IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 26, 2018, accepted January 6, 2019, date of publication January 11, 2019, date of current version February 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2892071

Riemannian Optimal Model Reduction of Stable Linear Systems

KAZUHIRO SATO, (Member, IEEE)
School of Regional Innovation and Social Design Engineering, Kitami Institute of Technology, Kitami 090-8507, Japan
e-mail: ksato@mail.kitami-it.ac.jp

This work was supported in part by JSPS KAKENHI under Grant JP18K13773, and in part by The Asahi Glass Foundation.

ABSTRACT In this paper, we develop a method for solving the problem of minimizing the H^2 error norm between the transfer functions of the original and reduced systems on the product set of the set of stable matrices and two Euclidean spaces. That is, we develop a method for identifying the optimal reduced system from all the asymptotically stable linear systems. However, it is difficult to develop an algorithm for solving this problem, because the set of stable matrices is highly non-convex. To overcome this issue, we show that the problem can be transformed into a tractable Riemannian optimization problem on the product manifold of the set of skew-symmetric matrices, the manifold of the symmetric positive-definite matrices, and two Euclidean spaces. The asymptotic stability of the reduced systems constructed using optimal solutions to our problem is preserved. To solve the reduced problem, the Riemannian gradient and Hessian are derived, and a Riemannian trust-region method is developed. The initial point in the proposed approach is selected using the output from the balanced truncation (BT) method. The numerical experiments demonstrate that our method considerably improves the results given by BT and other methods in terms of the H^2 norm and also provides the reduced systems that are globally near-optimal solutions to the problem of minimizing the H^∞ error norm. Moreover, we show that our method provides a better reduced model than the BT and other methods from the viewpoint of the frequency response.

INDEX TERMS H^2 optimal model reduction, Riemannian optimization.

I. INTRODUCTION

Accurate modeling is essential to various system control methods. However, the complexity of the controller is usually the same as that of the system [1]. That is, as the scale of the system to be controlled increases, the controller becomes more complex. This additional complexity can result in storage, accuracy, and computational speed problems [2]. Thus, we frequently need to approximate the original system as a small-scale model with high accuracy.

To produce a highly accurate reduced model, we use model reduction methods. The most famous approach is the balanced truncation (BT) method [1]–[4]. BT provides an asymptotically stable reduced model with guaranteed H^∞ bounds, as long as the original model is asymptotically stable. Another famous technique is the moment matching method [5]–[8], which produces a reduced system matching some coefficients of the transfer function of a given linear system. Moreover, for network systems, some model reduction methods that preserve the original network structure have been proposed [9]–[12]. Although the above mentioned methods do not guarantee any optimality, the H^2 optimal model reduction problem was studied for general asymptotically stable linear systems by formulating the optimization problem on the Stiefel manifold in [13] and [14]. However, the methods could be improved further as shown in Section V, because they only search for the optimal reduced model from a subset of all asymptotically stable linear systems.

In this study, we develop a novel H^2 optimal model reduction method for asymptotically stable linear systems. The problem is formulated as a minimization problem of the H^2 error norm between the transfer functions of the original and reduced systems on the product set of the set of stable matrices and two Euclidean spaces. That is, unlike [13] and [14], we search for the optimal reduced model with respect to all asymptotically stable linear systems. However, it is difficult to develop an algorithm for solving this problem, because the set of stable matrices is highly non-convex [15].

The contributions of this paper are summarized as follows.
1) We show that the original difficult problem can be transformed into a tractable Riemannian optimization problem

on the product manifold of the vector space of skew symmetric matrices, the manifold of symmetric positive-definite matrices, and two Euclidean spaces. To solve the problem, we propose a Riemannian trust-region method. To this end, we derive the Riemannian gradient and Hessian of the objective function. Moreover, we suggest applying the result of the BT method as an initial point of our algorithm.

2) Numerical experiments demonstrate that our proposed method improves the results of the BT method in the sense of the H^2 and H^∞ norms. That is, although the aim of our optimization problem is to minimize the H^2 error norm between the transfer functions of the original and reduced systems, the H^∞ error norm between those is also smaller than that of the BT method. Furthermore, we illustrate that our proposed method produces reduced systems that are globally near-optimal solutions to the problem of minimizing the H^∞ error norm. Moreover, we show that our method provides a better reduced model than the BT method from the viewpoint of the frequency response. In addition, we illustrate that our method using BT initialization is better than our method using random initializations and the method proposed in [13].

The remainder of this paper is organized as follows. In Section II, we formulate the H^2 optimal model reduction problem on the product set of the set of stable matrices and two Euclidean spaces. In Section III, we transform the problem into a tractable Riemannian optimization problem. In Section IV, we propose an optimization algorithm for solving our problem and a technique for choosing the initial point. In Section V, we demonstrate that our method is more effective than the BT and other methods. Finally, our conclusions are presented in Section VI.

Notation: The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The identity matrix of size n is denoted by I_n. The symbol $\mathrm{Skew}(n)$ denotes the vector space of skew-symmetric matrices in $\mathbb{R}^{n\times n}$. The manifold of symmetric positive-definite matrices in $\mathbb{R}^{n\times n}$ is denoted by $\mathrm{Sym}_+(n)$. The tangent space at x on a manifold X is denoted by T_xX. Given a matrix $A \in \mathbb{R}^{n\times n}$, $\mathrm{tr}(A)$ denotes the sum of the elements on the diagonal of A, and $A_{i,j}$ denotes the entry in row i and column j. Moreover, $\mathrm{sym}(A)$ and $\mathrm{sk}(A)$ denote the symmetric and skew-symmetric parts of A, respectively; i.e., $\mathrm{sym}(A) = \frac{A+A^T}{2}$ and $\mathrm{sk}(A) = \frac{A-A^T}{2}$. Here, A^T denotes the transpose of A. Given a vector $v \in \mathbb{C}^n$, $||v||$ denotes the Euclidean norm. The Hilbert space $L^2(\mathbb{R}^n)$ is defined by $L^2(\mathbb{R}^n) := \left\{f : [0,\infty) \to \mathbb{R}^n \;\middle|\; \int_0^\infty ||f(t)||^2\, dt < \infty\right\}$. Given a measurable function $f : [0,\infty) \to \mathbb{R}^n$, $||f||_{L^2}$ and $||f||_{L^\infty}$ denote the L^2 and L^∞ norms of f, respectively, i.e.,

$$||f||_{L^2} := \sqrt{\int_0^\infty ||f(t)||^2\, dt}, \quad ||f||_{L^\infty} := \sup_{t\geq 0} ||f(t)||.$$

Given a matrix $A \in \mathbb{C}^{n\times n}$, $||A||$ and $||A||_F$ denote the induced and Frobenius norms, i.e.,

$$||A|| := \sup_{v\in\mathbb{C}^n\setminus\{0\}} \frac{||Av||}{||v||}, \quad ||A||_F := \sqrt{\mathrm{tr}(A^*A)},$$

where the superscript $*$ denotes Hermitian conjugation, and $\mathrm{tr}(A)$ is the trace of A, i.e., the sum of the diagonal elements of A. For a matrix function $G(s) \in \mathbb{C}^{n\times n}$, $||G||_{H^2}$ and $||G||_{H^\infty}$ denote the H^2 and H^∞ norms of G, respectively, i.e.,

$$||G||_{H^2} := \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} ||G(i\omega)||_F^2\, d\omega},$$
$$||G||_{H^\infty} := \sup_{\omega\in\mathbb{R}} \bar{\sigma}(G(i\omega)),$$

where i is the imaginary unit, and $\bar{\sigma}(G(i\omega))$ denotes the maximum singular value of $G(i\omega)$.

II. PROBLEM SETUP

This section describes the formulation of our problem.

As the original system, we consider the linear continuous-time system

$$\begin{cases} \dot{x} = Ax + Bu, \\ y = Cx, \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $y \in \mathbb{R}^p$ are the state, input, and output, respectively. The matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{p\times n}$ are constant matrices. Throughout this paper, we assume that system (1) is asymptotically stable; i.e., the real parts of all the eigenvalues of the matrix A are negative. We also call the matrix A stable if system (1) is asymptotically stable. That is, we assume $A \in \mathbb{S}^{n\times n}$, where $\mathbb{S}^{n\times n}$ denotes the set of all stable matrices in $\mathbb{R}^{n\times n}$.

In this paper, we consider the following H^2 optimal model reduction problem of preserving the stability.

> Problem 1:
>
> minimize $\quad ||G - \hat{G}_r||_{H^2}$
>
> subject to $\quad (A_r, B_r, C_r) \in \mathbb{S}^{r\times r} \times \mathbb{R}^{r\times m} \times \mathbb{R}^{p\times r}$.

Here, G is the transfer function of system (1), i.e.,

$$G(s) := C(sI_n - A)^{-1}B, \quad s \in \mathbb{C},$$

and $\hat{G}_r$ is the transfer function of the reduced system

$$\begin{cases} \dot{\hat{x}}_r = A_r\hat{x}_r + B_r u, \\ \hat{y}_r = C_r\hat{x}_r. \end{cases} \tag{2}$$

Note that if $u \in L^2(\mathbb{R}^m)$, then the error $y - \hat{y}_r$ satisfies

$$||y - \hat{y}_r||_{L^\infty} \leq ||G - \hat{G}_r||_{H^2} \cdot ||u||_{L^2}. \tag{3}$$

The proof is shown in Appendix A. That is, if $||G - \hat{G}_r||_{H^2}$ is sufficiently small, then we can expect $||y - \hat{y}_r||_{L^\infty}$ to become almost zero for any u with a small $||u||_{L^2}$.

It is difficult to solve Problem 1 because the set $\mathbb{S}^{r\times r}$ is highly non-convex [15]. To develop an algorithm for solving Problem 1, we transform Problem 1 into an equivalent, tractable Riemannian optimization problem.

III. EQUIVALENT RIEMANNIAN OPTIMIZATION PROBLEM

This section proves that Problem 1 is equivalent to

Problem 2:

$$\begin{aligned}&\text{minimize} \quad f(J_r, R_r, B_r, C_r) := \|G - G_r\|_{H^2}^2\\ &\text{subject to} \quad (J_r, R_r, B_r, C_r) \in M.\end{aligned}$$

Here, G_r is the transfer function of the reduced system

$$\begin{cases}\dot{x}_r = (J_r - R_r)x_r + B_r u,\\ y_r = C_r x_r,\end{cases} \tag{4}$$

and

$$M := \mathrm{Skew}(r) \times \mathrm{Sym}_+(r) \times \mathbb{R}^{r\times m} \times \mathbb{R}^{p\times r}.$$

Note that system (4) is also asymptotically stable, because the real parts of all the eigenvalues of $J_r - R_r$ are negative.

To this end, we first note that asymptotically stable system (1) can be transformed into

$$\begin{cases}\dot{x} = (J - R)\mathcal{Q}x + Bu,\\ y = Cx,\end{cases} \tag{5}$$

where $\mathcal{Q} \in \mathrm{Sym}_+(n)$ and

$$J := \frac{1}{2}(A\mathcal{Q}^{-1} - \mathcal{Q}^{-1}A^T) \in \mathrm{Skew}(n),$$

$$R := -\frac{1}{2}(A\mathcal{Q}^{-1} + \mathcal{Q}^{-1}A^T) \in \mathrm{Sym}_+(n).$$

Although the proof can be found in [16, Proposition 1], we repeat it here for completeness. By definition, it is clear that $J \in \mathrm{Skew}(n)$ and $(J - R)\mathcal{Q} = A$. Thus, we show that $R \in \mathrm{Sym}_+(n)$. As shown in [3], the asymptotic stability of system (1) is equivalent to that for any $\mathcal{P} \in \mathrm{Sym}_+(n)$, there uniquely exists $\mathcal{Q} \in \mathrm{Sym}_+(n)$ such that

$$A^T\mathcal{Q} + \mathcal{Q}A = -\mathcal{P}.$$

That is, system (1) is asymptotically stable if and only if there exists $\mathcal{Q} \in \mathrm{Sym}_+(n)$ such that

$$-(A^T\mathcal{Q} + \mathcal{Q}A) \in \mathrm{Sym}_+(n). \tag{6}$$

Hence, the asymptotic stability of system (1) implies that

$$R = -\frac{1}{2}\mathcal{Q}^{-1}(A^T\mathcal{Q} + \mathcal{Q}A)\mathcal{Q}^{-1} \in \mathrm{Sym}_+(n).$$

Conversely, $A = (J-R)\mathcal{Q}$ with any $(J, R, \mathcal{Q}) \in \mathrm{Skew}(n) \times \mathrm{Sym}_+(n) \times \mathrm{Sym}_+(n)$ is stable. To see this, consider

$$\dot{x} = (J - R)\mathcal{Q}x, \tag{7}$$

and $H(x) := \frac{1}{2}x^T\mathcal{Q}x$. Then, the derivative of H along the trajectories of system (7) is evaluated by

$$\dot{H}(x) = x^T((J - R)\mathcal{Q})^T\mathcal{Q}x = -x^T\mathcal{Q}R\mathcal{Q}x.$$

That is, $H(0) = 0$, $H(x) > 0$ for $x \neq 0$, and $\dot{H}(x) < 0$ for $x \neq 0$. Thus, the function H is a Lyapunov function and $(J - R)\mathcal{Q}$ is stable. Hence, the set $\mathbb{S}^{n\times n}$ can be characterized by

$$\mathbb{S}^{n\times n} = \{(J - R)\mathcal{Q} \mid (J, R, \mathcal{Q}) \in \mathrm{Skew}(n) \times \mathrm{Sym}_+(n) \times \mathrm{Sym}_+(n)\}.$$

Note that we can easily find $\mathcal{Q} \in \mathrm{Sym}_+(n)$ satisfying (6). In fact, because the matrix A is stable, there uniquely exists $\mathcal{Q} \in \mathrm{Sym}_+(n)$ satisfying the Lyapunov equation

$$A^T\mathcal{Q} + \mathcal{Q}A + I_n = 0, \tag{8}$$

as shown in [3]. Lyapunov equation (8) can be efficiently solved using the Bartels–Stewart algorithm [17].

Because the transfer function of (5) coincides with that of (1), Problem 1 is equivalent to

Problem 3:

$$\begin{aligned}&\text{minimize} \quad \|G - \check{G}_r\|_{H^2}\\ &\text{subject to} \quad (J_r, R_r, \mathcal{Q}_r, B_r, C_r) \in N.\end{aligned}$$

Here, $\check{G}_r$ is the transfer function of the reduced system

$$\begin{cases}\dot{\check{x}}_r = (J_r - R_r)\mathcal{Q}_r\check{x}_r + B_r u,\\ \check{y}_r = C_r\check{x}_r,\end{cases}$$

and $N := \mathrm{Skew}(r) \times \mathrm{Sym}_+(r) \times \mathrm{Sym}_+(r) \times \mathbb{R}^{r\times m} \times \mathbb{R}^{p\times r}$.

Next, we show that Problem 3 can be transformed into Problem 2. To see this, we note that system (5) is equivalent to the form

$$\begin{cases}\dot{\tilde{x}} = (\tilde{J} - \tilde{R})\tilde{x} + \tilde{B}u,\\ y = \tilde{C}\tilde{x},\end{cases} \tag{9}$$

where $\tilde{x} \in \mathbb{R}^n$, $\tilde{J} \in \mathrm{Skew}(n)$, $\tilde{R} \in \mathrm{Sym}_+(n)$, $\tilde{B} \in \mathbb{R}^{n\times m}$, and $\tilde{C} \in \mathbb{R}^{p\times n}$. In fact, because $\mathcal{Q}$ is a positive-symmetric matrix, there exists a unique lower triangular $L \in \mathbb{R}^{n\times n}$ with positive diagonal entries such that $\mathcal{Q} = LL^T$. This is called the Cholesky decomposition of $\mathcal{Q}$. For a detailed explanation, see [18]. Thus, if we perform a coordinate transformation $\tilde{x} = (L^{-1})^T x$, we obtain (9), where $\tilde{J} = L^TJL$, $\tilde{R} = L^TRL$, $\tilde{B} = L^TB$, and $\tilde{C} = C(L^{-1})^T$. Because the transfer function of (9) coincides with that of (5), Problem 3 is equivalent to Problem 2.

From the above discussion, Problem 2 is equivalent to Problem 1, which completes the proof.

In contrast to Problem 1, we can develop an algorithm for solving Problem 2 using a Riemannian optimization method [19], as shown in the next section.

Remark 1: Reference [20] considered

$$\begin{aligned}&\text{minimize} \quad \|G - \hat{G}_r\|_{H^2}^2\\ &\text{subject to} \quad (A_r, B_r, C_r) \in \mathbb{R}^{r\times r} \times \mathbb{R}^{r\times m} \times \mathbb{R}^{p\times r},\end{aligned}$$

and proved that if reduced system (2) is controllable and observable, then at every stationary point of $||G - \hat{G}_r||^2_{H^2}$, we have that

$$A_r = W^T AV, \ B_r = W^T B, \ C_r = CV, \ W^T V = I_r.$$

Based on this fact, Antoulas *et al.* [5] and Gugercin *et al.* [7] developed an algorithm for finding such V and W. Although the algorithm can be applied to the model reduction of large-scale systems, a sequence produced by the algorithm does not generally converge to a local optimal solution, except for single-input–single-output symmetric systems [21].

Remark 2: References [13] and [14] considered

Problem 4:

$$\begin{aligned} &\text{minimize} \quad ||G - \bar{G}_r||^2_{H^2} \\ &\text{subject to} \quad U \in \mathrm{St}(r, n). \end{aligned}$$

Here, $\bar{G}_r$ is the transfer function of the reduced system

$$\begin{cases} \dot{\bar{x}}_r = U^T AU\bar{x}_r + U^T Bu, \\ \bar{y}_r = CU\bar{x}_r, \end{cases}$$

and $\mathrm{St}(r, n)$ is the Stiefel manifold defined by

$$\mathrm{St}(r, n) := \{U \in \mathbb{R}^{n\times r} \mid U^T U = I_r\}.$$

As explained in [13] and [14], if $A + A^T$ is negative-definite, then A and $U^T AU$ are stable, i.e., $A \in \mathbb{S}^{n\times n}$ and $U^T AU \in \mathbb{S}^{r\times r}$. Thus, if this is the case, a solution to Problem 4 is a feasible solution to Problem 1. That is, by solving Problem 4, we can obtain feasible solutions to Problem 1. However, in general, the optimal value of Problem 4 is larger than that of Problem 1, as shown in Section V. This fact has also been demonstrated for other model reduction problems [22], [23].

Remark 3: Instead of Problem 2, we can consider the following H^∞ optimal model reduction problem.

Problem 5:

$$\begin{aligned} &\text{minimize} \quad ||G - G_r||_{H^\infty} \\ &\text{subject to} \quad (J_r, R_r, B_r, C_r) \in M. \end{aligned}$$

However, in contrast to Problem 2, the objective function $||G - G_r||_{H^\infty}$ is not differentiable. Thus, it is difficult to develop an algorithm for solving Problem 5. In Section V, we demonstrate that there are examples for which we can obtain a globally near-optimal solution to Problem 5 by solving Problem 2.

Remark 4: Asymptotically stable system (9) with $\tilde{C} = \tilde{B}^T$ is an asymptotically stable port-Hamiltonian system [22]. In [22], the H^2 optimal model reduction method of preserving the port-Hamiltonian structure has been proposed using a Riemannian optimization.

IV. OPTIMIZATION ALGORITHM FOR PROBLEM 2

A. RIEMANNIAN GRADIENT, HESSIAN, AND EXPONENTIAL MAP

To develop an optimization algorithm for solving Problem 2, we derive the Riemannian gradient and Hessian of the objective function f, and compute the exponential map on the manifold M.

To this end, we first note that, because systems (9) and (4) are asymptotically stable, the objective function f can be expressed as

$$\begin{aligned} f(J_r, R_r, B_r, C_r) &= \mathrm{tr}(\tilde{C}\Sigma_c\tilde{C}^T + C_r P C_r^T - 2C_r X^T \tilde{C}^T) \\ &= \mathrm{tr}(\tilde{B}^T \Sigma_o \tilde{B} + B_r^T Q B_r + 2\tilde{B}^T Y B_r), \end{aligned}$$

where Σ_c, Σ_o, P, Q, X, and Y satisfy

$$(\tilde{J} - \tilde{R})\Sigma_c + \Sigma_c(\tilde{J} - \tilde{R})^T + \tilde{B}\tilde{B}^T = 0, \quad (10)$$

$$(\tilde{J} - \tilde{R})^T \Sigma_o + \Sigma_o(\tilde{J} - \tilde{R}) + \tilde{C}^T\tilde{C} = 0, \quad (11)$$

$$(J_r - R_r)P + P(J_r - R_r)^T + B_r B_r^T = 0, \quad (12)$$

$$(J_r - R_r)^T Q + Q(J_r - R_r) + C_r^T C_r = 0, \quad (13)$$

$$(\tilde{J} - \tilde{R})X + X(J_r - R_r)^T + \tilde{B}B_r^T = 0, \quad (14)$$

$$(\tilde{J} - \tilde{R})^T Y + Y(J_r - R_r) - \tilde{C}^T C_r = 0, \quad (15)$$

respectively. For a detailed derivation, see [13], [14], [20].

Let $\bar{f}$ denote the extension of the objective function f to the Euclidean space $\mathbb{R}^{r\times r} \times \mathbb{R}^{r\times r} \times \mathbb{R}^{r\times m} \times \mathbb{R}^{p\times r}$. In the same way as in [22]–[26], we then obtain

$$\begin{aligned} &\nabla\bar{f}(J_r, R_r, B_r, C_r) \\ &= 2(QP + Y^T X, -(QP + Y^T X), QB_r + Y^T B, C_r P - CX). \end{aligned} \quad (16)$$

To derive the Riemannian gradient and Hessian, we define the Riemannian metric of the manifold M as

$$\begin{aligned} &\langle(\xi_1, \eta_1, \zeta_1, \kappa_1), (\xi_2, \eta_2, \zeta_2, \kappa_2)\rangle_{(J_r,R_r,B_r,C_r)} \\ &:= \mathrm{tr}(\xi_1^T \xi_2) + \mathrm{tr}(R_r^{-1}\eta_1 R_r^{-1}\eta_2) + \mathrm{tr}(\zeta_1^T \zeta_2) + \mathrm{tr}(\kappa_1^T \kappa_2) \end{aligned} \quad (17)$$

for $(\xi_1, \eta_1, \zeta_1, \kappa_1), (\xi_2, \eta_2, \zeta_2, \kappa_2) \in T_{(J_r,R_r,B_r,C_r)}M$. It then follows from (22) in Appendix B and (16) that

$$\begin{aligned} &\mathrm{grad} f(J_r, R_r, B_r, C_r) \\ &= (2\mathrm{sk}(QP + Y^T X), -2R_r \mathrm{sym}(QP + Y^T X)R_r, \\ &\quad 2(QB_r + Y^T B), 2(C_r P - CX)). \end{aligned} \quad (18)$$

Furthermore, from (23) in Appendix B and (16), the Riemannian Hessian of f at (J_r, R_r, B_r, C_r) is given by

$$\begin{aligned} &\mathrm{Hess} f(J_r, R_r, B_r, C_r)[(J_r', R_r', B_r', C_r')] \\ &= (2\mathrm{sk}(Q'P + QP' + Y'^T X + Y^T X'), \\ &\quad -2R_r \mathrm{sym}(Q'P + QP' + Y'^T X + Y^T X')R_r \\ &\quad -2\mathrm{sym}(R_r' \mathrm{sym}(QP + Y^T X)R_r), \\ &\quad 2(Q'B_r + QB_r' + Y'^T B), 2(C_r' P + C_r P' - CX')), \end{aligned} \quad (19)$$

where P', Q', X', and Y' are the solutions to

$$
\begin{aligned}
&(J_r - R_r)P' + P'(J_r - R_r)^T + (J_r' - R_r')P + P(J_r' - R_r')^T \\
&\quad + B_r' B_r^T + B_r B_r'^T = 0, \\
&(J_r - R_r)^T Q' + Q'(J_r - R_r) + (J_r' - R_r')^T Q + Q(J_r' - R_r') \\
&\quad + C_r' C_r^T + C_r C_r'^T = 0, \\
&\quad (\tilde{J} - \tilde{R})^T X' + X'(J_r - R_r) + X(J_r' - R_r') + \tilde{B} B_r'^T = 0, \\
&\quad (\tilde{J} - \tilde{R})^T Y' + Y'(J_r - R_r) + Y(J_r' - R_r') - \tilde{C}^T C_r' = 0,
\end{aligned}
$$

respectively. Note that these equations are obtained by differentiating (12), (13), (14), and (15), respectively. Moreover, from (24) in Appendix A, we can define the exponential map on the manifold M as

$$
\begin{aligned}
&\mathrm{Exp}_{(J_r,R_r,B_r,C_r)}(\xi, \eta, \zeta, \kappa) \\
&\quad := (J_r + \xi, \mathrm{Exp}_{R_r}(\eta), B_r + \zeta, C_r + \kappa) \\
&\quad = (J_r + \xi, R_r \exp(R_r^{-1}\eta), B_r + \zeta, C_r + \kappa) \qquad (20)
\end{aligned}
$$

for any $(\xi, \eta, \zeta, \kappa) \in T_{(J_r,R_r,B_r,C_r)}M$.

B. TRUST-REGION METHOD FOR PROBLEM 2

Algorithm 1 describes the Riemannian trust-region method for solving Problem 2. At each iterate $p_r := (J_r, R_r, B_r, C_r) \in M$ in the Riemannian trust-region method, we evaluate the quadratic model $\hat{m}_{p_r}$ of the objective function f within a trust region:

$$
\begin{aligned}
&\hat{m}_{p_r}(\xi, \eta, \zeta, \kappa) \\
&= f(J_r, R_r, B_r, C_r) + \langle \mathrm{grad} f(J_r, R_r, B_r, C_r), (\xi, \eta, \zeta, \kappa)\rangle_{p_r} \\
&\quad + \frac{1}{2}\langle \mathrm{Hess} f(J_r, R_r, B_r, C_r)[(\xi, \eta, \zeta, \kappa)], (\xi, \eta, \zeta, \kappa)\rangle_{p_r}.
\end{aligned}
$$

Because we can construct the gradient and Hessian of f as in Section IV-A, we can construct $\hat{m}_{p_r}$. A trust region with a radius $\Delta > 0$ at $p_r \in M$ is defined as a ball in $T_{p_r}M$. The trust-region sub-problem at $p_r \in M$ with the radius Δ is thus defined as the problem of minimizing $\hat{m}_{p_r}(\xi, \eta, \zeta, \kappa)$ subject to

$$
||(\xi, \eta, \zeta, \kappa)||_{p_r} := \sqrt{\langle (\xi, \eta, \zeta, \kappa), (\xi, \eta, \zeta, \kappa)\rangle_{p_r}} \leq \Delta,
$$

where $(\xi, \eta, \zeta, \kappa) \in T_{p_r}M$. This sub-problem can be solved by the truncated conjugate gradient method [19]. We then compare the decrease in the objective function f and the model $\hat{m}_{p_r}$ attained by the resulting $(\xi_*, \eta_*, \zeta_*, \kappa_*)$, and use this to determine whether $(\xi_*, \eta_*, \zeta_*, \kappa_*)$ should be accepted and whether the trust region of radius Δ is appropriate. The constants $1/4$ and $3/4$ in the conditional expressions in Algorithm 1 are commonly used in the trust-region method for a general unconstrained optimization problem. These values ensure the convergence properties of the algorithm [19]. In fact, if the trust-region sub-problem is carefully solved, sequences generated by the Riemannian trust-region method converge quadratically under certain assumptions on the objective function in question [19].

Note that the reduced system attained by Algorithm 1 is asymptotically stable, because $(J_r, R_r, B_r, C_r) \in M$ at each iteration.

Algorithm 1 Trust-region method for Problem 2.

Choose an initial point $(p_r)_0 \in M$ and parameters $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$, $\gamma' \in [0, \frac{1}{4})$.
for $k = 0, 1, 2, \dots$ **do**
  Solve the following trust-region sub-problem for $(\xi, \eta, \zeta, \kappa)$ to obtain $(\xi_k, \eta_k, \zeta_k, \kappa_k) \in T_{(p_r)_k}M$:

$$
\begin{aligned}
&\text{minimize} \quad \hat{m}_{(p_r)_k}(\xi, \eta, \zeta, \kappa) \\
&\text{subject to} \quad ||(\xi, \eta, \zeta, \kappa)||_{(p_r)_k} \leq \Delta_k, \\
&\text{where} \quad (\xi, \eta, \zeta, \kappa) \in T_{(p_r)_k}M.
\end{aligned}
$$

  Evaluate

$$
\gamma_k := \frac{f(\mathrm{Exp}_{(p_r)_k}(0,0,0,0)) - f(\mathrm{Exp}_{(p_r)_k}(\xi_k,\eta_k,\zeta_k,\kappa_k))}{\hat{m}_{(p_r)_k}(0,0,0,0) - \hat{m}_{(p_r)_k}(\xi_k, \eta_k, \zeta_k, \kappa_k)}.
$$

  if $\gamma_k < \frac{1}{4}$ **then**
    $\Delta_{k+1} = \frac{1}{4}\Delta_k$.
  else if $\gamma_k > \frac{3}{4}$ and $||(\xi_k, \eta_k, \zeta_k, \kappa_k)||_{(p_r)_k} = \Delta_k$ **then**
    $\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$.
  else
    $\Delta_{k+1} = \Delta_k$.
  end if
  if $\gamma_k > \gamma'$ **then**
    $(p_r)_{k+1} = \mathrm{Exp}_{(p_r)_k}(\xi_k, \eta_k, \zeta_k, \kappa_k)$.
  else
    $(p_r)_{k+1} = (p_r)_k$.
  end if
end for

C. INITIAL POINT IN ALGORITHM 1

In this subsection, we describe a technique for choosing the initial point $(p_r)_0 \in M$ in Algorithm 1 using the output of the BT method [1]–[4]. The BT method can be implemented using the MATLAB command *balred* (i.e., we can easily implement the BT method), and provides satisfactory reduced models in many cases. That is, although there may be a lot of local minimizers for Problem 2, we can avoid bad local minimizers by using the BT method. In fact, we demonstrate that our proposed method performs better when BT initialization is used than when random initializations are used in Section V.

The BT method outputs the reduced matrices $(A_r)_{\mathrm{BT}}$, $(B_r)_{\mathrm{BT}}$, and $(C_r)_{\mathrm{BT}}$; the matrix $(A_r)_{\mathrm{BT}}$ is stable, because the original matrix A is stable [2]–[4]. Thus, there uniquely exists $\mathcal{Q}_r \in \mathrm{Sym}_+(r)$ satisfying

$$
(A_r)_{\mathrm{BT}}^T \mathcal{Q}_r + \mathcal{Q}_r (A_r)_{\mathrm{BT}} + I_r = 0,
$$

as explained in Section III. Next, we define

$$
\begin{aligned}
(J_r)_{\mathrm{BT}} &:= \frac{1}{2}\left((A_r)_{\mathrm{BT}}\mathcal{Q}_r^{-1} - \mathcal{Q}_r^{-1}(A_r)_{\mathrm{BT}}^T\right), \\
(R_r)_{\mathrm{BT}} &:= -\frac{1}{2}\left((A_r)_{\mathrm{BT}}\mathcal{Q}_r^{-1} + \mathcal{Q}_r^{-1}(A_r)_{\mathrm{BT}}^T\right).
\end{aligned}
$$

TABLE 1. $||G - G_r||_{H^2}$.

r	4	6	8	10	30
BT method	0.23248	0.11858	0.05526	0.02416	0.00002
The method proposed in [13]	0.10577	0.05953	0.03492	0.02139	0.00132
Proposed method (Random initialization)	0.03218	0.01229	0.01127	0.01054	0.00965
Proposed method (BT initialization)	0.03218	0.01061	0.00765	0.00552	0.00002

TABLE 2. $||G - G_r||_{H^\infty}$.

r	4	6	8	10	30
BT method	0.11669	0.05198	0.01646	0.00989	0.00008
The method proposed in [13]	0.12303	0.07946	0.05919	0.03559	0.00223
Proposed method (Random initialization)	0.04904	0.03230	0.03296	0.03278	0.03387
Proposed method (BT initialization)	0.04891	0.03182	0.01171	0.00908	0.00008
σ_{r+1}	0.02834	0.01198	0.00508	0.00262	0.00004

TABLE 3. $||\mathrm{grad}\, f(J_r, R_r, B_r, C_r)||_{(17)}$.

r	4	6	8	10	30
BT method	4.3×10^{-1}	1.5×10^{-1}	5.3×10^{-2}	1.8×10^{-2}	3.1×10^{-6}
Proposed method (Random initialization)	9.9×10^{-5}	9.6×10^{-5}	9.6×10^{-5}	9.9×10^{-5}	8.3×10^{-5}
Proposed method (BT initialization)	8.2×10^{-5}	9.8×10^{-5}	7.4×10^{-5}	7.4×10^{-5}	3.1×10^{-6}

Finally, we perform the Cholesky decomposition of $\mathcal{Q}_r = L_r L_r^T$, and set the initial point

$$(p_r)_0 = ((J_r)_0, (R_r)_0, (B_r)_0, (C_r)_0) = (L_r^T (J_r)_{\mathrm{BT}} L_r, L_r^T (R_r)_{\mathrm{BT}} L_r, L_r^T (B_r)_{\mathrm{BT}}, (C_r)_{\mathrm{BT}} (L_r^{-1})^T).$$

Note that, because transfer functions are invariant under coordinate transformations, we have that

$$(G_r)_{\mathrm{BT}} = (C_r)_{\mathrm{BT}} (sI_r - (A_r)_{\mathrm{BT}})^{-1} (B_r)_{\mathrm{BT}} = (C_r)_0 (sI_r - ((J_r)_0 - (R_r)_0))^{-1} (B_r)_0,$$

where $(G_r)_{BT}$ is the transfer function of the reduced system attained by the BT method.

V. NUMERICAL EXPERIMENTS

In this section, three examples are presented to illustrate that our method using BT initialization improves the result of BT in terms of the H^2 norm. Furthermore, we show that our method using BT initialization may provide better results for the H^∞ norm and the frequency response than BT. Moreover, we compare our method using BT initialization to our method using random initializations and the method proposed in [13]. To this end, we have used Manopt [27], which is a MATLAB toolbox for optimization on manifolds.

A. MASS-SPRING-DAMPER SYSTEM

We consider mass-spring-damper systems with masses m_i, spring constants k_i, and damping constants c_i ($i = 1, 2, \ldots, \frac{n}{2}$), where n is an even number. The inputs u_1 and u_2 are the external forces applied to the first two masses, m_1 and m_2. The output y_1 is the displacement of mass m_1. The state variables $\tilde{x}_j$ ($j = 1, 3, \ldots$) are the displacements of mass m_j and the state variables $\tilde{x}_k$ ($k = 2, 4, \ldots$) are the momentums of mass m_k. Here, we only consider the case

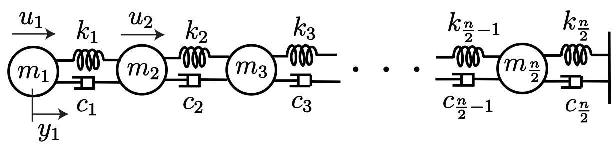

FIGURE 1. Mass-spring-damper system.

where $m_i = 4$, $k_i = 4$, and $c_i = 1$ ($i = 1, 2, \ldots, \frac{n}{2}$). The system can be described by (5) and the system matrices are given by $\tilde{J}_{1,2} = \tilde{J}_{3,4} = \cdots = \tilde{J}_{(n-1),n} = 1$, $\tilde{J}_{2,1} = \tilde{J}_{4,3} = \cdots = \tilde{J}_{n,(n-1)} = -1$, $\tilde{R}_{2,2} = \tilde{R}_{4,4} = \cdots = \tilde{R}_{n,n} = 1$, $\tilde{Q}_{1,1} = 4$, $\tilde{Q}_{2,2} = \tilde{Q}_{4,4} = \cdots = \tilde{Q}_{n,n} = \frac{1}{4}$, $\tilde{Q}_{3,3} = \tilde{Q}_{5,5} = \cdots = \tilde{Q}_{(n-1),(n-1)} = 8$, $\tilde{Q}_{1,3} = \tilde{Q}_{3,5} = \cdots = \tilde{Q}_{(n-3),(n-1)} = -4$, $\tilde{Q}_{3,1} = \tilde{Q}_{5,3} = \cdots = \tilde{Q}_{(n-1),(n-3)} = -4$, $\tilde{B}_{2,1} = \tilde{B}_{4,2} = 1$, $\tilde{C}_{1,1} = 1$, where the other entries of $\tilde{J}$, $\tilde{R}$, $\tilde{Q}$, $\tilde{B}$, and $\tilde{C}$ are zeros.

We reduced the dimension $n = 50$ to $r = 4, 6, 8, 10$, and 30, and compared our proposed method using BT and random initializations, the BT results, and results obtained using the method proposed by Sato and Sato [13]. Tables 1, 2, and 3 present the results for the H^2 error norm, H^∞ error norm, and gradient norm, respectively. Here, $||\cdot||_{(17)}$ denotes the induced norm from Riemannian metric (17). From Table 3, we can consider that we obtained local optimal solutions to Problem 2 when our method using BT and random initializations was applied. In contrast, the BT method did not provide a local optimal solution except $r = 30$. In Appendix D, we give the reduced matrices (J_r, R_r, B_r, C_r) obtained by our method using BT initialization in the case where $r = 4$. For each $r \in \{4, 6, 8, 10\}$, the H^2 and H^∞ error norms given by the BT method are greater than those of our method. In particular, for each $r \in \{4, 6, 8\}$, the H^2 error norms of our method are less than 1/7 of the corresponding error norms

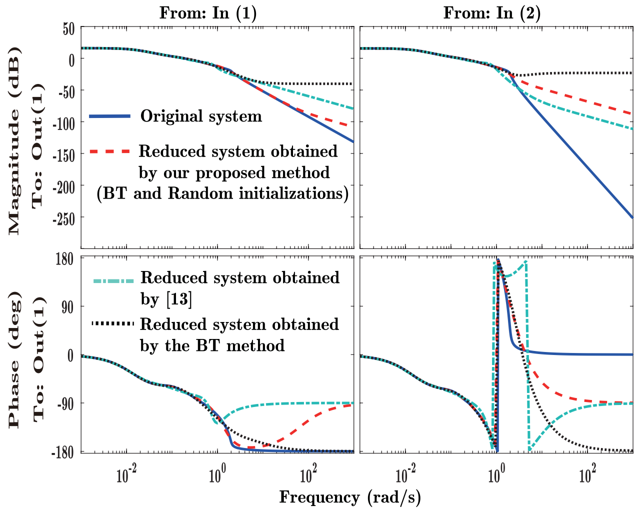

FIGURE 2. Bode diagram of original and reduced systems.

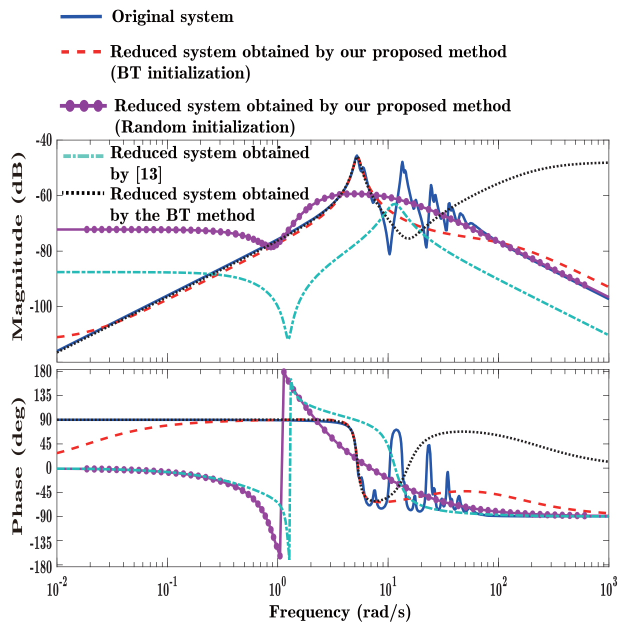

FIGURE 3. Bode diagram of original and reduced systems.

of the BT method. This is because the reduced models of the BT method are far from optimal, as can be seen from Table 3. Furthermore, Tables 1 and 2 show that the results of our method using BT initialization were better than those of our method using random initializations and the method proposed in [13]; however, the results of our method using BT and random initializations were almost the same when $r = 4$. Moreover, from Table 2 and Appendix C, we can conclude that our proposed method using BT initialization yielded a globally near-optimal solution to Problem 5. Here, σ_{r+1} is the $(r+1)$-th Hankel singular value of the original system.

Fig. 2 illustrates the Bode diagram of the original system and the reduced systems obtained by the proposed method using BT and random initializations, BT, and the method proposed by Sato and Sato [13] with $r = 4$. Here, note that the result of our proposed method using random initialization overlapped with the method using BT initialization. When the frequency is less than 1 rada/s, all reduced systems coincide with the original system. In contrast, when the frequency is greater than 1 rad/s, our reduced system is closer to the original system than the systems obtained using BT method and the method proposed in [13]. Thus, we can conclude that our proposed method produced better reduced systems than those produced using BT and the method proposed by Sato and Sato [13] in terms of the frequency response.

B. BUILDING SYSTEM

We consider the building model of the Los Angeles University Hospital reported in [28]. This model can be described by (1), and has $n = 48$ and $m = p = 1$. For $r = 3$, we obtained the following results:

- BT method

$$\begin{cases} ||G - G_r||_{H^2} = 0.0416, \\ ||G - G_r||_{H^\infty} = 0.0079. \end{cases}$$

- The method proposed in [13]

$$\begin{cases} ||G - G_r||_{H^2} = 0.0427, \\ ||G - G_r||_{H^\infty} = 0.0052. \end{cases}$$

- Proposed method (Random initialization)

$$\begin{cases} ||G - G_r||_{H^2} = 0.0036, \\ ||G - G_r||_{H^\infty} = 0.0042. \end{cases}$$

- Proposed method (BT initialization)

$$\begin{cases} ||G - G_r||_{H^2} = 0.0030, \\ ||G - G_r||_{H^\infty} = 0.0039. \end{cases}$$

Note that

$$||G - G||_{H^\infty} \geq \sigma_4 = 0.0019.$$

Fig. 3 illustrates the Bode diagram of the original and reduced systems with $r = 3$. The results given by the BT method and the proposed method using BT initialization were similar to that of the original system in the low-frequency region. In contrast, the results obtained using our method that used random initialization and the method proposed by Sato and Sato [13] were far from that of the original system in the low-frequency region. In addition, our proposed method using BT initialization yielded considerably better results than that of BT in the high-frequency region.

C. RANDOM SYSTEM

We also consider an asymptotically stable random system with $n = 100$ and $m = p = 1$. For $r = 5$, we obtained the following results:

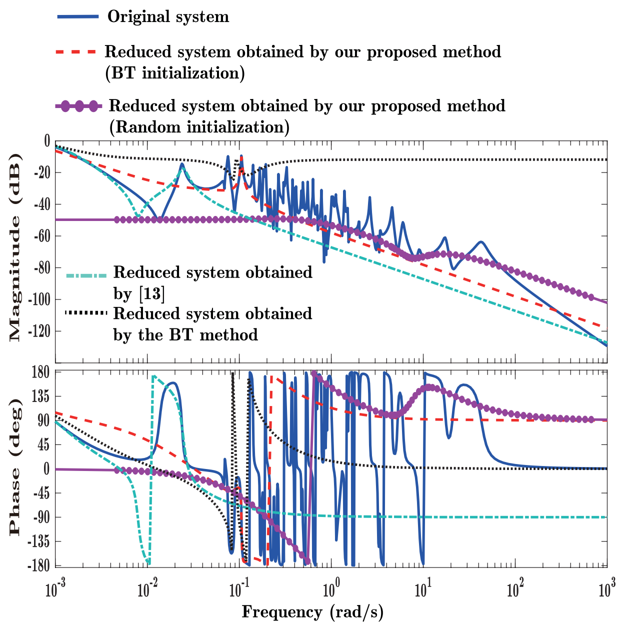

FIGURE 4. **Bode diagram of original and reduced systems.**

- BT method

$$\begin{cases} ||G - G_r||_{H^2} = 0.1042, \\ ||G - G_r||_{H^\infty} = 0.4914. \end{cases}$$

- The method proposed in [13]

$$\begin{cases} ||G - G_r||_{H^2} = 0.0230, \\ ||G - G_r||_{H^\infty} = 0.3700. \end{cases}$$

- Proposed method (Random initialization)

$$\begin{cases} ||G - G_r||_{H^2} = 0.0315, \\ ||G - G_r||_{H^\infty} = 1.3345. \end{cases}$$

- Proposed method (BT initialization)

$$\begin{cases} ||G - G_r||_{H^2} = 0.0213, \\ ||G - G_r||_{H^\infty} = 0.3039. \end{cases}$$

Note that

$$||G - G||_{H^\infty} \geq \sigma_6 = 0.1483.$$

Fig. 4 illustrates the Bode diagram of the original and reduced systems with $r = 5$. Because the original system is a complicated system, the BT method provided the reduced system that is far from the original. In the other methods, it was seemed that our method using BT initialization was better than the results of the method using random initialization and the method proposed in [13] in terms of the magnitude, although the results of all methods were not satisfactory near the original in terms of the phase.

D. DISCUSSIONS

We can see from Table 3 that the BT method may provide locally optimal solutions to Problem 2 if the reduced model dimension is sufficiently large, because $||\mathrm{grad} f(J_r, R_r, B_r, C_r)||_{(17)}$ of BT is sufficiently close to zero when $r = 30$. However, from the viewpoint of controller design, it is preferable for the dimension of the state of a plant to be as small as possible. Thus, from the results presented in Sections V-A, V-B, and V-C, we can conclude that our proposed method using BT initialization will be useful for improving the results obtained by the BT method to reduce the original asymptotically stable linear system to a small-dimensional system in terms of the H^2 and H^∞ norms. Furthermore, the results showed that our method using BT initialization is better than our method using random initialization and the method proposed in [13] in terms of the H^2 and H^∞ norms.

Figs. 2, 3, and 4 showed that our method using BT initialization improved the magnitude characteristics of the BT method. In contrast, Fig. 3 illustrated that our method using random initialization and the method proposed in [13] are not always better than that of the BT method in terms of the frequency response. Thus, even if the results of our method using BT and random initializations are almost the same in terms of the H^2 and H^∞ norms (such as the case presented in Section V-B), our method performs significantly better when using BT initialization than it does when using random initialization. In fact, although we made a number of random systems in Section V-C and compared our method using BT and random initializations, our method using BT initialization always performed better in terms of the frequency response. That is, we can also conclude that our method using BT initialization performs better than our method using random initialization, BT, and the method proposed in [13] in terms of the frequency response.

VI. CONCLUSION

We have proposed a Riemannian optimal model reduction method for asymptotically stable linear systems. The model reduction problem was formulated as a minimization problem of the H^2 error norm between the transfer functions of the original and reduced systems on the product manifold of the set of skew-symmetric matrices, the manifold of the symmetric positive-definite matrices, and two Euclidean spaces. The asymptotic stability of the reduced systems constructed using the optimal solutions to our problem is preserved. Moreover, we proposed that the initial point in our algorithm should be the output of the BT method, because BT produces satisfactory reduced models and is easily implemented in MATLAB. Numerical experiments demonstrated that, in terms of the H^2 norm, our method is better than the BT and other methods. Furthermore, we illustrated that our method provides globally near-optimal solutions to the minimization problem of the H^∞ error norm. Moreover, Bode diagrams showed that our method is better than the BT and other methods.

A. PROOF OF (3)

For convenience, we prove (3), although a similar discussion can be found in [25].

Because systems (1) and (2) are both asymptotically stable, they are L^2-stable. That is, $u \in L^2(\mathbb{R}^m)$ implies that $y, \hat{y}_r \in L^2(\mathbb{R}^p)$, and thus, there exist Fourier transformations U, Y, and $\hat{Y}_r$ of u, y, and $\hat{y}_r$, respectively. Hence, we have that

$$
\begin{aligned}
&||y - \hat{y}_r||_{L^\infty} \\
&= \sup_{t\geq 0} ||y(t) - \hat{y}_r(t)|| \\
&= \sup_{t\geq 0} ||\frac{1}{2\pi}\int_{-\infty}^{\infty} (Y(i\omega) - \hat{Y}_r(i\omega))e^{i\omega t} d\omega|| \\
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty} ||Y(i\omega) - \hat{Y}_r(i\omega)|| d\omega \\
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty} ||G(i\omega) - \hat{G}_r(i\omega)|| \cdot ||U(i\omega)|| d\omega \\
&\leq \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} ||G(i\omega) - \hat{G}_r(i\omega)||^2 d\omega}\sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} ||U(i\omega)||^2 d\omega} \\
&\leq ||G - \hat{G}_r||_{H^2} \cdot ||u||_{L^2},
\end{aligned}
$$

where the second equality follows from the inverse Fourier transformations of Y and $\hat{Y}_r$, the fifth inequality is from the Cauchy–Schwarz inequality, and the final inequality follows from $||G(i\omega) - \hat{G}_r(i\omega)|| \leq ||G(i\omega) - \hat{G}_r(i\omega)||_F$ and Parseval's theorem. This completes the proof.

B. GEOMETRY OF THE MANIFOLD $\mathrm{Sym}_+(r)$

We review the geometry of $\mathrm{Sym}_+(r)$ to develop an optimization algorithm for solving Problem 1. For a detailed explanation, see [23].

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

$$\langle \xi_1, \xi_2 \rangle_S := \mathrm{tr}(S^{-1}\xi_1 S^{-1}\xi_2). \tag{21}$$

Let $g : \mathrm{Sym}_+(r) \to \mathbb{R}$ be a smooth function and $\bar{g}$ be the extension of g to Euclidean space $\mathbb{R}^{r\times r}$. Riemannian gradient $\mathrm{grad}\, g(S)$ with respect to Riemannian metric (21) is given by

$$\mathrm{grad}\, g(S) = S\mathrm{sym}(\nabla \bar{g}(S))S, \tag{22}$$

where $\nabla \bar{g}(S)$ denotes the Euclidean gradient of $\bar{g}$ at $S \in \mathrm{Sym}_+(r)$. Riemannian Hessian $\mathrm{Hess}\, g(S) : T_S\mathrm{Sym}_+(r) \to T_S\mathrm{Sym}_+(r)$ of function g at $S \in \mathrm{Sym}_+(r)$ is given by

$$
\begin{aligned}
\mathrm{Hess}\, g(S)[\xi] = &S\mathrm{sym}(\mathrm{D}\nabla \bar{g}(S)[\xi])S \\
&+\mathrm{sym}(\xi\,\mathrm{sym}(\nabla \bar{g}(S))S).
\end{aligned} \tag{23}
$$

The exponential map on $\mathrm{Sym}_+(r)$ is given by

$$
\begin{aligned}
\mathrm{Exp}_S(\xi) &= S^{\frac{1}{2}} \exp(S^{-\frac{1}{2}}\xi S^{-\frac{1}{2}})S^{\frac{1}{2}} \\
&= S\exp(S^{-1}\xi),
\end{aligned} \tag{24}
$$

where exp is the matrix exponential function.

C. LIMITATION OF MODEL REDUCTION IN TERMS OF THE HANKEL SINGULAR VALUES

We review the relation between a lower bound of $||G - G_r||_{H^\infty}$ and the Hankel singular values of G [3].

Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq \sigma_{r+1} \geq \cdots \geq \sigma_n > 0$ be the Hankel singular values of the transfer function G. Here, the Hankel singular values are the square roots of the eigenvalues of $\Sigma_o\Sigma_c$, where Σ_c and Σ_o denote the controllability and observability gramians satisfying (10) and (11), respectively. Then, for any reduced transfer function G_r of order r, we have that

$$||G - G_r||_{H^\infty} \geq \sigma_{r+1}.$$

That is, all approximation errors are not smaller than the $(r+1)$-th Hankel singular value of G in terms of the H^∞ norm.

D. REDUCED MATRICES (J_r, R_r, B_r, C_r) IN THE CASE WHERE $r = 4$ IN SECTION V-A

We present the reduced matrices (J_r, R_r, B_r, C_r) produced by our proposed method using BT initialization in the case where $r = 4$ in Section V-A. Let $J_r = \big((J_r)^1\ (J_r)^2\big)$, $R_r = \big((R_r)^1\ (R_r)^2\big)$, $C_r = \big((C_r)^1\ (C_r)^2\big)$. Then, we obtained

$$
(J_r)^1 = \begin{pmatrix} 0.000000000000000 & -0.049530743507566 \\ 0.049530743507566 & 0.000000000000000 \\ -0.018625039127746 & 0.626524211054092 \\ 0.007106890495913 & -1.083765311671058 \end{pmatrix},
$$

$$
(J_r)^2 = \begin{pmatrix} 0.018625039127746 & -0.007106890495913 \\ -0.626524211054092 & 1.083765311671058 \\ 0.000000000000000 & 0.066881602488369 \\ -0.066881602488369 & 0.000000000000000 \end{pmatrix},
$$

$$
(R_r)^1 = \begin{pmatrix} 0.020979798103068 & 0.008729495305520 \\ 0.008729495305520 & 0.296162218193050 \\ -0.026753473825891 & 0.016509857981159 \\ -0.003019900398660 & -0.169695898367632 \end{pmatrix},
$$

$$
(R_r)^2 = \begin{pmatrix} -0.026753473825891 & -0.003019900398660 \\ 0.016509857981159 & -0.169695898367632 \\ 0.277287705425208 & -0.447429037737505 \\ -0.447429037737505 & 1.303620534440710 \end{pmatrix},
$$

$$
B_r = \begin{pmatrix} 1.087281955207546 & 1.075128712585373 \\ 0.019632883027025 & -0.081897882654859 \\ -0.060704161404099 & -0.031902870273656 \\ 0.013609328117831 & -0.011572768539278 \end{pmatrix},
$$

$$
(C_r)^1 = \begin{pmatrix} 0.079020553332377 & 0.648595865888539 \end{pmatrix},
$$

$$
(C_r)^2 = \begin{pmatrix} 0.877453660076422 & -3.055799879863735 \end{pmatrix}.
$$

REFERENCES

[1] G. Obinata and B. D. O. Anderson, *Model Reduction for Control System Design*. London, U.K.: Springer-Verlag, 2001.

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

[3] G. Dullerud and F. Paganini, *A Course in Robust Control Theory*. New York, NY, USA: Springer-Verlag, 2000.

[4] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. AC-26, no. 1, pp. 17–32, Feb. 1981.

[5] A. C. Antoulas, C. A. Beattie, and S. Gugercin, "Interpolatory model reduction of large-scale dynamical systems," in *Efficient Modeling and Control of Large-Scale Systems*. New York, NY, USA: Springer-Verlag, 2010, pp. 3–58.
[6] A. Astolfi, "Model reduction by moment matching for linear and nonlinear systems," *IEEE Trans. Autom. Control*, vol. 55, no. 10, pp. 2321–2336, Oct. 2010.
[7] S. Gugercin, A. C. Antoulas, and C. Beattie, "H_2 model reduction for large-scale linear dynamical systems," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 609–638, Jun. 2008.
[8] T. C. Ionescu, A. Astolfi, and P. Colaneri, "Families of moment matching based, low order approximations for linear systems," *Syst. Control Lett.*, vol. 64, pp. 47–56, Feb. 2014.
[9] T. Ishizaki, K. Kashima, A. Girard, J. I. Imura, L. Chen, and K. Aihara, "Clustered model reduction of positive directed networks," *Automatica*, vol. 59, pp. 238–247, Sep. 2015.
[10] N. Monshizadeh, H. L. Trentelman, and M. K. Camlibel, "Projection-based model reduction of multi-agent systems using graph partitions," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 2, pp. 145–154, Jun. 2014.
[11] N. Monshizadeh, C. De Persis, A. J. van der Schaft, and J. M. A. Scherpen, "A novel reduced model for electrical networks with constant power loads," *IEEE Trans. Autom. Control*, vol. 63, no. 5, pp. 1288–1299, May 2017.
[12] H. Sandberg and R. M. Murray, "Model reduction of interconnected linear systems," *Optim. Control Appl. Methods*, vol. 30, no. 3, pp. 225–245, May/Jun. 2009.
[13] H. Sato and K. Sato, "Riemannian trust-region methods for H^2 optimal model reduction," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, Dec. 2015, pp. 4648–4655.
[14] W.-Y. Yan and J. Lam, "An approximate approach to H^2 optimal model reduction," *IEEE Trans. Autom. Control*, vol. 44, no. 7, pp. 1341–1358, Jul. 1999.
[15] F.-X. Orbandexivry, Y. Nesterov, and P. Van Dooren, "Nearest stable system using successive convex approximations," *Automatica*, vol. 49, no. 5, pp. 1195–1203, May 2013.
[16] S. Prajna, A. van der Schaft, and G. Meinsma, "An LMI approach to stabilization of linear port-controlled Hamiltonian systems," *Syst. Control Lett.*, vol. 45, no. 5, pp. 371–385, Apr. 2002.
[17] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation AX+ XB=C [F4]," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, Sep. 1972.
[18] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2012.
[19] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2008.
[20] P. Van Dooren, K. A. Gallivan, and P.-A. Absil, "H_2-optimal model reduction of MIMO systems," *Appl. Math. Lett.*, vol. 21, no. 12, pp. 1267–1273, Dec. 2008.
[21] G. Flagg, C. Beattie, and S. Gugercin, "Convergence of the iterative rational Krylov algorithm," *Syst. Control Lett.*, vol. 61, no. 6, pp. 688–691, Jun. 2012.
[22] K. Sato, "Riemannian optimal model reduction of linear port-Hamiltonian systems," *Automatica*, vol. 98, pp. 428–434, Jul. 2018.
[23] K. Sato and H. Sato, "Structure-preserving H^2 optimal model reduction based on the Riemannian trust-region method," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 505–511, Feb. 2018.
[24] H. Sato and K. Sato, "A new H^2 optimal model reduction method based on Riemannian conjugate gradient method," in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, Dec. 2016, pp. 5762–5768.
[25] K. Sato, "Riemannian optimal control and model matching of linear port-Hamiltonian systems," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6575–6581, Dec. 2017.
[26] K. Sato, "Riemannian optimal model reduction of linear second-order systems," *IEEE Control Syst. Lett.*, vol. 1, no. 1, pp. 2–7, Jul. 2017.
[27] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a MATLAB toolbox for optimization on manifolds," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1455–1459, Jan. 2014.
[28] Y. Chahlaoui and P. Van Dooren, "Benchmark examples for model reduction of linear time invariant dynamical systems," in *Dimension Reduction of Large-Scale Systems* (Lecture Notes in Computational Science and Engineering), vol. 45. Berlin, Germany: Springer-Verlag, 2005, pp. 381–395.

KAZUHIRO SATO was born in Hokkaido, Japan, in 1985. He received the B.S., M.S., and Ph.D. degrees from Kyoto University, Japan, in 2009, 2011, and 2014, respectively. From 2014 to 2017, he was a Postdoctoral Fellow with Kyoto University. He is currently an Assistant Professor with the Kitami Institute of Technology. His research interests include mathematical control theory and optimization. He is a member of SICE and IEEE.