\mathbb{R}^{n \times n}$. The Riemannian gradient of l at J is

$$\text{grad } l(J) = \text{skew}(\nabla \bar{l}(J)) \in \text{Skew}(n), \quad (2)$$

where $\nabla \bar{l}(J)$ is the Euclidean gradient of \bar{l} at J . The exponential mapping at J is given by $\text{Exp}_J(\xi_J) = J + \xi_J$ for all $\xi_J \in \text{Skew}(n)$.

Let us denote the set of symmetric positive definite matrices in $\mathbb{R}^{n \times n}$ by $\text{Sym}_{++}(n)$, which is an open submanifold of $\mathbb{R}^{n \times n}$. The tangent space at $P \in \text{Sym}_{++}(n)$, denoted by $T_P \text{Sym}_{++}(n)$, is canonically identified with the set of symmetric matrices denoted by $\text{Sym}(n)$. Under $T_P \text{Sym}_{++}(n) \simeq \text{Sym}(n)$, we equip $\text{Sym}_{++}(n)$ with a Riemannian metric at $P \in \text{Sym}_{++}(n)$ defined by

$$\langle \xi_P, \eta_P \rangle_P := \text{tr}(P^{-1} \xi_P P^{-1} \eta_P) \quad (3)$$

for any $\xi_P, \eta_P \in \text{Sym}(n)$ [26]. Let us define an operator $\text{sym}: \mathbb{R}^{n \times n} \rightarrow \text{Sym}(n)$ by $\text{sym}(Y) := \frac{Y + Y^\top}{2}$. Let $m: \text{Sym}_{++}(n) \rightarrow \mathbb{R}$ be a twice continuously differentiable function and \bar{m} be the smooth extension of m to the Euclidean space $\mathbb{R}^{n \times n}$. The Riemannian gradient of m at P is

$$\text{grad } m(P) = P \text{sym}(\nabla \bar{m}(P)) P, \quad (4)$$

where $\nabla \bar{m}(P)$ is the Euclidean gradient of \bar{m} at P . On $\text{Sym}_{++}(n)$ equipped with (3), we introduce retraction [26] by $\text{Retr}_P(\xi_P) = P + \xi_P + \frac{1}{2} \xi_P P^{-1} \xi_P$ for all $\xi_P \in \text{Sym}(n)$. See [26, 27] for the details of the concepts and the notations on optimization on Riemannian manifolds and the geometries of the $\text{Skew}(n)$ and $\text{Sym}_{++}(n)$.

B. Characterizations of stability

Let us start with the definitions of the stability: $A \in \mathbb{R}^{n \times n}$ is stable if the real parts of all the eigenvalues of the matrix A are negative. We say that the system (1) is asymptotically stable if A is stable.

In the following proposition, a useful result for the stability is provided. The proof can be found in the literature, e.g., [22, Section III] and [21].

Proposition II.1. *A matrix $A \in \mathbb{R}^{n \times n}$ is stable if and only if there exists $(J, R, Q) \in \text{Skew}(n) \times \text{Sym}_{++}(n) \times \text{Sym}_{++}(n)$ such that*

$$A = (J - R)Q. \quad (5)$$

This proposition motivates us to consider optimization problems with respect to (J, R, Q) to ensure the stability of the system. Note that, for any stable matrix A , the triplet (J, R, Q) satisfying (5) is not unique.

Define $\mathcal{M} := \text{Skew}(n) \times \text{Sym}_{++}(n) \times \text{Sym}_{++}(n)$ and write $\theta := (J, R, Q) \in \mathcal{M}$. Clearly, \mathcal{M} is a product Riemannian manifold and its tangent space at θ is expressed as $T_\theta \mathcal{M} \simeq \text{Skew}(n) \times \text{Sym}(n) \times \text{Sym}(n)$, on which a Riemannian metric at θ is defined as $\langle \xi_\theta, \eta_\theta \rangle_\theta := \text{tr}(\xi_J^\top \eta_J) + \text{tr}(R^{-1} \xi_R R^{-1} \eta_R) + \text{tr}(Q^{-1} \xi_Q Q^{-1} \eta_Q)$ for any $\xi_\theta = (\xi_J, \xi_R, \xi_Q), \eta_\theta = (\eta_J, \eta_R, \eta_Q) \in T_\theta \mathcal{M}$. We use the retraction $\text{Retr}_\theta(\xi_\theta) = (J + \xi_J, R + \xi_R + \frac{1}{2} \xi_R R^{-1} \xi_R, Q + \xi_Q + \frac{1}{2} \xi_Q Q^{-1} \xi_Q)$ for all $\xi_\theta = (\xi_J, \xi_R, \xi_Q) \in T_\theta \mathcal{M}$. We refer readers to [27] for the geometry of a product manifold.

III. PROBLEM SETUP

In this section, we formulate optimization problems for the identification of a stable linear system with prior knowledge and derive optimality conditions.

A. Problem formulations

In this paper, we assume that a system to be identified is of the form (1) with stable A . Thus, from Proposition II.1, a discretized model with noise ε_k of the system by the Euler method is expressed by

$$x_{k+1} = (I + h(J - R)Q)x_k + \varepsilon_k, \quad (6)$$

where $h > 0$ is the sampling interval. That is, the noise ε_k can be interpreted as the prediction error at time k . From (6), we can define the least-square function

$$\begin{aligned} f(\theta) &:= \frac{1}{N} \sum_{k=0}^{N-1} \|\varepsilon_k\|^2 \\ &= \frac{1}{N} \|X' - (I + h(J - R)Q)X\|_{\mathbb{F}}^2, \end{aligned} \quad (7)$$

where $X := (x_0 \ x_1 \ \cdots \ x_{N-1}) \in \mathbb{R}^{n \times N}$ and $X' := (x_1 \ x_2 \ \cdots \ x_N) \in \mathbb{R}^{n \times N}$. Note that the system (6) is a nonlinear regression model, because there exists a multiplication of the parameters $J - R$ and Q .

Using the least-square function (7), we consider the following optimization problem:

$$\underset{\theta = (J, R, Q) \in \mathcal{M}}{\text{minimize}} \quad f(\theta) \quad (8a)$$

$$\text{subject to} \quad g_{ij}(\theta) \leq 0, \quad ((i, j) \in \mathcal{I}), \quad (8b)$$

$$h_{ij}(\theta) = 0, \quad ((i, j) \in \mathcal{E}). \quad (8c)$$

Here, $\mathcal{I}, \mathcal{E} \subseteq \{1, \dots, n\} \times \{1, \dots, n\}$ are the index sets of inequality and equality constraints, respectively. Moreover, $\{g_{ij}\}_{(i,j) \in \mathcal{I}}, \{h_{ij}\}_{(i,j) \in \mathcal{E}}: \mathcal{M} \rightarrow \mathbb{R}$ are continuously differentiable functions. The problem (8) is a Riemannian nonlinear optimization problem (RNLO) [23], that is, a constrained nonlinear optimization problem over a Riemannian manifold. It should be noted that this formulation is novel for a stable linear system identification with prior knowledge.

Any inequality and equality constraints are acceptable in (8) as long as they are continuously differentiable with respect to θ ; for example, we can deal with nonnegativity (or nonpositivity) of the elements $e_i^\top (J - R)Q e_j \stackrel{\leq}{\geq} 0$ and an element equality, $e_i^\top (J - R)Q e_j = c_{ij}$ for a given constant $c_{ij} \in \mathbb{R}$. In Section V, we consider box-type inequalities.

B. Optimality conditions

In what follows, we define the Karush-Kuhn-Tucker (KKT) conditions and relevant concepts for RNLO. We say that $\theta^\bullet \in \mathcal{M}$ satisfies the KKT conditions of RNLO (8) if there exist Lagrange multipliers $\{\mu_{ij}^\bullet\}_{(i,j) \in \mathcal{I}} \subseteq \mathbb{R}$ and $\{\lambda_{ij}^\bullet\}_{(i,j) \in \mathcal{E}} \subseteq \mathbb{R}$ such that the following hold:

$$\begin{aligned} &\text{grad } f(\theta^\bullet) \\ &\quad + \sum_{(i,j) \in \mathcal{I}} \mu_{ij}^\bullet \text{grad } g_{ij}(\theta) + \sum_{(i,j) \in \mathcal{E}} \lambda_{ij}^\bullet \text{grad } h_{ij}(\theta) = 0, \\ &\mu_{ij}^\bullet \geq 0, g_{ij}(\theta^\bullet) \leq 0, \mu_{ij}^\bullet g_{ij}(\theta^\bullet) = 0, \quad ((i, j) \in \mathcal{I}), \\ &h_{ij}(\theta^\bullet) = 0, \quad ((i, j) \in \mathcal{E}), \end{aligned} \quad (9)$$

where the operator grad denotes the Riemannian gradient as in Section I. We call θ^\bullet a KKT point of (8). For brevity, we often write $\mu^\bullet \in \mathbb{R}^{|\mathcal{I}|}$ and $\lambda^\bullet \in \mathbb{R}^{|\mathcal{E}|}$ for $\{\mu_{ij}^\bullet\}_{(i,j) \in \mathcal{I}}$ and $\{\lambda_{ij}^\bullet\}_{(i,j) \in \mathcal{E}}$, respectively. It is known that, under some conditions called constraint qualifications, the KKT conditions are necessary ones for the optimality; that is, a local minimizer satisfies the KKT conditions under constraint qualifications [23, 28].

IV. OPTIMIZATION METHODS

In this section, we introduce a specific method for solving (8), Riemannian sequential quadratic optimization (RSQO) that was originally presented in [25].

A. Geometry of problems

In the following theorem, we derive the Riemannian gradient of f specifically.

Theorem IV.1. *Given $\theta = (J, R, Q) \in \mathcal{M}$, the Riemannian gradient of f in (8) is*

$$\begin{aligned} &\text{grad } f(\theta) \\ &= (\text{skew}(\nabla_J \bar{f}(\theta)), R \text{sym}(\nabla_R \bar{f}(\theta))R, Q \text{sym}(\nabla_Q \bar{f}(\theta))Q), \end{aligned} \quad (10)$$

where \bar{f} is the smooth extension of f to the Euclidean space and

$$\nabla_J \bar{f}(\theta) = -\frac{2h}{N}(X' - (I + h(J - R)Q)X)X^\top Q^\top, \quad (11)$$

$$\nabla_R \bar{f}(\theta) = -\nabla_J \bar{f}(\theta), \quad (12)$$

$$\nabla_Q \bar{f}(\theta) = -\frac{2h}{N}(J^\top - R^\top)(X' - (I + h(J - R)Q)X)X^\top \quad (13)$$

are the Euclidean gradients on J, R , and Q , respectively.

Proof. Note that the Riemannian gradient of the function on the product manifold is organized component-wisely [27]. Using this fact, we derive the Riemannian gradient of our problems. Hereafter, we abbreviate $\mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n}$ as $\mathbb{R}^{n \times n \otimes 3}$ for brevity.

Let \bar{f} be the smooth extension of f to the Euclidean space. We define $\bar{s}_k: \mathbb{R}^{n \times n \otimes 3} \rightarrow \mathbb{R}^n$ by

$$\bar{s}_k(\bar{\theta}) := x_{k+1} - Zx_k, \quad (14)$$

where $\bar{\theta} := (\bar{J}, \bar{R}, \bar{Q}) \in \mathbb{R}^{n \times n \otimes 3}$ and $Z := I + h(\bar{J} - \bar{R})\bar{Q}$. By substituting (14) into (7), we have $\bar{f}(\bar{\theta}) = \frac{1}{N} \sum_{k=1}^{N-1} \bar{s}_k(\bar{\theta})^\top \bar{s}_k(\bar{\theta})$ and its directional derivative at $\bar{\theta} \in \mathbb{R}^{n \times n \otimes 3}$ along $\bar{\xi}_{\bar{\theta}} = (\bar{\xi}_{\bar{J}}, \bar{\xi}_{\bar{R}}, \bar{\xi}_{\bar{Q}}) \in \mathbb{R}^{n \times n \otimes 3}$ is

$$D\bar{f}(\bar{\theta})[\bar{\xi}_{\bar{\theta}}] = \frac{2}{N}(D\bar{s}_k(\bar{\theta})[\bar{\xi}_{\bar{\theta}}])^\top \bar{s}_k(\bar{\theta}). \quad (15)$$

Now, we derive the explicit form of $D\bar{s}_k(\bar{\theta})[\bar{\xi}_{\bar{\theta}}]$ in (15). To this end, we first calculate the directional derivative along $(\bar{\xi}_{\bar{J}}, 0, 0)$ as $D\bar{s}_k(\bar{\theta})[(\bar{\xi}_{\bar{J}}, 0, 0)] = -h\bar{\xi}_{\bar{J}}\bar{Q}x_k$. Similarly, we have $D\bar{s}_k(\bar{\theta})[(0, \bar{\xi}_{\bar{R}}, 0)] = h\bar{\xi}_{\bar{R}}\bar{Q}x_k$ and $D\bar{s}_k(\bar{\theta})[(0, 0, \bar{\xi}_{\bar{Q}})] = -h(\bar{J} - \bar{R})\bar{\xi}_{\bar{Q}}x_k$. Thus, the explicit form of $D\bar{s}_k(\bar{\theta})[\bar{\xi}_{\bar{\theta}}]$ is

$$D\bar{s}_k(\bar{\theta})[\bar{\xi}_{\bar{\theta}}] = -h((\bar{\xi}_{\bar{J}} - \bar{\xi}_{\bar{R}})\bar{Q} + (\bar{J} - \bar{R})\bar{\xi}_{\bar{Q}})x_k. \quad (16)$$

Using the result, we derive the Riemannian gradient of f . By substituting (14) and (16) into (15), we have

$$\begin{aligned} D\bar{f}(\bar{\theta})[\bar{\xi}_{\bar{\theta}}] &= \text{tr} \left(\bar{\xi}_{\bar{J}}^\top \sum_{k=1}^{N-1} -\frac{2h}{N}(x_{k+1} - Zx_k)x_k^\top \bar{Q}^\top \right) \\ &+ \text{tr} \left(\bar{\xi}_{\bar{R}}^\top \sum_{k=1}^{N-1} \frac{2h}{N}(x_{k+1} - Zx_k)x_k^\top \bar{Q}^\top \right) \\ &+ \text{tr} \left(\bar{\xi}_{\bar{Q}}^\top \sum_{k=1}^{N-1} -\frac{2h}{N}(\bar{J}^\top - \bar{R}^\top)(x_{k+1} - Zx_k)x_k^\top \right). \end{aligned} \quad (17)$$

Here, it follows from $x_{k+1} = X'e_k$ and $x_k = Xe_k$ that

$$\begin{aligned} \sum_{k=1}^{N-1} (x_{k+1} - Zx_k)x_k^\top &= (X' - ZX) \left(\sum_{k=1}^{N-1} e_k e_k^\top \right) X^\top \\ &= (X' - ZX)X^\top, \end{aligned} \quad (18)$$

where the second equality holds from $\sum_{k=1}^{N-1} e_k e_k^\top = I$. Thus, since $\mathcal{M} \subseteq \mathbb{R}^{n \times n \otimes 3}$ holds, combining (17) and (18) provides (11), (12), and (13) for any $\theta \in \mathcal{M}$. By projecting (11), (12), and (13) onto $T_\theta \mathcal{M}$ according to (2) and (4), we obtain the Riemannian gradient of the form (10). \square

In a similar manner, we can derive the Riemannian gradient of constraints although we focus on that of the objective function here.

B. Riemannian sequential quadratic optimization

In this subsection, we briefly explain the RSQO for solving (8) using notations and terminologies we have set up so far. See [25, Section 3] for more details.

1) Description of RSQO: RSQO is an iterative method. Let $\theta^k \in \mathcal{M}$ be a current iterate. RSQO first solves the following quadratic subproblem to generate a search direction $\Delta\theta^{k\bullet} \in T_{\theta^k} \mathcal{M}$:

$$\begin{aligned} &\text{minimize} && \tilde{f}^k(\Delta\theta^k) \\ &\Delta\theta^k \in T_{\theta^k} \mathcal{M} \\ &\text{subject to} && \tilde{g}_{ij}^k(\Delta\theta^k) \leq 0, \quad ((i, j) \in \mathcal{I}), \\ &&& \tilde{h}_{ij}^k(\Delta\theta^k) = 0, \quad ((i, j) \in \mathcal{E}), \end{aligned} \quad (19)$$

where

$$\tilde{f}^k(\Delta\theta^k) := \frac{1}{2} \langle B^k[\Delta\theta^k], \Delta\theta^k \rangle_{\theta^k} + \langle \text{grad} f(\theta^k), \Delta\theta^k \rangle_{\theta^k}, \quad (20)$$

$$\tilde{g}_{ij}^k(\Delta\theta^k) := g_{ij}(\theta^k) + \langle \text{grad} g_{ij}(\theta^k), \Delta\theta^k \rangle_{\theta^k}, \quad (21)$$

$$\tilde{h}_{ij}^k(\Delta\theta^k) := h_{ij}(\theta^k) + \langle \text{grad} h_{ij}(\theta^k), \Delta\theta^k \rangle_{\theta^k}, \quad (22)$$

and $B^k: T_{\theta^k} \mathcal{M} \rightarrow T_{\theta^k} \mathcal{M}$ is a symmetric positive-definite linear operator. Since the problem can be expressed as a Euclidean convex quadratic optimization problem, it can be solved by, for example, interior-point methods or active set methods. We denote the Lagrange multipliers at the solution $\Delta\theta^{k\bullet}$ corresponding to the inequality and equality constraints by $\mu^{k\bullet} \in \mathbb{R}^{|\mathcal{I}|}$ and $\lambda^{k\bullet} \in \mathbb{R}^{|\mathcal{E}|}$, respectively.

Next, RSQO determines the step length by using the ℓ_1 penalty function defined as

$$P_\rho(\theta) := f(\theta) + \rho \left(\sum_{(i,j) \in \mathcal{I}} \max\{0, g_{ij}(\theta)\} + \sum_{(i,j) \in \mathcal{E}} |h_{ij}(\theta)| \right),$$

where $\rho > 0$ is a penalty parameter [24]. RSQO first sets

$$\rho^k = \begin{cases} \rho^{k-1}, & \text{if } \rho^{k-1} \geq v^k, \\ v^k + \varepsilon, & \text{otherwise,} \end{cases} \quad (23)$$

where $v^k := \max\{\max_{(i,j) \in \mathcal{I}} \mu_{ij}^{k\bullet}, \max_{(i,j) \in \mathcal{E}} |\lambda_{ij}^{k\bullet}|\}$ and $\varepsilon > 0$ is a prescribed algorithmic parameter. Then, by using P_{ρ^k} as a merit function, we find the smallest nonnegative integer t satisfying

$$\begin{aligned} &\gamma\beta^t \langle B^k[\Delta\theta^{k\bullet}], \Delta\theta^{k\bullet} \rangle_{\theta^k} \\ &\leq P_{\rho^k}(\theta^k) - P_{\rho^k}(\text{Retr}_{\theta^k}(\beta^t \Delta\theta^{k\bullet})) \end{aligned} \quad (24)$$

and set $\alpha^k = \beta^t$. Using the search direction and the step length, RSQO updates the iterate. The above procedure is formalized as in Algorithm 1.

Algorithm 1: Riemannian sequential quadratic optimization (RSQO)

- 1 **Input** : Initial point $\theta^0 \in \mathcal{M}$; initial linear operator $B^0: T_{\theta^0}\mathcal{M} \rightarrow T_{\theta^0}\mathcal{M}$; hyperparameters $\rho^{-1} > 0, \varepsilon > 0, \beta \in (0, 1), \gamma \in (0, 1)$;
 - 2 **for** $k = 0, 1, \dots$ **do**
 - 3 Compute $\Delta\theta^{k\bullet}$ with associated Lagrange multipliers $\mu_{ij}^{k\bullet}$ and $\lambda_{ij}^{k\bullet}$ by solving (19);
 - 4 Update ρ^k according to (23);
 - 5 Determine the integer t according to the line search rule (24) and set $\alpha^k = \beta^t$;
 - 6 Update $\theta^{k+1} = \text{Retr}_{\theta^k}(\alpha^k \Delta\theta^{k\bullet})$, $\mu_{ij}^{k+1} = \mu_{ij}^{k\bullet}$ for $(i, j) \in \mathcal{I}$, and $\lambda_{ij}^{k+1} = \lambda_{ij}^{k\bullet}$ for $(i, j) \in \mathcal{E}$.
 - 7 **end for**
-

2) *Convergence properties:* First, the line search terminates within finitely many trials in RSQO, because the search direction is a descent direction for $P_{\rho^k} \circ \text{Retr}_{\theta^k}$. For details, see [25, Proposition 3.9, Remark 3.10]. Next, we show that RSQO has the global convergence property under the following assumptions, standard in nonlinear optimization.

Assumption 1. *There exist $m, M > 0$ such that, for any k ,*

$$m \|\Delta\theta^k\|_{\theta^k}^2 \leq \langle B^k[\Delta\theta^k], \Delta\theta^k \rangle_{\theta^k} \leq M \|\Delta\theta^k\|_{\theta^k}^2$$

for all $\Delta\theta^k \in T_{\theta^k}\mathcal{M}$.

Assumption 2. *The subproblem (19) is feasible at each iteration.*

Assumption 3. *The generated sequence $\left\{ \left(\theta^k, \mu^k, \lambda^k \right) \right\}_k$ is bounded. Here, the boundedness of $\{\theta^k\}$ is with respect to the distance induced from the Riemannian metric, while those of $\{\mu^k\}$ and $\{\lambda^k\}$ are in the sense of the Euclidean distance.*

Assumptions 2 and 3 are not easy to check in prior. But, we observed that they were not violated in the numerical experiments. By Assumption 3, any accumulation point of $\{\theta^k\}$ stays at $\text{Sym}_{++}(n)$.

Theorem IV.2. [25, Theorem 3.12] *Suppose Assumptions 1-3. Let $(\theta^\bullet, \mu^\bullet, \lambda^\bullet)$ be any accumulation point of $\left\{ \left(\theta^k, \mu^{k+1}, \lambda^{k+1} \right) \right\}$ generated by RSQO. Then, $(\theta^\bullet, \mu^\bullet, \lambda^\bullet)$ satisfies the KKT conditions of RNLO (8).*

V. NUMERICAL SIMULATIONS

In this section, we demonstrate the effectiveness of our RNLO modeling and RSQO. We conduct numerical experiments on a random system and make comparisons. In Section V-A, we introduce a problem setting as well as other modelings for the comparisons. In Section V-B, we describe the solver settings and an evaluation index for the experiments. In Section V-C, we show the numerical results and discuss the effect of choice of the modelings and the algorithms. All the experiments are implemented in Matlab_R2023a and Manopt [29] on a Windows 10 Pro with 2.60 GHz Core i9-11980HK CPU and 64.0 GB memory.

A. Problem setting: synthetic system

In this experiment, we consider the following synthetic system:

$$\theta = (J, R, Q) \in \mathcal{M} \quad \text{minimize} \quad f(\theta) = \frac{1}{N} \|X' - (I + h(J - R)Q)X\|_{\mathbb{F}}^2 \quad (25a)$$

$$\text{subject to} \quad l_{ij} \leq e_i^\top (J - R)Qe_j \leq r_{ij}, \quad ((i, j) \in \mathcal{I}_1 \cup \mathcal{I}_2), \quad (25b)$$

$$k_{ij}^2 \leq (e_i^\top (J - R)Qe_j - c_{ij})^2, \quad ((i, j) \in \mathcal{I}_2), \quad (25c)$$

where $\mathcal{I}_1, \mathcal{I}_2$ are disjoint subsets of the whole indices $\{1, \dots, n\} \times \{1, \dots, n\}$. \mathcal{I}_1 is the index set of 1-box constraints, for each (i, j) of which the (i, j) -th component of the true system belongs to $[l_{ij}, r_{ij}] \subseteq \mathbb{R}$. Similarly, \mathcal{I}_2 is the one of 2-box constraints, for each (i, j) of which the true (i, j) -th element is assumed to lie in $[l_{ij}, c_{ij} - k_{ij}] \cup [c_{ij} + k_{ij}, r_{ij}]$. In a similar manner to Theorem IV.1, we derive the Riemannian gradients of the constraints.

We will compare the following two modelings with the above RNLO modeling. One is the Euclidean version of the prediction error method with prior knowledge (Euclidean nonlinear optimization; ENLO):

$$\text{minimize}_{A \in \mathbb{R}^{n \times n}} \quad \frac{1}{N} \|X' - (I + hA)X\|_{\mathbb{F}}^2 \quad (26a)$$

$$\text{subject to} \quad l_{ij} \leq e_i^\top A e_j \leq r_{ij}, \quad ((i, j) \in \mathcal{I}_1 \cup \mathcal{I}_2), \quad (26b)$$

$$k_{ij}^2 \leq (e_i^\top A e_j - c_{ij})^2, \quad ((i, j) \in \mathcal{I}_2), \quad (26c)$$

which is obtained by replacing $(J - R)Q$ with A in (25), respectively. Since this problem does not impose the stability condition on A , the solutions are not necessarily stable. The other is the Riemannian modeling that minimizes (25a) under the absence of any constraints.

In the experiments, we consider the case $n = 10, h = 0.02$, and $N = 40$. Indices of \mathcal{I}_1 are randomly picked up from $\{1, 2, 3, \dots, 10\}^2$, and then those of \mathcal{I}_2 are from $\{1, 2, 3, \dots, 10\}^2 \setminus \mathcal{I}_1$. We randomly generate $(J^*, R^*, Q^*) \in \mathcal{M}$ using Manopt command `M.rand()` and then let the true system $A^* := (J^* - R^*)Q^* \in \mathbb{R}^{n \times n}$. For each $(i, j) \in \mathcal{I}_1$, l_{ij} and r_{ij} are randomly generated so that $e_i^\top A^* e_j \in [l_{ij}, r_{ij}]$. As well, for each $(i, j) \in \mathcal{I}_2$, l_{ij}, r_{ij}, k_{ij} and c_{ij} are randomly generated so that $e_i^\top A^* e_j \in [l_{ij}, c_{ij} - k_{ij}] \cup [c_{ij} + k_{ij}, r_{ij}]$. We set the ratios of 1-box and 2-box constraints as 0.2 and 0.1, respectively. Each component of x_0 is generated uniformly at random in the range $(-1000, 1000)$, and $\{x_k\}_{k=1}^N$ is determined according to $x_{k+1} = \exp(Ah)x_k$. Additive white Gaussian noise with a Signal-to-Noise Ratio (SNR) of 10 dB or 20 dB is then added to these values. We also scale the data by dividing all the elements by $\|x_0\|$ for the sake of the numerical stability. We adopt randomly generated $(J, R, Q) \in \mathcal{M}$ as the initial point.

B. Experimental environment

We compare the following six pairs of modelings and algorithms:

- **RNLO-by-RSQO**: our RNLO modeling (25) and RSQO method proposed in [25].
- **EE**: the Euclidean nonlinear modeling (26) and `fmincon SQO`, an Euclidean SQO or SQP solver.
- **EET**: the Euclidean nonlinear modeling (26) and `fmincon SQO` with the projection to the stable matrix set proposed in [30].
- **LS**: the Euclidean nonlinear modeling (26a) without the constraints and `fmincon SQO`.
- **GD**: the Riemannian modeling (25a) without the constraints and the Riemannian gradient decent method [27].
- **RTR**: the Riemannian modeling (25a) without the constraints and the Riemannian trust region method [27].

We evaluate the quality of solutions of the optimization problems in view of the first largest real part of eigenvalues of A , because it will govern the behavior of the system. We also monitor the optimization process of each algorithm with respect to the max violation; let

$$\begin{aligned} g_{ij}^l(\theta) &:= -e_i^\top (J - R)Qe_j + l_{ij}, \quad ((i, j) \in \mathcal{I}_1 \cup \mathcal{I}_2), \\ g_{ij}^r(\theta) &:= e_i^\top (J - R)Qe_j - r_{ij}, \quad ((i, j) \in \mathcal{I}_1 \cup \mathcal{I}_2), \\ g_{ij}^c(\theta) &:= -(e_i^\top (J - R)Qe_j - c_{ij})^2 + k_{ij}^2, \quad ((i, j) \in \mathcal{I}_2) \end{aligned}$$

be the constraints in (25b) and (25c). The max violation at θ is defined as

$$\max \left\{ 0, \max_{(i,j) \in \mathcal{I}_2} \{g_{ij}^c(\theta)\}, \max_{(i,j) \in \mathcal{I}_1 \cup \mathcal{I}_2} \{g_{ij}^l(\theta), g_{ij}^r(\theta)\} \right\}.$$

Throughout the experiments, we employ the identity mapping as B^k in (20), which clearly satisfies Assumption 1.

C. Numerical results and discussion

We solved 50 problem instances by the six pairs of modelings and algorithms. All the algorithms were run from the identical initial point, which was produced for each problem instance, and terminated after 100 seconds of CPU time. The results are depicted in Figs. 1, 2, 3, and 4. Note that the scales in the subfigures within each figure significantly differ, and the result shown in Fig. 1 was obtained by randomly selecting a solution from one of the 50 problem instances. Although SNR was 10 dB for Fig. 1, it was 20 dB for the other figures. Figs. 1 and 2 demonstrate the eigenvalue distributions obtained by the various methods. Fig. 3 denotes the boxplots of the relative errors of the first largest real part of eigenvalue between true system and numerical solutions among 50 trials, respectively. Fig. 4 denotes the boxplots of the max violation among 50 trials, respectively.

As shown in Figs. 1, 2, and 3, RNLO-by-RSQO was considerably better than the other methods in terms of the identification of the eigenvalues. In particular, Fig. 3 showed that only RNLO-by-RSQO can approximately estimate the most important eigenvalue, that is, the eigenvalue with the first largest real part. Here, we omitted the relative error attained

by EE, because the error was considerably larger than those attained by the other methods.

While we haven't shown the figure due to page limitations, an important observation needs to be highlighted. Using the same initial point, EE's objective value was superior to that of RNLO-by-RSQO. However, in terms of eigenvalues, RNLO-by-RSQO considerably outperformed EE. Furthermore, when EE was initialized with the matrix A obtained from RNLO-by-RSQO, its resulting eigenvalues were inferior to those generated by RNLO-by-RSQO. These findings emphasize EE's tendency to overfit noisy data, contrasting with the robustness demonstrated by RNLO-by-RSQO.

According to Fig. 4, RNLO-by-RSQO was significantly better than the other methods except for EE in terms of the max violation. This result can be expected, because only RNLO-by-RSQO and EE incorporate prior information in the form of equality and inequality constraints on the parameters. The abundance of outliers implies that EE is susceptible to noise. Note that the max violation of EET, which projects the unstable eigenvalues obtained by EE onto the imaginary axis, is larger than that of EE. This is because the projection does not take into account equality or inequality constraints.

It is remarkable that RNLO-by-RSQO is superior to GD and RTR in terms of the eigenvalues, as shown in Figs. 1, 2, and 3. This results from the fact that GD and RTR consider to solve (25a) without the constraints unlike RNLO-by-RSQO. In other words, the problem for GD and RTR does not incorporate any prior knowledge other than stability unlike RNLO-by-RSQO. Thus, GD and RTR generated convergence sequences to some bad solutions as in the results of Figs. 1, 2, and 3.

In summary, the superiority of RNLO-by-RSQO can be attributed to the parametrization based on Proposition II.1 and the incorporation of prior knowledge concerning equality and inequality constraints. A particularly noteworthy point is that RNLO-by-RSQO is more robust to noise compared to EE. The only difference between these methods is whether or not they utilize the parametrization based on Proposition II.1.

VI. CONCLUSION AND FUTURE WORK

We developed a prediction error method that ensures the stability of a linear system and meets with the prior knowledge. The method employs RNLO, a class of the nonlinear optimization on a Riemannian manifold. For solving this RNLO, we proposed to use RSQO. Numerical experiments with comparisons declare the effectiveness of our RNLO formulation in terms of the stability of the system and prior knowledge together with the superiority of RSQO.

An initial point generally affects numerical results in nonlinear optimization, and a sampling time h also does the accuracy of the system identification. Hence, a good proposal for the selection of an initial point and a sampling time is a possible future research. Moreover, it should be noted that our approach presented in this paper does not always identify a stable system with prior knowledge information and sufficient accuracy. In fact, for example, in situations where there is little data and a large amount of noise, the estimated matrix using our approach can be a stable matrix near the imaginary axis on the complex

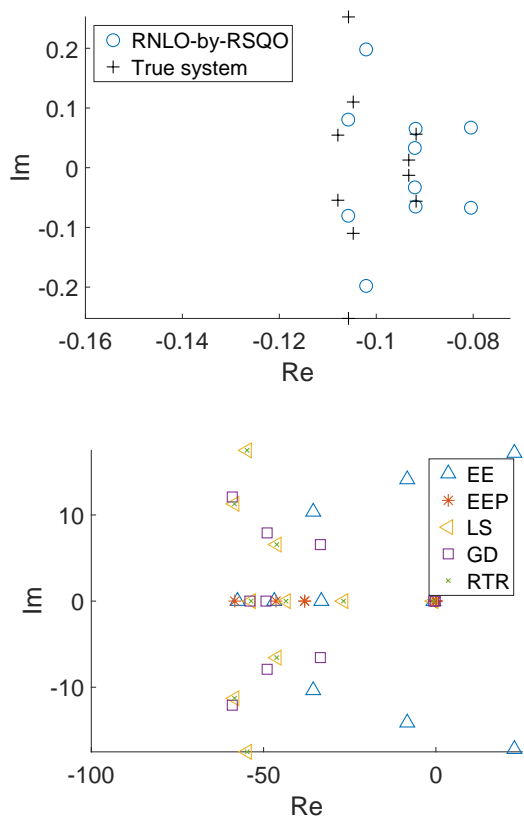


Fig. 1. Eigenvalue distribution (SNR = 10 dB)

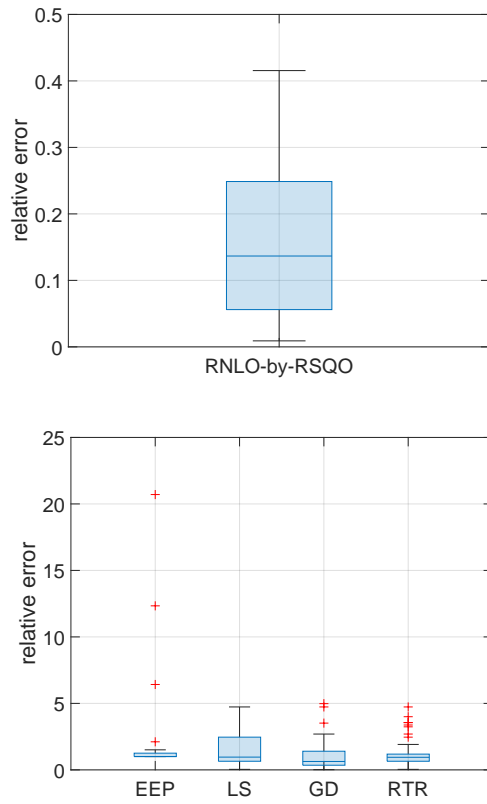


Fig. 3. Relative error of the first largest real part of the eigenvalue among 50 trials (SNR = 20 dB)

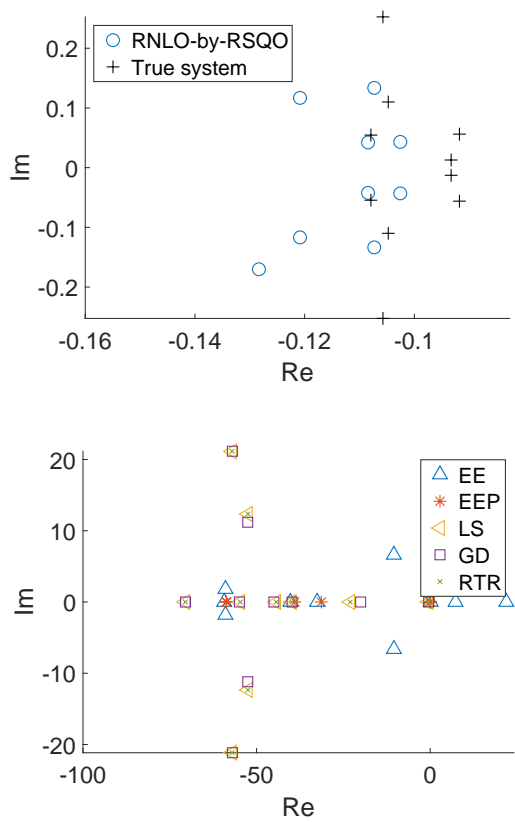


Fig. 2. Eigenvalue distribution (SNR = 20 dB)

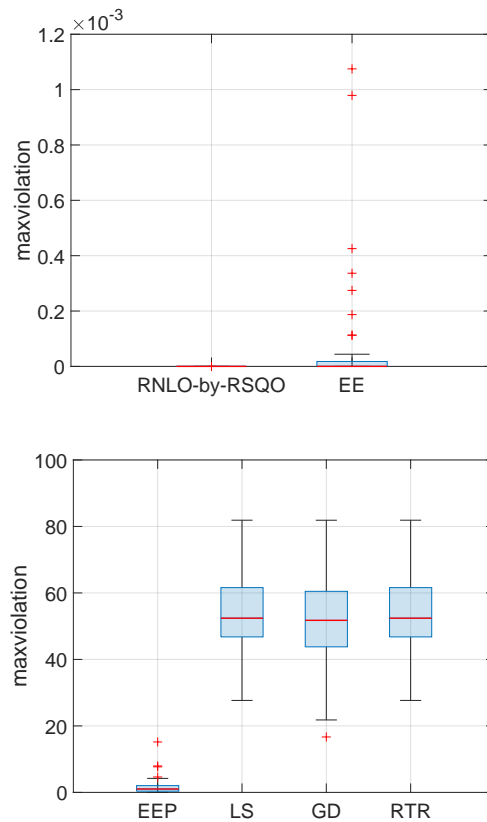


Fig. 4. The max violation among 50 trials (SNR = 20 dB)

plane even if the true eigenvalues are far away from the axis. This issue should be resolved in the future research.

REFERENCES

- [1] L. Ljung: *System identification: theory for the user*. Second. Prentice Hall, 1999.
- [2] K. Sato, H. Sato, and T. Damm: Riemannian optimal identification method for linear systems with symmetric positive-definite matrix. *IEEE Transactions on Automatic Control* 65.11 (2020), pp. 4493–4508.
- [3] H. Sato and K. Sato: Riemannian optimal system identification algorithm for linear MIMO systems. *IEEE Control Systems Letters* 1.2 (2017), pp. 376–381.
- [4] T. McKelvey, A. Helmersson, and T. Ribarits: Data driven local coordinates for multivariable linear systems and their application to system identification. *Automatica* 40 (2004), pp. 1629–1635.
- [5] A. Wills and B. Ninness: On gradient-based search for multivariable system estimates. *IEEE Transactions on Automatic Control* 53.1 (2008), pp. 298–306.
- [6] T. Katayama: *Subspace methods for system identification*. Springer-Verlag, 2005.
- [7] M. C. Campi, T. Sugie, and F. Sakai: An iterative identification method for linear continuous-time systems. *IEEE Transactions on Automatic Control* 53.7 (2008), pp. 1661–1669.
- [8] P. V. Overschee and B. D. Moor: N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica* 30.1 (1994), pp. 75–93.
- [9] M. Verhaegen and A. Hansson: N2SID: Nuclear norm subspace identification of innovation models. *Automatica* 72 (2016), pp. 57–63.
- [10] M. Verhaegen: Identification of the deterministic part of MIMO state space models given in innovation s form from input-output data. *Automatica* 30.1 (1994), pp. 61–74.
- [11] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor: *Dynamic mode decomposition: data-driven modeling of complex systems*. Society for Industrial and Applied Mathematics, 2016.
- [12] J. Heiland and B. Unger: *Identification of linear time-invariant systems with dynamic mode decomposition*. arXiv:2109.06765. 2021.
- [13] S. T. M. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley: Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids* 57.42 (2016).
- [14] T. V. Gestel, J. A. K. Suykens, P. V. Dooren, and B. D. Moor: Identification of stable models in subspace identification by using regularization. *IEEE Transactions on Automatic Control* 46.9 (2001), pp. 1416–1420.
- [15] S. L. Lacy and D. S. Bernstein: Subspace identification with guaranteed stability using constrained optimization. *IEEE Transactions on Automatic Control* 48.7 (2003), pp. 1259–1263.
- [16] D. N. Miller and R. A. D. Callafon: Subspace identification with eigenvalue constraints. *Automatica* 49 (2013), pp. 2468–2473.
- [17] S. Siddiqi, B. Boots, and G. J. Gordon: A constraint generation approach to learning stable linear dynamical systems. *Advances in neural information processing systems* 20 (2007).
- [18] J. Umenberger, J. Wågberg, I. R. Manchester, and T. B. Schön: Maximum likelihood identification of stable linear dynamical systems. *Automatica* 96 (2018), pp. 280–292.
- [19] J. Umenberger and R. Manchester: Convex bounds for equation error in stable nonlinear identification. *IEEE Control Systems Letters* 3.1 (2019), pp. 73–78.
- [20] J. Umenberger and R. Manchester: Specialized interior-point algorithm for stable nonlinear system identification. *IEEE Transactions on Automatic Control* 64.6 (2019), pp. 2442–2456.
- [21] S. Prajna, A. van der Schaft, and G. Meinsma: An LMI approach to stabilization of port-controlled Hamiltonian systems. *Systems & Control Letters* 45 (2002), pp. 371–385.
- [22] K. Sato: Riemannian optimal model reduction of stable linear systems. *IEEE Access* 7 (2019), pp. 14689–14698.
- [23] W. H. Yang, L.-H. Zhang, and R. Song: Optimality conditions for the nonlinear programming problems on Riemannian manifolds. *Pacific Journal of Optimization* 10.2 (2014), pp. 415–434.
- [24] C. Liu and N. Boumal: Simple algorithms for optimization on Riemannian manifolds with constraints. *Applied Mathematics & Optimization* (2019), pp. 1–33.
- [25] M. Obara, T. Okuno, and A. Takeda: Sequential quadratic optimization for nonlinear optimization problems on Riemannian manifolds. *SIAM Journal on Optimization* 32.3 (2022), pp. 822–853.
- [26] B. Jeuris, R. Vandebril, and B. Vandereycken: A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis* 39 (2012), pp. 379–402.
- [27] P.-A. Absil, R. Mahony, and R. Sepulchre: *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.
- [28] R. Bergmann and R. Herzog: Intrinsic formulation of KKT conditions and constraint qualifications on smooth manifolds. *SIAM Journal on Optimization* 29.4 (2019), pp. 2423–2444.
- [29] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre: Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research* 15.1 (2014), pp. 1455–1459.
- [30] V. Noferini and F. Poloni: Nearest Ω -stable matrix via Riemannian optimization. *Numerische Mathematik* 148.4 (2021), pp. 817–851.