2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

$\mathcal{H}_2$ Pseudo-Optimality in Model Order Reduction by Krylov Subspace Methods

Thomas Wolf, Heiko K. F. Panzer and Boris Lohmann

Abstract— **New sufficient conditions for $\mathcal{H}_2$ pseudo-optimality in Krylov-based model reduction of linear dynamical systems are presented. The conditions are easy to evaluate and permit first applications: a new algorithm to generate $\mathcal{H}_2$ pseudo-optimal reduced models with respect to the projecting subspace and a procedure to generate superior local optima in iterative methods. Numerical examples illustrate the contributions.**

I. INTRODUCTION

Model order reduction (MOR) aims at the approximation of large-scale dynamical systems by another model of reduced order. For the reduction of linear time invariant (LTI) systems different approaches have been shown to be well-suited, such as the *Krylov subspace methods*, [1], [9], or the *Truncated Balanced Realization* (TBR), [1].

The (relative) approximation error in the dynamics of the reduced system is commonly measured in the $\mathcal{H}_2$ or $\mathcal{H}_\infty$ norm. Recently, in [10], [11] algorithms for computing $\mathcal{H}_2$ (pseudo-)optimal reduced systems were proposed. However, these methods are of iterative nature, and convergence can be proven only for particular dynamical systems [6].

A necessary condition for a (local) $\mathcal{H}_2$ optimality is so-called $\mathcal{H}_2$ *pseudo-optimality*. It turns out that—although no analytic solution to $\mathcal{H}_2$ optimality is available—at least the analytic solution to $\mathcal{H}_2$ pseudo-optimality can be found.

In this work, $\mathcal{H}_2$ pseudo-optimality is investigated based on the reduction by Krylov subspaces. The theoretical insight leads to a new algorithm to enforce $\mathcal{H}_2$ pseudo-optimality in the reduced system, with the advantage over existing approaches, that the connection to the projecting Krylov subspace is maintained. This can be of benefit in certain cases. Furthermore, a restart strategy for the iterative methods in [10], [11] is derived from the proposed sufficient conditions.

In Section II preliminaries are given. The main theorem, which introduces new sufficient conditions for $\mathcal{H}_2$ pseudo-optimality, can be found in Section III. Section IV discusses the potential of the theoretical results, together with possible applications. Numerical examples are given in Section V.

II. PRELIMINARIES

We consider LTI systems of the form

$$\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t), \qquad y(t) = \mathbf{c}\mathbf{x}(t), \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^N$, $u(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$ denote the states, input and output of the system, respectively. The dynamics of the system are described by $\mathbf{E}, \mathbf{A} \in \mathbb{R}^{N\times N}$ and $\mathbf{b} \in \mathbb{R}^N$, $\mathbf{c} \in \mathbb{R}^{1\times N}$. It is assumed that $\mathbf{E}$ is non-singular: $\det(\mathbf{E}) \neq 0$. With the usual abuse of notation, let $G(s)$ denote the transfer function of system (1) in the Laplace domain as well as the dynamical system itself.

The goal of model order reduction is to approximate a large-scale system (1), e. g. with $N = 10^2 \dots 10^7$, by one of much smaller dimension $n \ll N$. In projection based model reduction, this is carried out by appropriate projection matrices $\mathbf{V}, \mathbf{W} \in \mathbb{R}^{N\times n}$, leading to the reduced dynamics $G_r(s)$ with the state-space representation

$$\mathbf{E}_r\dot{\mathbf{x}}_r(t) = \mathbf{A}_r\mathbf{x}_r(t) + \mathbf{b}_r u(t), \qquad y_r(t) = \mathbf{c}_r\mathbf{x}_r(t), \tag{2}$$

where $\mathbf{A}_r = \mathbf{W}^T\mathbf{A}\mathbf{V}$, $\mathbf{E}_r = \mathbf{W}^T\mathbf{E}\mathbf{V}$, $\mathbf{b}_r = \mathbf{W}^T\mathbf{b}$, and $\mathbf{c}_r = \mathbf{c}\mathbf{V}$. Note that the reduced system can be interpreted as a projection of (1) by the projector $\boldsymbol{\mathcal{P}} = \mathbf{E}\mathbf{V}\left(\mathbf{W}^T\mathbf{E}\mathbf{V}\right)^{-1}\mathbf{W}^T$, mapping onto span($\mathbf{E}\mathbf{V}$) orthogonally to span($\mathbf{W}$). The reduced system (2) is called a *Petrov-Galerkin* projection of (1), [9]. In this paper the projection matrix $\mathbf{V}$ (or $\mathbf{W}$) is chosen to span certain rational Krylov subspaces.

A. Rational Krylov Subspaces

A Krylov subspace is generally defined as

$$\mathcal{K}^k(\mathbf{A}, \mathbf{b}) = \text{span}\left\{\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{k-1}\mathbf{b}\right\}. \tag{3}$$

By selecting a complex valued expansion point $s_i \in \mathbb{C}$ and a desired multiplicity $m_i \in \mathbb{N}^+$ the rational input Krylov subspace reads as

$$\mathcal{K}_{s_i}^{m_i} := \mathcal{K}^{m_i}\left((\mathbf{A} - s_i\mathbf{E})^{-1}\mathbf{E}, (\mathbf{A} - s_i\mathbf{E})^{-1}\mathbf{b}\right). \tag{4}$$

If the projection matrix $\mathbf{V}$ then is computed to span the union of certain Krylov subspaces

$$\text{span}(\mathbf{V}) \supseteq \mathcal{K}_{s_i}^{m_i}, \qquad i = 1, \dots, q \tag{5}$$

the reduced model (2) matches m_i moments around the respective expansion points s_i, if s_i is neither an eigenvalue of (1) nor an eigenvalue of (2), [1], [7], [9]. Moments are defined as the coefficients of the Taylor series expansion of the transfer function around s_i. As the rational transfer function $G_r(s)$ interpolates $G(s)$ at the given frequencies s_i, this method for model reduction is referred to as *Rational Interpolation* or *Rational Krylov* (RK).

B. Equivalence to Sylvester Equations

It was first shown in [8], and related to other results in [15], that any basis of an arbitrary input rational Krylov subspace (5) can equivalently be interpreted as the solution $\mathbf{V}$ of a Sylvester equation

$$\mathbf{A}\mathbf{V} - \mathbf{E}\mathbf{V}\mathbf{S} = \mathbf{b}\widehat{\mathbf{c}}_r, \tag{6}$$

All authors are with the Institute of Automatic Control, Technische Universität München, Boltzmannstr. 15, D-85748 Garching, Germany.
Corresponding author's email: `thomas.wolf@tum.de`

978-3-952-41734-8/

where the eigenvalues of $\mathbf{S} \in \mathbb{R}^{n\times n}$ correspond to the expansion points s_i defining the subspace span($\mathbf{V}$). Furthermore, the vector $\widehat{\mathbf{c}}_r \in \mathbb{R}^{1\times n}$ basically contains information on the orthogonalization process that was used to construct $\mathbf{V}$. From (6) a second Sylvester equation can be derived

$$\mathbf{A}\mathbf{V} - \mathbf{E}\mathbf{V}\mathbf{E}_r^{-1}\mathbf{A}_r = \mathbf{b}_\perp \widehat{\mathbf{c}}_r, \tag{7}$$

where $\mathbf{b}_\perp := (\mathbf{I} - \boldsymbol{\mathcal{P}})\mathbf{b} \in \mathbb{R}^N$ closes the vector chain from $\mathbf{b}$ to its projection $\boldsymbol{\mathcal{P}}\mathbf{b}$. The matrix $\mathbf{S}$ and the vectors $\mathbf{b}_\perp$ and $\widehat{\mathbf{c}}_r$ are crucial for the formulation of the results in this paper. The following equations describe one possible way for their computation (for details please refer to [15]):

$$\mathbf{b}_\perp = \mathbf{b} - \mathbf{E}\mathbf{V}\mathbf{E}_r^{-1}\mathbf{b}_r, \tag{8}$$

$$\widehat{\mathbf{c}}_r = (\mathbf{b}_\perp^T\mathbf{b}_\perp)^{-1}\mathbf{b}_\perp^T\left(\mathbf{A}\mathbf{V} - \mathbf{E}\mathbf{V}\mathbf{E}_r^{-1}\mathbf{A}_r\right), \tag{9}$$

$$\mathbf{S} = \mathbf{E}_r^{-1}\left(\mathbf{A}_r - \mathbf{b}_r\widehat{\mathbf{c}}_r\right). \tag{10}$$

C. The Error System

The approximation error in model reduction is defined as the difference in the output of the original and the reduced system, $y_e(t) := y(t) - y_r(t)$, and is described by the transfer function $G_e(s) := G(s) - G_r(s)$. In [15] it was shown, that when the reduced system stems from a Krylov projection, the error system can be factorized as follows:

$$G_e(s) = G_\perp(s)\cdot\widehat{G}_r(s), \tag{11}$$

$$G_\perp(s) := \mathbf{c}\left(s\mathbf{E} - \mathbf{A}\right)^{-1}\mathbf{b}_\perp, \tag{12}$$

$$\widehat{G}_r(s) := \widehat{\mathbf{c}}_r\left(s\mathbf{E}_r - \mathbf{A}_r\right)^{-1}\mathbf{b}_r + 1. \tag{13}$$

Please note, that $G_\perp(s)$ shares $\mathbf{A}, \mathbf{E}$ and $\mathbf{c}$ with the original system and differs from it only in the input $\mathbf{b}_\perp$. In addition, $\widehat{G}_r(s)$ shares $\mathbf{A}_r, \mathbf{E}_r$ and $\mathbf{b}_r$ with the reduced system and differs only in the output $\widehat{\mathbf{c}}_r$ and the feedthrough term.

D. Gramian and $\mathcal{H}_2$ Norm

The solutions of the generalized Lyapunov equations

$$\mathbf{A}\mathbf{P}\mathbf{E}^T + \mathbf{E}\mathbf{P}\mathbf{A}^T + \mathbf{b}\mathbf{b}^T = \mathbf{0}, \tag{14}$$

$$\mathbf{A}^T\mathbf{Q}\mathbf{E} + \mathbf{E}^T\mathbf{Q}\mathbf{A} + \mathbf{c}^T\mathbf{c} = \mathbf{0}, \tag{15}$$

define the *Controllability Gramian* $\mathbf{P}$ and the *Observability Gramian* $\mathbf{E}^T\mathbf{Q}\mathbf{E}$ of $G(s)$, which play an important role in the analysis and reduction of dynamical systems, [1]. We will use the Gramians to compute the $\mathcal{H}_2$ norm, defined as

$$\|G\|_{\mathcal{H}_2}^2 := \frac{1}{2\pi}\int_{-\infty}^{+\infty}|G(j\omega)|^2 d\omega. \tag{16}$$

The $\mathcal{H}_2$ norm of a system can equivalently be formulated with respect to $\mathbf{P}$ and $\mathbf{Q}$ by $\|G\|_{\mathcal{H}_2}^2 = \mathbf{b}^T\mathbf{Q}\mathbf{b} = \mathbf{c}\mathbf{P}\mathbf{c}^T$, [5]. Applying this formulation to the error system $G_e(s)$, leads to the following statement for the approximation error, [1],

$$\|G_e\|_{\mathcal{H}_2}^2 = \mathbf{b}^T\mathbf{Q}\mathbf{b} + 2\mathbf{b}^T\mathbf{Y}\mathbf{b}_r + \mathbf{b}_r^T\mathbf{Q}_r\mathbf{b}_r \tag{17}$$

$$= \mathbf{c}\mathbf{P}\mathbf{c}^T - 2\mathbf{c}\mathbf{X}\mathbf{c}_r^T + \mathbf{c}_r\mathbf{P}_r\mathbf{c}_r^T, \tag{18}$$

where $\mathbf{X}$ and $\mathbf{Y}$ solve the Sylvester equations

$$\mathbf{A}\mathbf{X}\mathbf{E}_r^T + \mathbf{E}\mathbf{X}\mathbf{A}_r^T + \mathbf{b}\mathbf{b}_r^T = \mathbf{0}, \tag{19}$$

$$\mathbf{A}^T\mathbf{Y}\mathbf{E}_r + \mathbf{E}^T\mathbf{Y}\mathbf{A}_r - \mathbf{c}^T\mathbf{c}_r = \mathbf{0}, \tag{20}$$

and $\mathbf{P}_r$ and $\mathbf{E}_r^T\mathbf{Q}_r\mathbf{E}_r$ denote the reduced Gramians, i. e. $\mathbf{P}_r$ and $\mathbf{Q}_r$ solve the reduced Lyapunov equations

$$\mathbf{A}_r\mathbf{P}_r\mathbf{E}_r^T + \mathbf{E}_r\mathbf{P}_r\mathbf{A}_r^T + \mathbf{b}_r\mathbf{b}_r^T = \mathbf{0}, \tag{21}$$

$$\mathbf{A}_r^T\mathbf{Q}_r\mathbf{E}_r + \mathbf{E}_r^T\mathbf{Q}_r\mathbf{A}_r + \mathbf{c}_r^T\mathbf{c}_r = \mathbf{0}. \tag{22}$$

E. $\mathcal{H}_2$ (Pseudo-)Optimality

$\mathcal{H}_2$ optimal model reduction is the search for the reduced system $G_r(s)$ of fixed order n, such that

$$\|G - G_r\|_{\mathcal{H}_2} = \min_{\dim(G_r^*)=n}\|G - G_r^*\|_{\mathcal{H}_2} \tag{23}$$

In [10] it was shown, that a necessary first-order condition to (23) is that $G_r(s)$ interpolates $G(s)$ and its derivative $G'(s)$ at the mirror images of the reduced poles. Let λ_i be the poles of $G_r(s)$, then this reads as

$$G(-\lambda_i) = G_r(-\lambda_i), \quad i = 1,\ldots,n \tag{24}$$

$$G'(-\lambda_i) = G_r'(-\lambda_i), \quad i = 1,\ldots,n. \tag{25}$$

In other words, $G_r(s)$ matches two moments at the mirror images of its poles λ_i. However, the set $\{\lambda_i\} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \subset \mathbb{C}$ of reduced poles, that can fulfill both (24) and (25), is not known a priori, which is why only iterative methods can be stated that hopefully converge to such a set. Nevertheless, the *Iterative Rational Krylov Algorithm* (IRKA), presented in [10], shows good convergence behavior and leads to locally optimal reduced models in the $\mathcal{H}_2$ norm.

Indeed, it is possible to enforce merely (24) and disregard (25) for arbitrary sets $\{\lambda_i\}$ of asymptotically stable reduced poles, which induces the concept of pseudo-optimality.

Definition 1. Let $\{\lambda_i\} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \subset \mathbb{C}$ be a given set of asymptotically stable reduced poles and define the set $\mathcal{T}(\lambda)$ of all transfer functions of order n sharing the pole-configuration $\{\lambda_i\}$. Then, the pseudo-optimal reduced system $G_r(s)$ solves the following minimization problem

$$\|G - G_r\|_{\mathcal{H}_2} = \min_{G_r^* \in \mathcal{T}(\lambda)}\|G - G_r^*\|_{\mathcal{H}_2} \tag{26}$$

Theorem 1 ([11])**.** *The reduced system $G_r(s)$ is pseudo-optimal, i. e. it minimizes* (26) *in the set of all reduced systems sharing the pole-configuration $\{\lambda_i\}$, if and only if*

$$G(-\lambda_i) = G_r(-\lambda_i), \quad i = 1,\ldots,n. \tag{27}$$

Therefore, pseudo-optimality is a necessary condition for an $\mathcal{H}_2$ optimal reduced system (23). In [11], Theorem 1 was used to propose an *Iterative SVD-Rational Krylov Based Method* (ISRK), that combines the IRKA algorithm with guaranteed preservation of stability.

F. The Problem Formulation

The starting point for this work is to note, that a sufficient condition for $\mathcal{H}_2$ pseudo-optimality can be stated as

$$\Lambda(\mathbf{S}) = -\Lambda(\mathbf{A}_r, \mathbf{E}_r), \tag{28}$$

where $\Lambda(\mathbf{S})$ denotes the set of eigenvalues of $\mathbf{S}$ from (6) and $\Lambda(\mathbf{A}_r, \mathbf{E}_r)$ denotes the set of eigenvalues $\{\lambda_i\}$ of the generalized eigenvalue problem: $\det(\mathbf{A}_r - \lambda_i\mathbf{E}_r) = 0$. Please note, that the set $\Lambda(\mathbf{S})$ corresponds to the set of expansion points that was used to construct the projection matrix $\mathbf{V}$, see [8], [15] for details. Therefore, by moment matching, (28) guarantees that condition (27) is fulfilled.

Remark. Condition (28) is sufficient but not necessary for (27) to hold. However, since $G_r(s)$ is unique with respect to the transfer behavior, it is not restrictive to concentrate on (28) to find the solution to (27). (The uniqueness is due to the fact that n expansion points for moment matching and n reduced poles are fixed in the reduced system.)

The contribution of this work is to analyze (pseudo-) optimality based on the matrix formulation (28) rather than on the interpolation formulation (27). Towards this aim, five new and easy-to-evaluate conditions are shown to be equivalent to (28). Furthermore, first ideas how to benefit from these new conditions are presented in Section IV.

III. MAIN RESULTS

We are now ready to state the main theorem.

Theorem 2. *Let* $\mathbf{V}$ *span a rational input Krylov subspace, i. e.* (5) – (10) *hold, and let* $\{\lambda_i\} = \{\lambda_1, \lambda_2, \dots, \lambda_n\} \subset \mathbb{C}$ *be a given set of asymptotically stable reduced poles,* $\{\lambda_i\} = \Lambda(\mathbf{A}_r, \mathbf{E}_r)$*, that is closed under conjugation. Then, the following statements are equivalent:*

i) $\Lambda(\mathbf{S}) = -\Lambda(\mathbf{A}_r, \mathbf{E}_r)$
ii) $\mathbf{S} = -\mathbf{P}_r\mathbf{A}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1}$
iii) $\mathbf{E}_r^{-1}\mathbf{b}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T = \mathbf{0}$
iv) $\mathbf{S}\mathbf{P}_r + \mathbf{P}_r\mathbf{S}^T - \mathbf{P}_r\widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r\mathbf{P}_r = \mathbf{0}$
v) $\mathbf{X} = \mathbf{V}\mathbf{P}_r$
vi) $\mathbf{P}_r^{-1} = \mathbf{E}_r^T\widehat{\mathbf{Q}}_r\mathbf{E}_r$

where $\mathbf{E}_r^T\widehat{\mathbf{Q}}_r\mathbf{E}_r$ *defines the Observability Gramian of the system* $\widehat{G}_r(s)$*, i. e.* $\widehat{\mathbf{Q}}_r$ *solves the reduced Lyapunov equation*

$$\mathbf{A}_r^T\widehat{\mathbf{Q}}_r\mathbf{E}_r + \mathbf{E}_r^T\widehat{\mathbf{Q}}_r\mathbf{A}_r + \widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r = \mathbf{0}. \tag{29}$$

The proof is given in the appendix. Please note, that Theorem 2 only requires a rational Krylov subspace $\mathbf{V}$, whereas $\mathbf{W}$ can be arbitrary (as long as $\mathbf{A}_r$, $\mathbf{E}_r$ are nonsingular). For the sake of convenience we assumed $\{\lambda_i\}$ to be closed under conjugation, because then a real basis $\mathbf{V}$ can be found, leading to exclusively real matrices in the theorem.

Naturally, dual conditions to Theorem 2 can be stated if $\mathbf{W}$ spans an output rational Krylov subspace and $\mathbf{V}$ is arbitrary (as long as $\mathbf{A}_r$, $\mathbf{E}_r$ are nonsingular). In this regard, the dual Sylvester equations read as [15]

$$\mathbf{A}^T\mathbf{W} - \mathbf{E}^T\mathbf{W}\mathbf{S}_W^T = \mathbf{c}^T\widehat{\mathbf{b}}_r, \tag{30}$$

$$\mathbf{A}^T\mathbf{W} - \mathbf{E}^T\mathbf{W}\mathbf{E}_r^{-T}\mathbf{A}_r^T = \mathbf{c}_\perp^T\widehat{\mathbf{b}}_r, \tag{31}$$

where $\mathbf{c}_\perp$, $\widehat{\mathbf{b}}_r$ and $\mathbf{S}_W$ can be computed by

$$\mathbf{c}_\perp^T = \mathbf{c}^T - \mathbf{E}^T\mathbf{W}\mathbf{E}_r^{-T}\mathbf{c}_r^T, \tag{32}$$

$$\widehat{\mathbf{b}}_r = (\mathbf{c}_\perp\mathbf{c}_\perp^T)^{-1}\mathbf{c}_\perp\left(\mathbf{A}^T\mathbf{W} - \mathbf{E}^T\mathbf{W}\mathbf{E}_r^{-T}\mathbf{A}_r^T\right), \tag{33}$$

$$\mathbf{S}_W^T = \mathbf{E}_r^{-T}\left(\mathbf{A}_r^T - \mathbf{c}_r^T\widehat{\mathbf{b}}_r^T\right). \tag{34}$$

With these definitions, the dual conditions for $\mathcal{H}_2$ pseudo-optimality can be stated.

Theorem 3. *Let* $\mathbf{W}$ *span a rational output Krylov subspace, i. e.* (30) – (34) *hold, and let* $\{\lambda_i\} = \{\lambda_1, \lambda_2, \dots, \lambda_n\} \subset \mathbb{C}$ *be a given set of asymptotically stable reduced poles,* $\{\lambda_i\} = \Lambda(\mathbf{A}_r, \mathbf{E}_r)$*, that is closed under conjugation. Then, the following statements are equivalent:*

i) $\Lambda(\mathbf{S}_W) = -\Lambda(\mathbf{A}_r, \mathbf{E}_r)$
ii) $\mathbf{S}_W^T = -\mathbf{Q}_r\mathbf{A}_r\mathbf{E}_r^{-1}\mathbf{Q}_r^{-1}$
iii) $\mathbf{E}_r^{-T}\mathbf{c}_r^T + \mathbf{Q}_r\widehat{\mathbf{b}}_r = \mathbf{0}$
iv) $\mathbf{S}_W^T\mathbf{Q}_r + \mathbf{Q}_r\mathbf{S}_W - \mathbf{Q}_r\widehat{\mathbf{b}}_r\widehat{\mathbf{b}}_r^T\mathbf{Q}_r = \mathbf{0}$
v) $\mathbf{Y} = -\mathbf{W}\mathbf{Q}_r$
vi) $\mathbf{Q}_r^{-1} = \mathbf{E}_r\widehat{\mathbf{P}}_r\mathbf{E}_r^T$

where $\widehat{\mathbf{P}}_r$ *solves the reduced Lyapunov equation*

$$\mathbf{A}_r\widehat{\mathbf{P}}_r\mathbf{E}_r^T + \mathbf{E}_r\widehat{\mathbf{P}}_r\mathbf{A}_r^T + \widehat{\mathbf{b}}_r\widehat{\mathbf{b}}_r^T = \mathbf{0}. \tag{35}$$

Proof. The proof is dual to the one of Theorem 2, hence omitted. □

Corollary 1. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *span input and output rational Krylov subspaces, respectively. If both the conditions from Theorem 2 and from Theorem 3 are fulfilled, then the first-order necessary conditions for* $\mathcal{H}_2$ *optimality are satisfied.*

Proof. Due to condition *i)* of Theorem 2 and Theorem 3, the expansion points used to construct $\mathbf{V}$ and $\mathbf{W}$ are equal and therefore, two moments are matched at the mirror images of the reduced poles—which is equivalent to the first-order necessary conditions for $\mathcal{H}_2$ optimality (24) and (25). □

Remark. Applying Theorem 2 to the Sylvester equations (6) and (7) leads to the $\mathcal{H}_2$ pseudo-optimal Sylvester equations

$$\mathbf{A}\mathbf{V}\mathbf{P}_r\mathbf{E}_r^T + \mathbf{E}\mathbf{V}\mathbf{P}_r\mathbf{A}_r^T = -\mathbf{b}\mathbf{b}_r^T, \tag{36}$$

$$\mathbf{A}\mathbf{V}\mathbf{P}_r\mathbf{E}_r^T + \mathbf{E}\mathbf{V}\mathbf{P}_r\mathbf{S}^T\mathbf{E}_r^T = \mathbf{b}_\perp\mathbf{b}_r^T. \tag{37}$$

(Dual Sylvester equations for $\mathbf{W}$ can also be formulated.) Please note, that (36) was already presented in [4] (for the special case $\mathbf{E} = \mathbf{I}$), using a completely different proof.

IV. APPLICATIONS

Although Theorems 2 and 3 reveal new insight on $\mathcal{H}_2$ pseudo-optimality, the benefit of the new conditions remains open; first applications are still presented in the following.

A. $\mathcal{H}_2$ Pseudo-Optimal Reduction

Based on Theorem 2, a new iteration-free algorithm can be designed for the *Pseudo-Optimal Reduction by Krylov* (PORK) with a given set $\{s_i\}$, see Algorithm 1.

Indeed, the outcome of PORK could also have been computed in a more straightforward fashion, as sketched in the following: the goal is to find a reduced model of order n which not only has poles at a given set $\{\lambda_i\}$, but also interpolates the original model at $\{-\lambda_i\}$. Now assume $\mathbf{E}_r\dot{\mathbf{x}}_r = \mathbf{A}_r\mathbf{x}_r + \mathbf{b}_r u$ to be given, such that $\Lambda(\mathbf{A}_r, \mathbf{E}_r) = \{\lambda_i\}$ and $\mathbf{b}_r$ controllable (e. g. choose $\mathbf{A}_r$ diagonal, $\mathbf{E}_r = \mathbf{I}$ and $\mathbf{b}_r = [1, \dots, 1]^T$). Furthermore, define

$$\mathbf{M} := \Big[(-\lambda_1\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r, \dots, (-\lambda_n\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r\Big] \tag{38}$$

$$\mathbf{N} := [G(-\lambda_1), \dots, G(-\lambda_n)]. \tag{39}$$

Algorithm 1 Pseudo-Optimal Reduction by Krylov (PORK)

```
Input: $\mathbf{E}$, $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$, $\{s_i\}$
Output: $\mathcal{H}_2$ pseudo-optimal reduced system $G_r(s)$
$\mathbf{V} = \text{arnoldi}(\mathbf{E}, \mathbf{A}, \mathbf{b}, \{s_i\})$
    // Algorithm for computing the rational Krylov subspace
Choose $\mathbf{W}$ arbitrary (e.g. $\mathbf{W} = \mathbf{V}$)
$\widetilde{\mathbf{A}}_r = \mathbf{W}^T\mathbf{A}\mathbf{V}$, $\widetilde{\mathbf{E}}_r = \mathbf{W}^T\mathbf{E}\mathbf{V}$, $\widetilde{\mathbf{b}}_r = \mathbf{W}^T\mathbf{b}$
$\mathbf{b}_\perp = \mathbf{b} - \mathbf{E}\mathbf{V}\widetilde{\mathbf{E}}_r^{-1}\widetilde{\mathbf{b}}_r$    // eq. (8)
$\widehat{\mathbf{c}}_r = (\mathbf{b}_\perp^T\mathbf{b}_\perp)^{-1}\mathbf{b}_\perp^T\left(\mathbf{A}\mathbf{V} - \mathbf{E}\mathbf{V}\widetilde{\mathbf{E}}_r^{-1}\widetilde{\mathbf{A}}_r\right)$    // eq. (9)
$\mathbf{S} = \widetilde{\mathbf{E}}_r^{-1}\left(\widetilde{\mathbf{A}}_r - \widetilde{\mathbf{b}}_r\widehat{\mathbf{c}}_r\right)$    // eq. (10)
$\mathbf{P}_r^{-1} = \text{lyap}(-\mathbf{S}^T, \widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r)$    // condition iv)
$\mathbf{b}_r = -(\mathbf{P}_r^{-1})^{-1}\widehat{\mathbf{c}}_r^T$    // condition iii)
$\mathbf{A}_r = \mathbf{S} + \mathbf{b}_r\widehat{\mathbf{c}}_r$, $\mathbf{E}_r = \mathbf{I}$, $\mathbf{c}_r = \mathbf{c}\mathbf{V}$
```

Then, the interpolation condition $G(-\lambda_i) = G_r(-\lambda_i)$, $i = 1, \ldots, n$, can be rewritten as

$$\mathbf{c}_r\mathbf{M} = \mathbf{N}, \tag{40}$$

which is an n-by-n linear system that can be solved for $\mathbf{c}_r$ such that $G_r(s) = \mathbf{c}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r$ is pseudo-optimal.

A possible application of this approach or PORK is the following: assume e. g. that an arbitrary model order reduction (Krylov, TBR, etc.) is performed leading to an intermediate reduced system $G_r^*(s)$. Using the reduced poles of $G_r^*(s)$ in order to compute the pseudo-optimal reduced system $G_r(s)$ must lower the $\mathcal{H}_2$ error due to Theorem 1. In this setting, the numerical efforts for PORK and the simpler approach (40) are equivalent: the main effort in PORK is the computation of the Krylov subspace, as this is the only operation in high dimension N; this however, is comparable to the computation of $\mathbf{N}$. Nevertheless, the approach (40) is more straightforward than PORK and was already generalized to multiple inputs/outputs in [2].

However, the advantage of PORK is that the reduced system stems from a projection onto $\mathbf{V}$, whereas the connection to the Krylov subspace $\mathbf{V}$ is irrecoverable by the approach (40). This can be of benefit as discussed in the following.

The error system can be factorized for arbitrary reduced order models by (11) – (13): $G_e(s) = G_\perp(s) \cdot \widehat{G}_r(s)$. The only prerequisite is that the reduced system stems from a Krylov based projection, which is why the factorization for a pseudo-optimal reduced system is not possible by the simple approach (40). This is due to the fact that the projecting subspace $\mathbf{V}$ is mandatory for computing $\mathbf{b}_\perp$ by (8) in order to perform the factorization. Therefore, only PORK is capable of constructing a pseudo-optimal reduced system for which the error system can be factorized.

This is of particular benefit, because then $\widehat{G}_r(s)$ is all-pass and the remaining error moves to $G_\perp(s)$, [14]. This leads to the idea that in a second step $G_\perp(s)$ can be further reduced, which is pursued by the SPARK algorithm in [12]. There, the reduced models are iteratively constructed and therefore, only converge up to a certain tolerance towards a local $\mathcal{H}_2$ optimum. Forcing pseudo-optimality by PORK instead, can significantly improve the performance of SPARK.

Furthermore, bounds on system norms of $G_\perp(s)$ can be used in order to bound the approximation error in model reduction for a special class of systems, as presented in [13]. There, due to the all-pass nature of $\widehat{G}_r(s)$, a pseudo-optimal reduced system by PORK can lower the overestimation inherent in the bounds. Therefore, PORK is a natural extension to the SPARK algorithm and the bounds in [13], which will be tackled in future work.

B. Restarting IRKA/ISRK

The final local optimum after convergence of IRKA/ISRK heavily depends on the initial selection of expansion points. To our knowledge, only heuristics are available for this choice, [10], and no strategy to guarantee a good local optimum is known. In the following, we present a new approach for this choice. The basic idea is to take a simple set of initial shifts, say random or zero, and perform IRKA/ISRK; after convergence, the reduced system is used to find new initial shifts, in order to restart the algorithm. Towards this aim, the update rule of IRKA, $\{s_i\} \leftarrow -\Lambda(\mathbf{A}_r, \mathbf{E}_r)$, is reformulated in Algorithm 2, based on the conditions of Theorem 2: the algorithm computes the desired $\widehat{\mathbf{c}}_r$ for pseudo-optimality in step 2, which can be interpreted as a "feedback" vector that places the reduced poles at its mirror images in step 3. Therefore, Algorithm 2 is nothing else than the IRKA update $\{s_i\} \leftarrow -\Lambda(\mathbf{A}_r, \mathbf{E}_r)$, but provides a new degree of freedom: the update is carried out by the "feedback" $\widehat{\mathbf{c}}_{r,d}$ that can be arbitrarily scaled; if now $\alpha \cdot \widehat{\mathbf{c}}_{r,d}$ is used once instead of $\widehat{\mathbf{c}}_{r,d}$, the scalar $\alpha \in \mathbb{R}$ is the degree of freedom, that allows to judiciously restart an already converged IRKA by choosing $\alpha \neq 1$. Admittedly, this approach is heuristically motivated, but performs well and can also ensure convergence even if IRKA does not converge in the first place, as will be shown by the numerical examples.

Algorithm 2 Update in IRKA

```
Input: $\mathbf{E}_r$, $\mathbf{A}_r$, $\mathbf{b}_r$, $\mathbf{c}_r$
Output: next shifts $\{s_i\}$
1: $\mathbf{P}_r = \text{lyap}(\mathbf{E}_r^{-1}\mathbf{A}_r, \mathbf{E}_r^{-1}\mathbf{b}_r\mathbf{b}_r^T\mathbf{E}_r^{-T})$
2: $\widehat{\mathbf{c}}_{r,d} = -\mathbf{P}_r^{-1}\mathbf{E}_r^{-1}\mathbf{b}_r$    // desired $\widehat{\mathbf{c}}_r$ by condition iii)
3: $\mathbf{S}_d = \mathbf{E}_r^{-1}(\mathbf{A}_r - \mathbf{b}_r\widehat{\mathbf{c}}_{r,d})$    // desired $\mathbf{S}$ by eq. (10)
4: $\{s_i\} \leftarrow \Lambda(\mathbf{S}_d)$    // update
```

The approach will be compared to another restart strategy, where—in the style of the original IRKA—the reduced eigenvalues are directly scaled by α: $\{s_i\} \leftarrow -\alpha \cdot \Lambda(\mathbf{A}_r, \mathbf{E}_r)$.

C. $\mathcal{H}_2$ Optimal Convergence Criterion

Another application of Theorems 2 and 3 is to track convergence in IRKA/ISRK. Condition *iii)* comprises a vector that is zero in the $\mathcal{H}_2$ (pseudo-)optimum. Therefore, the relative norms $R_{\mathcal{H}_2}^{\mathbf{V}} := \|\mathbf{E}_r^{-1}\mathbf{b}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\|_2 / \|\mathbf{E}_r^{-1}\mathbf{b}_r\|_2$ and $R_{\mathcal{H}_2}^{\mathbf{W}} := \|\mathbf{E}_r^{-T}\mathbf{c}_r^T + \mathbf{Q}_r\widehat{\mathbf{b}}_r\|_2 / \|\mathbf{E}_r^{-T}\mathbf{c}_r^T\|_2$ measure convergence, i. e. indicate the distance to a pseudo-optimum. This allows to divide the information into the input and output side by $R_{\mathcal{H}_2}^{\mathbf{V}}$ and $R_{\mathcal{H}_2}^{\mathbf{W}}$, respectively. Note, that for ISRK only $R_{\mathcal{H}_2}^{\mathbf{V}} = 0$ holds upon convergence, whereas for IRKA $R_{\mathcal{H}_2}^{\mathbf{V}} = R_{\mathcal{H}_2}^{\mathbf{W}} = 0$.

The original convergence criterion is the relative change in the reduced eigenvalues, [10]. However, it is unclear, how the pairing of the k-th with the k–1^{st} eigenvalues should be done. Nevertheless, in most of the numerical tests it was sufficient to use the command `cplxpair` in MATLAB, which is why no significant benefit of $R_{\mathcal{H}_2}^{\mathbf{V}}$ and $R_{\mathcal{H}_2}^{\mathbf{W}}$ was observed.

V. NUMERICAL EXAMPLES

We illustrate the applications of the previous section by two numerical examples from the benchmark collection [3]. The first one is the artificial single-input single-output model FOM of order $N = 1006$. The second example is a structural model of one component of the International Space Station (ISS) of order $N = 270$. We choose the dynamical system from the first input to the first output. For details on the models please refer to the benchmark collection.

In the first scenario we reduce the FOM model by IRKA to the order $n = 8$ with initial expansion points $s_i = 0$, $i = 1, \dots, n$. From iteration 20 on, IRKA alternates between two sets of expansion points (reduced poles). Due to this "limit cycle" behavior, IRKA has to be stopped. A remedy would be to take the poles of one of the reduced systems and compute the corresponding $\mathcal{H}_2$ pseudo-optimal reduced model. Here, the projecting subspace $\mathbf{V}$ of a previous iteration can be recycled, which is why PORK takes in this setting only one fourth of the time compared to the solution of (40). Another remedy is to restart IRKA by the approach from Section IV-B, because taking $2 \cdot \widehat{\mathbf{c}}_{r,d}$ once, leads to convergence. The resulting reduced system is compared to the $\mathcal{H}_2$ pseudo-optimal one in Figure 1. The relative $\mathcal{H}_2$ errors are 0.54 for the pseudo-optimal system and 0.017 for restarted IRKA.

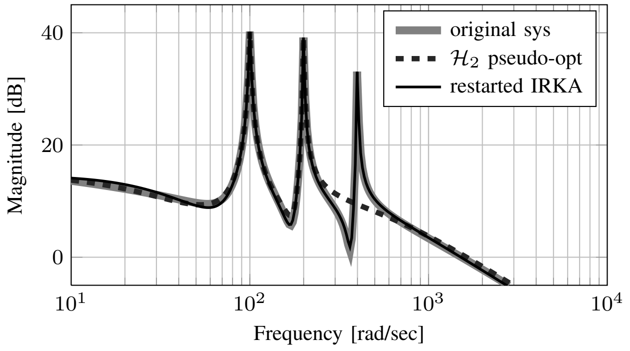

Fig. 1. Magnitude plot, FOM model, $n = 8$

In the second scenario we reduce the ISS model by IRKA to the order $n = 8$ with initial expansion points $s_i = 0$, $i = 1, \dots, n$. Although IRKA converged, restarting it can lead to better reduced systems as shown in Figure 2, where the relative $\mathcal{H}_2$ and $\mathcal{H}_\infty$ errors are plotted after convergence, when IRKA is restarted with different values of α for both strategies form Section IV-B. Note, that the error for IRKA without restart corresponds to the value $\alpha = 1$. Comparing the results reveals that the new restart strategy from Section IV-B is more robust, because, if it converges to a new local optimum, it always finds the best local optimum in this example. In contrast, scaling the reduced eigenvalues for restart causes a rather random behavior, and the best optimum is found only for isolated values of α. This leads to the conclusion, that $\widehat{\mathbf{c}}_{r,d}$ points to a judicious direction, and therefore, it is reasonable to scale this direction. The magnitude plots for the different strategies are compared in Figure 3 for the value $\alpha = 20$.

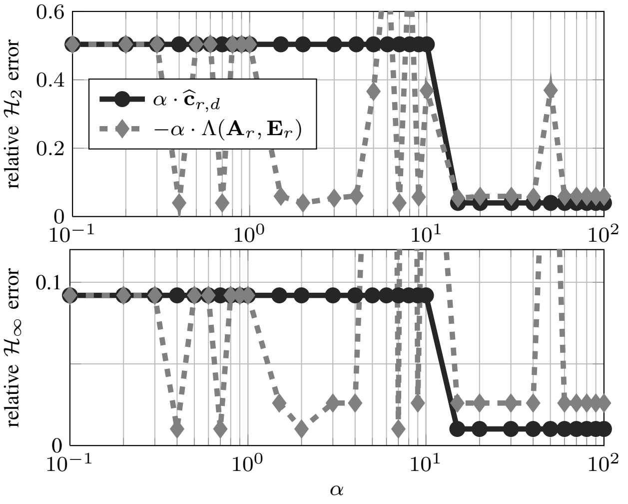

Fig. 2. Relative $\mathcal{H}_2$ and $\mathcal{H}_\infty$ error, ISS model, $n = 8$

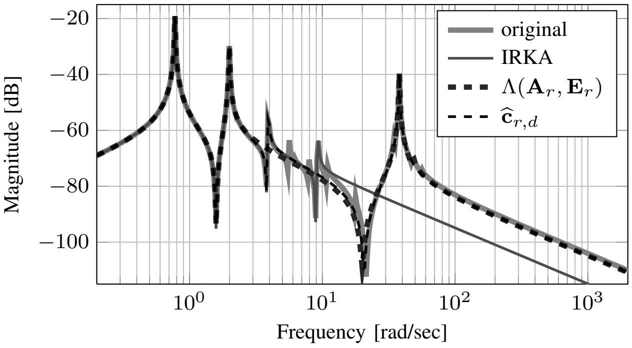

Fig. 3. Magnitude plot, ISS model, $n = 8$

VI. CONCLUSIONS

New conditions for $\mathcal{H}_2$ pseudo-optimal reduction by Krylov subspaces are presented. The conditions give theoretical insight on $\mathcal{H}_2$ optimality, allow to design a new algorithm for the direct $\mathcal{H}_2$ pseudo-optimal reduction, and furthermore, permit the judicious restart of IRKA and ISRK. Future work will cover further applications, the combination with SPARK, and the generalization to multiple inputs/outputs.

APPENDIX

Proof of the Main Theorem. Let $\mathbf{V}$ be the basis of the rational input Krylov subspace with expansion points $\{-\lambda_i\}$. Then, (10) and (21) hold and can be rewritten as:

$$\mathbf{E}_r^{-1}\mathbf{A}_r\mathbf{P}_r - \mathbf{S}\mathbf{P}_r - \mathbf{E}_r^{-1}\mathbf{b}_r\widehat{\mathbf{c}}_r\mathbf{P}_r = \mathbf{0}, \qquad (41)$$

$$\mathbf{E}_r^{-1}\mathbf{A}_r\mathbf{P}_r + \mathbf{P}_r\mathbf{A}_r^T\mathbf{E}_r^{-T} + \mathbf{E}_r^{-1}\mathbf{b}_r\mathbf{b}_r^T\mathbf{E}_r^{-T} = \mathbf{0}. \qquad (42)$$

ii) ⇔ *iii)*: We start with condition *ii)*, being equivalent to

$$-\mathbf{S}\mathbf{P}_r - \mathbf{P}_r\mathbf{A}_r^T\mathbf{E}_r^{-T} = \mathbf{0}. \tag{43}$$

Substituting $\mathbf{P}_r\mathbf{A}_r^T\mathbf{E}_r^{-T}$ by (42) yields

$$\Leftrightarrow \quad \mathbf{E}_r^{-1}\mathbf{A}_r\mathbf{P}_r - \mathbf{S}\mathbf{P}_r + \mathbf{E}_r^{-1}\mathbf{b}_r\mathbf{b}_r^T\mathbf{E}_r^{-T} = \mathbf{0}, \tag{44}$$

Comparing (44) with (41) leads to condition *iii)*, which shows the equivalence of *ii)* and *iii)*.

iii) ⇔ *iv)*: First, please note that

$$\left(\mathbf{E}_r^{-1}\mathbf{b}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\right)\left(\mathbf{E}_r^{-1}\mathbf{b}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\right)^T = \mathbf{E}_r^{-1}\mathbf{b}_r\mathbf{b}_r^T\mathbf{E}_r^{-T} + \mathbf{E}_r^{-1}\mathbf{b}_r\widehat{\mathbf{c}}_r\mathbf{P}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\mathbf{b}_r^T\mathbf{E}_r^{-T} + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r\mathbf{P}_r. \tag{45}$$

Now substituting $\mathbf{E}_r^{-1}\mathbf{A}_r\mathbf{P}_r$ in (42) by (41) leads to

$$\mathbf{S}\mathbf{P}_r + \mathbf{P}_r\mathbf{S} + \mathbf{E}_r^{-1}\mathbf{b}_r\widehat{\mathbf{c}}_r\mathbf{P}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\mathbf{b}_r^T\mathbf{E}_r^{-T} + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\mathbf{b}_r^T\mathbf{E}_r^{-T} + \mathbf{E}_r^{-1}\mathbf{b}_r\mathbf{b}_r^T\mathbf{E}_r^{-T} = \mathbf{0}, \tag{46}$$

and with (45) it follows that

$$\mathbf{S}\mathbf{P}_r + \mathbf{P}_r\mathbf{S} - \mathbf{P}_r\widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r\mathbf{P}_r = \left(\mathbf{E}_r^{-1}\mathbf{b}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\right)\left(\mathbf{E}_r^{-1}\mathbf{b}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\right)^T. \tag{47}$$

If the left hand side is zero, so is the right hand side and vice versa.

i) ⇐ *ii)*: This part is obvious, when identifying $\mathbf{P}_r$ as a similarity transformation.

i) ⇒ *ii)*: To prove this part, we again consider (10)

$$-\mathbf{S} = -\mathbf{E}_r^{-1}\mathbf{A}_r + \mathbf{E}_r^{-1}\mathbf{b}_r\widehat{\mathbf{c}}_r, \tag{48}$$

which is true for any Krylov based reduction. Condition *i)* leads to the interpretation of (48) as a pole placement problem in control theory: we are searching for the "feedback" $\widehat{\mathbf{c}}_r$ such that the eigenvalues of $\mathbf{E}_r^{-1}\mathbf{A}_r$ are mirrored along the imaginary axis. Because condition *i)* requires that all eigenvalues are assigned, it follows that the pair $(\mathbf{E}_r^{-1}\mathbf{A}_r, \mathbf{E}_r^{-1}\mathbf{b}_r)$ must be controllable and therefore, the $\widehat{\mathbf{c}}_r$ we are searching for is unique. Now rewriting (42), we find that

$$\mathbf{P}_r\mathbf{A}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1} = -\mathbf{E}_r^{-1}\mathbf{A}_r - \mathbf{E}_r^{-1}\mathbf{b}_r\mathbf{b}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1}. \tag{49}$$

Due to $\Lambda(\mathbf{A}_r, \mathbf{E}_r) = \Lambda(\mathbf{P}_r\mathbf{A}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1})$, condition *i)* requires that the left hand sides of (48) and (49)—and consequently also the right hand sides—share the same eigenvalues. However, since both right hand sides differ only in $\widehat{\mathbf{c}}_r$ and $-\mathbf{b}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1}$, respectively, and since $\widehat{\mathbf{c}}_r$ is unique, we can identify $\widehat{\mathbf{c}}_r = -\mathbf{b}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1}$. This is equal to *iii)* which is equivalent to *ii)* and proofs that *i)* ⇒ *ii)*.

iii) ⇔ *v)*: Noting that $\mathbf{X}$ in (19) is unique, we will insert *v)* in (19) and show that this is equivalent to *iii)*:

$$\mathbf{A}\mathbf{V}\mathbf{P}_r\mathbf{E}_r^T + \mathbf{E}\mathbf{V}\mathbf{P}_r\mathbf{A}_r^T + \mathbf{b}\mathbf{b}_r^T = \mathbf{0} \tag{50}$$

$$\Leftrightarrow \mathbf{A}\mathbf{V} + \mathbf{E}\mathbf{V}\mathbf{P}_r\mathbf{A}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1} + \mathbf{b}\mathbf{b}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1} = \mathbf{0} \tag{51}$$

Using (42) for $\mathbf{P}_r\mathbf{A}_r^T\mathbf{E}_r^{-T}$ yields

$$\Leftrightarrow \mathbf{A}\mathbf{V} + \mathbf{E}\mathbf{V}\mathbf{E}_r^{-1}\mathbf{A}_r^{-1} - \mathbf{E}\mathbf{V}\mathbf{E}_r^{-1}\mathbf{b}_r\mathbf{b}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1} + \mathbf{b}\mathbf{b}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1} = \mathbf{0}. \tag{52}$$

Substituting $\mathbf{b} - \mathbf{E}\mathbf{V}\mathbf{E}_r^{-1}\mathbf{b}_r = \mathbf{b}_\perp$ we get

$$\Leftrightarrow \mathbf{A}\mathbf{V} + \mathbf{E}\mathbf{V}\mathbf{E}_r^{-1}\mathbf{A}_r^{-1} = \mathbf{b}_\perp\left(-\mathbf{b}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1}\right). \tag{53}$$

Matching (53) with (7), equivalence of *iii)* and (53) follows.

iv) ⇔ *vi)*: Starting from condition *iv)*

$$\mathbf{P}_r^{-1}\mathbf{S} + \mathbf{S}^T\mathbf{P}_r^{-1} - \widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r = \mathbf{0}, \tag{54}$$

we will insert (10)

$$\mathbf{S} = \mathbf{E}_r^{-1}\mathbf{A}_r - \mathbf{E}_r^{-1}\mathbf{b}_r\widehat{\mathbf{c}}_r \overset{iii)}{=} \mathbf{E}_r^{-1}\mathbf{A}_r + \mathbf{P}_r\widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r, \tag{55}$$

to show that this is equivalent to *vi)*:

$$\Leftrightarrow \mathbf{P}_r^{-1}\mathbf{E}_r^{-1}\mathbf{A}_r + \mathbf{A}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1} + \widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r = \mathbf{0} \tag{56}$$

$$\Leftrightarrow \mathbf{E}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1}\mathbf{E}_r^{-1}\mathbf{A}_r + \mathbf{A}_r^T\mathbf{E}_r^{-T}\mathbf{P}_r^{-1}\mathbf{E}_r^{-1}\mathbf{E}_r + \widehat{\mathbf{c}}_r^T\widehat{\mathbf{c}}_r = \mathbf{0} \tag{57}$$

Comparing the Lyapunov equations (57) and (29), we can identify $\widehat{\mathbf{Q}}_r = \mathbf{E}_r^{-T}\mathbf{P}_r^{-1}\mathbf{E}_r^{-1}$, which completes the proof that conditions *i)* - *vi)* are equivalent. ☐

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their valuable remarks.

REFERENCES

[1] A. C. Antoulas. *Approximation of large-scale dynamical systems.* Advances in Design and Control, SIAM, Philadelphia, 2005.

[2] C. Beattie and S. Gugercin. Realization-independent H2-approximation. In *51st IEEE Conference on Decision and Control*, pages 4953–4958, Maui, Hawaii, USA, 2012.

[3] Y. Chahlaoui and P. Van Dooren. A collection of benchmark examples for model reduction of linear time invariant dynamical systems. Slicot working note, 2002.

[4] P. Van Dooren, K.A. Gallivan, and P.-A. Absil. $\mathcal{H}_2$-optimal model reduction of mimo systems. *Applied Mathematics Letters*, 21:1267–1273, 2008.

[5] John C. Doyle, B. Francis, and A. Tannembaum. *Feedback control theory*. Maxwell/Macmillan, 1992.

[6] Garret Flagg, Christopher Beattie, and Serkan Gugercin. Convergence of the iterative rational Krylov algorithm. *System & Control Letters*, 61:688–691, 2012.

[7] Roland W. Freund. Krylov-subspace methods for reduced-order modeling in circuit simulation. *Journal of Computational and Applied Mathematics*, 123:395–421, 2000.

[8] K. Gallivan, A. Vandendorpe, and P. Van Dooren. Sylvester equations and projection-based model reduction. *Journal of Computational and Applied Mathematics*, 162(1):213–229, 2004.

[9] E. J. Grimme. *Krylov projection methods for model reduction.* PhD thesis, Uni. Illinois at Urbana Champaign, 1997.

[10] S. Gugercin, A. C. Antoulas, and C.A. Beattie. H2 model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications,*, 30(2):609–638, 2008.

[11] Serkan Gugercin. An iterative svd-krylov based method for model reduction of large-scale dynamical systems. *Linear Algebra and its Applications*, 428:1964–1986, 2008.

[12] Heiko K. F. Panzer, Stefan Jaensch, Thomas Wolf, and Boris Lohmann. A greedy rational Krylov method for H2-pseudooptimal model order reduction with preservation of stability. In *American Control Conference*, Washington, DC, USA, 2013.

[13] Heiko K. F. Panzer, Thomas Wolf, and Boris Lohmann. H2 and Hinf error bounds for model order reduction of second order systems by Krylov subspace methods. In *European Control Conference*, Zurich, Switzerland, 2013.

[14] Thomas Wolf, Heiko Panzer, and Boris Lohmann. Gramian-based error bound in model reduction by Krylov-subspace methods. In *18th IFAC World Congress*, pages 3587–3592, Milano, Italy, August 2011.

[15] Thomas Wolf, Heiko Panzer, and Boris Lohmann. Sylvester equations and a factorization of the error system in Krylov-based model reduction. In *Vienna Conference on Mathematical Modelling*, 2012.