$\mathcal{H}_2$-OPTIMAL MODEL REDUCTION WITH HIGHER ORDER POLES*

PAUL VAN DOOREN†‡, KYLE A. GALLIVAN¶‖, AND P.-A. ABSIL†§

Second revision, compiled July 16, 2010, 15:51

Abstract. We revisit the problem of approximating a multiple-input multiple-output (MIMO) $p \times m$ rational transfer function $H(s)$ of high degree by another $p \times m$ rational transfer function $\widehat{H}(s)$ of much smaller degree, so that the $\mathcal{H}_2$ norm of the approximation error is minimized. We show that in the general case of higher order poles in the reduced order model, called the defective case, the stationary points of the $\mathcal{H}_2$ norm of the approximation error can still be characterized by tangential interpolation conditions. We also indicate that the sensitivity of the solution of this problem depends on the parameterization used.

Key words. Model reduction, optimal $\mathcal{H}_2$ approximation, tangential interpolation.

AMS subject classifications. 41A05, 65D05, 93B40

1. Introduction. In this paper, we consider the problem of approximating a real $p \times m$ rational transfer function $H(s)$ of McMillan degree N by a real $p \times m$ rational transfer function $\widehat{H}(s)$ of lower McMillan degree n using the $\mathcal{H}_2$-norm as the approximation criterion. We refer to [2] for the relevant background on model reduction and linear system theory.

Since a transfer function has an unbounded $\mathcal{H}_2$-norm if it is not strictly proper, we will constrain both $H(s)$ and $\widehat{H}(s)$ to be strictly proper (i.e., they are zero at $s = \infty$). Such transfer functions have minimal (i.e., controllable and observable) state-space realizations $(A, B, C) \in \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times m} \times \mathbb{R}^{p\times N}$ and $(\widehat{A}, \widehat{B}, \widehat{C}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times m} \times \mathbb{R}^{p\times n}$ satisfying

$$\left\{ \begin{array}{l} \dot{x} = Ax + Bu, \\ y = Cx, \end{array} \right. \qquad H(s) := C(sI_N - A)^{-1}B, \tag{1.1}$$

and

$$\left\{ \begin{array}{l} \dot{\widehat{x}} = \widehat{A}\widehat{x} + \widehat{B}u, \\ \widehat{y} = \widehat{C}\widehat{x}, \end{array} \right. \qquad \widehat{H}(s) := \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}, \tag{1.2}$$

where $u \in \mathbb{R}^m$, $y, \widehat{y} \in \mathbb{R}^p$, $x \in \mathbb{R}^N$, $\widehat{x} \in \mathbb{R}^n$. Expressions for the gradients of the squared $\mathcal{H}_2$-norm error function

$$\mathcal{J}_{(A,B,C)} : (\widehat{A}, \widehat{B}, \widehat{C}) \mapsto \|C(sI_N - A)^{-1}B - \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}\|^2_{\mathcal{H}_2}$$

have been known since the work of Wilson [16] (these expressions are recalled in Theorem 3.2). One can object, however, that the *full parameterization*

$$(\widehat{A}, \widehat{B}, \widehat{C}) \mapsto \widehat{H}(s) = \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B} \tag{1.3}$$

*This paper presents research supported by the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office and by the National Science Foundation under contract OCI-03-24944. The scientific responsibility rests with its authors.

†Université catholique de Louvain, Dept. Math. Eng., B-1348 Louvain-la-Neuve, Belgium.

‡http://www.inma.ucl.ac.be/~vdooren/

¶Department of Mathematics, Florida State University, Tallahassee FL 32306, USA.

‖http://www.math.fsu.edu/~gallivan/

§http://www.inma.ucl.ac.be/~absil/

is not one to one, since the triple

$$(\widehat{A}_T, \widehat{B}_T, \widehat{C}_T) := (T^{-1}\widehat{A}T, T^{-1}\widehat{B}, \widehat{C}T)$$

for any matrix $T \in GL(n, \mathbb{R})$ defines the same transfer function :

$$\widehat{H}(s) = \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B} = \widehat{C}_T(sI_n - \widehat{A}_T)^{-1}\widehat{B}_T.$$

If one could eliminate the n^2 degrees of freedom of the invertible transformation T, one could hope to fully parameterize the target system $\widehat{H}(s)$ with only $n(m+p)$ independent parameters, and to turn Wilson's conditions into $n(m+p)$ non redundant scalar conditions. Concerning the parameterization task, Byrnes and Falb [6, Th. 4.7] show that the set $\mathrm{Rat}^n_{p,m}$ of $p \times m$ strictly proper rational transfer functions of degree n can be parameterized with only $n(m+p)$ real parameters in a locally smooth manner; but it is also shown there that there exists *no globally smooth parameterization* of $\mathrm{Rat}^n_{p,m}$ if $\min(p, m) > 1$. An obvious candidate for a minimal parameterization is the Jordan canonical form, which happens to be smooth almost everywhere *except at points where A is degenerate or non-diagonalizable.* When one approaches such matrices, the diagonalizing similarity transformation tends to a singular matrix, which explains the non-smoothness of that representation. Therefore, minimal parameterizations should be avoided and over-parameterizations considered.

When they have only first-order poles the diagonalizable case, the stationary points $\widehat{H}(s)$ of the $\mathcal{H}_2$-norm error function (i.e., the points where the gradient of $J_{(A,B,C)}$ vanishes) can be characterized via interpolation conditions at the negative of the poles of $\widehat{H}(s)$. These results are, in fact, a consequence of the relation between the equations of the gradients of the $\mathcal{H}_2$-norm error (as derived originally by Wilson in [16]) and tangential interpolation based on Sylvester equations (as derived in [3], [7], [8]). These links were obtained by several authors, using different approaches ([12], [11], [10], [5]).

In this paper, we characterize the stationary points $\widehat{H}(s)$ of the $\mathcal{H}_2$-norm error function *without* the assumption that they have only first-order poles and use the Jordan canonical form to describe the solution. The stationarity conditions elegantly generalize to higher-order tangential interpolation conditions of degree $k_i - 1$ (in the sense of [8]), where k_i is the size of the ith Jordan block. The interpolation points remain the negative of the poles $\widehat{\lambda}_i$ of $\widehat{H}(s)$, and the interpolation directions are polynomial vectors of degree $k_i - 1$, built from the rows of $\widehat{B}$ and the columns of $\widehat{C}$. We can also show (see [15]) that these tangential interpolation conditions contain $n(m + p)$ non-redundant scalar conditions. The result in Theorem 4.5 has several precursors: Aigrain and Williams [1] for the SISO case with simple real poles, Meier and Luenberger [13] for the general SISO case, Halevi [11] for the square MIMO case (but only for the interpolation points and not for the tangential directions) and Gugercin et al. [10], Van Dooren et al. [14], and Bunse-Gerstner et al. [5] for the MIMO case without high order poles.

Since the set of systems with higher-order poles is nowhere dense in $\mathrm{Rat}^n_{p,m}$, the generalization of the stationarity conditions to higher-order poles (or the defective case) is chiefly of theoretical interest. However, we also argue in this paper that the non-smoothness of the Jordan canonical form at the systems with higher-order poles is a source of difficulties that should not be dismissed. Indeed, model reduction problems occur in practice where the reduced-order system is close to having higher-order poles. A typical example is position control of an (essentially) undamped system

(such as a robot arm) using a force actuator. Since position is the double integral of acceleration (or force), the ideal (linearized) model for such a system is one with a double pole at the origin. Any reduced order model that tries to capture such a specific dynamic is bound to be close to having a second order pole at the origin. We also argue that the case of higher-order poles is important for both understanding the nature of the approximation problem and for computational purposes. First, we show on an example that $\mathcal{H}_2$-optimal reduced-order models with higher-order poles do occur. Second, we point out that because the minimal parameterization changes in a non-smooth manner at the higher-order poles, the minimal tangential interpolation conditions for $\mathcal{H}_2$-norm stationary points become ill conditioned around systems $\widehat{H}(s)$ with higher-order poles. When the influence of a nearby higher-order pole becomes problematic, it is therefore better to use a smooth parameterization such as the full reduced order model (1.3). We illustrate this with a simple numerical example.

The paper is organized as follows. After presenting in Section 2 the necessary background material on the $\mathcal{H}_2$ approximation problem, in Section 3 we recall Wilson's formulas for the gradient of the $\mathcal{H}_2$-norm error function. In Section 4, Wilson's first-order optimality conditions are expressed in a tangential interpolation form obtained by representing the reduced-order model in Jordan canonical form—thus covering the case of higher-order poles in the reduced-order model. The link to tangential interpolation by means of projection matrices that solve Sylvester equations is discussed in Section 5. The importance of dealing with the case of higher-order poles is illustrated in Section 6. The discrete-time case is covered in Section 7, and conclusions are drawn in Section 8.

2. The $\mathcal{H}_2$ approximation problem. Much of the material in this section is standard and can be found in [2]. Let $E(s)$ be an arbitrary strictly proper transfer function, with realization triple (A_e, B_e, C_e). If $E(s)$ is unstable, its $\mathcal{H}_2$-norm is defined to be ∞. Otherwise, its squared $\mathcal{H}_2$-norm is defined as the trace of a matrix integral :

$$\|E(s)\|_{\mathcal{H}_2}^2 := \operatorname{tr} \int_{-\infty}^{\infty} E(j\omega)E(j\omega)^H \frac{d\omega}{2\pi} = \operatorname{tr} \int_0^{\infty} [C_e \exp^{A_e t} B_e][C_e \exp^{A_e t} B_e]^T dt. \tag{2.1}$$

This can be related to an expression involving the gramians P_e and Q_e defined as

$$P_e := \int_0^{\infty} [\exp^{A_e t} B_e][\exp^{A_e t} B_e]^T dt, \quad Q_e := \int_0^{\infty} [C_e \exp^{A_e t}]^T [C_e \exp^{A_e t}] dt,$$

which are also known to be the solutions of the Lyapunov equations

$$A_e P_e + P_e A_e^T + B_e B_e^T = 0, \quad Q_e A_e + A_e^T Q_e + C_e^T C_e = 0. \tag{2.2}$$

Using these, it easily follows that the squared $\mathcal{H}_2$-norm of $E(s)$ can be expressed as

$$\|E(s)\|_{\mathcal{H}_2}^2 = \operatorname{tr} B_e^T Q_e B_e = \operatorname{tr} C_e P_e C_e^T. \tag{2.3}$$

We now apply this to the error function

$$E(s) := H(s) - \widehat{H}(s) = C(sI_N - A)^{-1}B - \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}.$$

A realization of $E(s)$ in partitioned form is given by

$$(A_e, B_e, C_e) := \left(\begin{bmatrix} A & \\ & \widehat{A} \end{bmatrix}, \begin{bmatrix} B \\ \widehat{B} \end{bmatrix}, \begin{bmatrix} C & -\widehat{C} \end{bmatrix} \right), \tag{2.4}$$

and the Lyapunov equations (2.2) become

$$P_e := \begin{bmatrix} P & X \\ X^T & \widehat{P} \end{bmatrix}, \quad \begin{bmatrix} A & \\ & \widehat{A} \end{bmatrix} \begin{bmatrix} P & X \\ X^T & \widehat{P} \end{bmatrix} + \begin{bmatrix} P & X \\ X^T & \widehat{P} \end{bmatrix} \begin{bmatrix} A^T & \\ & \widehat{A}^T \end{bmatrix} + \begin{bmatrix} B \\ \widehat{B} \end{bmatrix} \begin{bmatrix} B^T & \widehat{B}^T \end{bmatrix} = 0, \tag{2.5}$$

and

$$Q_e := \begin{bmatrix} Q & Y \\ Y^T & \widehat{Q} \end{bmatrix}, \quad \begin{bmatrix} A^T & \\ & \widehat{A}^T \end{bmatrix} \begin{bmatrix} Q & Y \\ Y^T & \widehat{Q} \end{bmatrix} + \begin{bmatrix} Q & Y \\ Y^T & \widehat{Q} \end{bmatrix} \begin{bmatrix} A & \\ & \widehat{A} \end{bmatrix} + \begin{bmatrix} C^T \\ -\widehat{C}^T \end{bmatrix} \begin{bmatrix} C & -\widehat{C} \end{bmatrix} = 0. \tag{2.6}$$

To obtain a low-order approximation $\widehat{H}(s) = \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}$ of a given full-order model $H(s) = C(sI_N - A)^{-1}B$, we need to minimize the $\mathcal{H}_2$-distance $\|H(s) - \widehat{H}(s)\|^2_{\mathcal{H}_2}$ also given by the function

$$\mathcal{J}_{(A,B,C)}(\widehat{A}, \widehat{B}, \widehat{C}) = \|C(sI_N - A)^{-1}B - \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}\|^2_{\mathcal{H}_2}. \tag{2.7a}$$

We will frequently omit the subscript in $\mathcal{J}_{(A,B,C)}(\widehat{A}, \widehat{B}, \widehat{C})$ when the full-order model is clear from the context. In view of (2.3), $\mathcal{J}(\widehat{A}, \widehat{B}, \widehat{C})$ admits the formulation

$$\mathcal{J}(\widehat{A}, \widehat{B}, \widehat{C}) = \operatorname{tr}\left(\begin{bmatrix} B^T & \widehat{B}^T \end{bmatrix} \begin{bmatrix} Q & Y \\ Y^T & \widehat{Q} \end{bmatrix} \begin{bmatrix} B \\ \widehat{B} \end{bmatrix} \right) = \operatorname{tr}\left(B^T Q B + 2B^T Y \widehat{B} + \widehat{B}^T \widehat{Q} \widehat{B} \right), \tag{2.7b}$$

where Q, Y and $\widehat{Q}$ depend on A, $\widehat{A}$, C and $\widehat{C}$ through the Lyapunov equation (2.6), or equivalently

$$\mathcal{J}(\widehat{A}, \widehat{B}, \widehat{C}) = \operatorname{tr}\left(\begin{bmatrix} C & -\widehat{C} \end{bmatrix} \begin{bmatrix} P & X \\ X^T & \widehat{P} \end{bmatrix} \begin{bmatrix} C^T \\ -\widehat{C}^T \end{bmatrix} \right) = \operatorname{tr}\left(CPC^T - 2CX\widehat{C}^T + \widehat{C}\widehat{P}\widehat{C}^T \right), \tag{2.7c}$$

where P, X and $\widehat{P}$ depend on A, $\widehat{A}$, B and $\widehat{B}$ through the Lyapunov equation (2.5). Note that the terms B^TQB and CPC^T in the above expressions are constant, and hence can be discarded in the optimization.

3. Gradients of the squared $\mathcal{H}_2$-norm error function. The derivations above can be used to obtain formulas for the gradients of the squared $\mathcal{H}_2$-norm error function $\mathcal{J}$ versus $\widehat{A}$, $\widehat{B}$, and $\widehat{C}$. We define the gradients as follows.

DEFINITION 3.1. *The gradient of a real-valued function $f(M)$ of a real matrix variable $M \in \mathbb{R}^{k\times l}$, is the real matrix $\nabla_M f(M) \in \mathbb{R}^{k\times l}$, defined by*

$$[\nabla_M f(M)]_{i,j} = \frac{\partial}{\partial M_{i,j}} f(M), \quad i = 1, \ldots, k, \quad j = 1, \ldots, l. \tag{3.1}$$

We will write $\nabla_{\widehat{A}} f$ as a compact notation for $\nabla_{\widehat{A}} f(\widehat{A}, \widehat{B}, \widehat{C})$ when the argument is clear from the context.

Starting from the characterizations (2.5,2.7c) and (2.6,2.7b) of the $\mathcal{H}_2$ norm, one can derive succinct forms of the gradients. This theorem is originally due to Wilson [16], but we state here the version derived in [14], where a proof based on inner products and traces is given.

LEMMA 3.2. *The gradients $\nabla_{\widehat{A}}\mathcal{J}$, $\nabla_{\widehat{B}}\mathcal{J}$ and $\nabla_{\widehat{C}}\mathcal{J}$ of the squared $\mathcal{H}_2$-norm error $\mathcal{J}$ (2.7), where both (A, B, C) and $(\widehat{A}, \widehat{B}, \widehat{C})$ are minimal (i.e., controllable and observable), are given by*

$$\nabla_{\widehat{A}}\mathcal{J} = 2(\widehat{Q}\widehat{P} + Y^T X), \quad \nabla_{\widehat{B}}\mathcal{J} = 2(\widehat{Q}\widehat{B} + Y^T B), \quad \nabla_{\widehat{C}}\mathcal{J} = 2(\widehat{C}\widehat{P} - CX), \tag{3.2}$$

where

$$A^TY + Y\widehat{A} - C^T\widehat{C} = 0, \quad \widehat{A}^T\widehat{Q} + \widehat{Q}\widehat{A} + \widehat{C}^T\widehat{C} = 0, \tag{3.3}$$
$$X^TA^T + \widehat{A}X^T + \widehat{B}B^T = 0, \quad \widehat{P}\widehat{A}^T + \widehat{A}\widehat{P} + \widehat{B}\widehat{B}^T = 0. \tag{3.4}$$

4. Stationarity conditions in Jordan form. We will assume that both transfer functions $H(s)$ and $\widehat{H}(s)$ have real minimal (controllable and observable) realizations (A, B, C) and $(\widehat{A}, \widehat{B}, \widehat{C})$ and we allow $\widehat{H}(s)$ to have multiple and higher-order poles. The main result is given in Theorem 4.5, where we show that the stationary points of the $\mathcal{H}_2$-norm error function are characterized by tangential interpolation conditions whose degree depends on the size of the Jordan blocks of $\widehat{H}(s)$.

Let $\widehat{H}(s)$ have the following minimal representation

$$\widehat{H}(s) = \sum_{i=1}^{\ell} \widehat{H}_i(s), \quad \widehat{H}_i(s) := \widehat{C}_i(sI - \widehat{A}_i)^{-1}\widehat{B}_i^H, \quad \widehat{A}_i := \begin{bmatrix} \widehat{\lambda}_i & -1 & & \\ & \widehat{\lambda}_i & \ddots & \\ & & \ddots & -1 \\ & & & \widehat{\lambda}_i \end{bmatrix}, \tag{4.1}$$

where $\widehat{A}_i \in \mathbb{C}^{k_i \times k_i}$, $\widehat{B}_i^H \in \mathbb{C}^{k_i \times m}$, $\widehat{C}_i \in \mathbb{C}^{p \times k_i}$ and where $\{(\widehat{A}_i, \widehat{B}_i^H, \widehat{C}_i) : i = 1, \ldots, \ell\}$ is a self-conjugate set. Notice that this is essentially the partial fraction expansion of $\widehat{H}(s)$ and that there may be more than one Jordan block $\widehat{A}_i$ associated with the same complex eigenvalue $\widehat{\lambda}_i$. The minimality of the representation implies linear independence of the leading columns in each block $\widehat{B}_i$ and of the trailing rows in each block $\widehat{C}_i$ that correspond to the same eigenvalue $\widehat{\lambda}_i$, since these blocks appear as subblocks of a minimal realization of $\widehat{H}(s)$.

We will need S_i, T_i^H, the (complex) left and right eigenspaces of the (real) matrix $\widehat{A}$ corresponding to the (complex) eigenvalue $\widehat{\lambda}_i$. From the expansion (4.1), we have :

$$\widehat{A}S_i = S_i\widehat{A}_i, \quad \widehat{C}S_i = \widehat{C}_i, \quad T_i^H\widehat{A} = \widehat{A}_iT_i^H, \quad T_i^H\widehat{B} = \widehat{B}_i^H, \quad T_i^HS_i = I_k. \tag{4.2}$$

Note also that the matrices S_i and T_i^H are not unique. When there is only one Jordan block associated with an eigenvalue $\widehat{\lambda}_i$, its degree of freedom is just a block scaling S_iD_i and $D_i^{-1}T_i^H$ with $D_i \in \mathbb{C}^{k_i \times k_i}$ invertible. When there is more than one Jordan block associated with $\widehat{\lambda}_i$, the degrees of freedom are more involved. Below, we associate right and left bases S_i, T_i with each individual Jordan block A_i.

We will also need the following lemmas in preparation for the main theorem.

LEMMA 4.1. *If $-\lambda$ is not an eigenvalue of A, the solution of the matrix equation*

$$A^TY + YF - C^TL = 0 \quad \text{with} \quad F := \begin{bmatrix} \lambda & -1 & & \\ & \lambda & \ddots & \\ & & \ddots & -1 \\ & & & \lambda \end{bmatrix} \in \mathbb{C}^{k \times k}, \tag{4.3}$$

and $L := \begin{bmatrix}\ell_0 & \ell_1 & \dots & \ell_{k-1}\end{bmatrix}$, *is given by*

$$Y = \mathcal{C}_{A,C}(-\lambda)\begin{bmatrix}\ell_0 & \ell_1 & \dots & \ell_{k-1}\\ & \ell_0 & \ddots & \vdots\\ & & \ddots & \ell_1\\ & & & \ell_0\end{bmatrix},$$

where

$$\mathcal{C}_{A,C}(-\lambda) := \begin{bmatrix}(A^T+\lambda I)^{-1}C^T & \dots & (A^T+\lambda I)^{-k}C^T\end{bmatrix}.$$

Moreover, let $\phi_\lambda(s) := \begin{bmatrix}1 & (s+\lambda) & \dots & (s+\lambda)^{k-1}\end{bmatrix}^T$, $y(s) := Y\phi_\lambda(s)$, *then*

$$y(s) = (A^T - sI)^{-1}C^T L\phi_\lambda(s) + O((s+\lambda)^k)$$

which means that the ith column y_i *of* Y *is also the coefficient of* $(s+\lambda)^{i-1}$ *in the Taylor expansion of* $(A^T - sI)^{-1}C^T L\phi_\lambda(s)$.

Proof. The first part easily follows from $(A^T+\lambda I)y_1 = C^T\ell_0$ and $(A^T+\lambda I)y_i = C^T\ell_{i-1} + y_{i-1}$, $i > 1$. The second part follows from the identity

$$(A^T - sI)^{-1}C^T = \sum_{i=1}^{\infty}(s+\lambda)^{i-1}(A^T+\lambda I)^{-i}C^T$$

and from the convolution of this formal series with the polynomial vector $L\phi_\lambda(s)$. □

We also give the dual version of this lemma without proof.

LEMMA 4.2. *If* $-\lambda$ *is not an eigenvalue of* A, *the solution of the matrix equation*

$$X^H A^T + FX^H - R^H B^T = 0$$

with $F \in \mathbb{C}^{k\times k}$ *as in* (4.3) *and* $R := \begin{bmatrix}r_{k-1} & r_{k-2} & \dots & r_0\end{bmatrix}$, *is given by*

$$X^H = \begin{bmatrix}r_0^H & r_1^H & \dots & r_{k-1}^H\\ & r_0^H & \ddots & \vdots\\ & & \ddots & r_1^H\\ & & & r_0^H\end{bmatrix}\mathcal{O}_{A,B}(-\lambda), \text{ where } \mathcal{O}_{A,B}(-\lambda) := \begin{bmatrix}B^T(A^T+\lambda I)^{-k}\\ \vdots\\ B^T(A^T+\lambda I)^{-1}\end{bmatrix}.$$

Moreover, let $\psi_\lambda(s) := \begin{bmatrix}(s+\lambda)^{k-1} & \dots & (s+\lambda) & 1\end{bmatrix}$, $x^H(s) := \psi_\lambda(s)X^H$, *then*

$$x^H(s) = \psi_\lambda(s)R^H B^T(A^T - sI)^{-1} + O((s+\lambda)^k)$$

which means that the ith row x_i^H *of* X^H *is also the coefficient of* $(s+\lambda)^{i-1}$ *in the Taylor expansion of* $\psi_\lambda(s)R^H B^T(A^T - sI)^{-1}$.

We first obtain expressions for $\nabla_{\widehat{B}}\mathcal{J}$ and $\nabla_{\widehat{C}}\mathcal{J}$ that exploits the Jordan canonical form.

THEOREM 4.3. *Let* $H(s) = C(sI_N - A)^{-1}B$ *and* $\widehat{H}(s) = \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}$ *be real minimal realizations, and let* $\widehat{A}_i$, $\widehat{B}_i$, $\widehat{C}_i$, S_i, *and* T_i, $i = 1,\dots,\ell$, *describe the Jordan canonical form of* $\widehat{H}(s)$ *as in* (4.1) *and* (4.2). *Assume that* $-\widehat{\lambda}_i$ *is not a pole of* $H(s)$, $i = 1,\dots,\ell$. *Define*

$$\psi_{\widehat{\lambda}_i}(s) := \begin{bmatrix}(s+\widehat{\lambda}_i)^{k_i-1} & \dots & (s+\widehat{\lambda}_i) & 1\end{bmatrix}, \quad \phi_{\widehat{\lambda}_i}(s) := \begin{bmatrix}1 & (s+\widehat{\lambda}_i) & \dots & (s+\widehat{\lambda}_i)^{k_i-1}\end{bmatrix}^T.$$

Then we have

$$\frac{1}{2}(\nabla_{\widehat{B}}\mathcal{J})^T S_i \phi_{\widehat{\lambda}_i}(s) = [H^T(s) - \widehat{H}^T(s)]\widehat{C}_i \phi_{\widehat{\lambda}_i}(s) + O(s+\widehat{\lambda}_i)^{k_i}, \tag{4.4}$$

$$\frac{1}{2}\psi_{\widehat{\lambda}_i}(s) T_i^H (\nabla_{\widehat{C}}\mathcal{J})^T = \psi_{\widehat{\lambda}_i}(s)\widehat{B}_i^H [H^T(s) - \widehat{H}^T(s)] + O(s+\widehat{\lambda}_i)^{k_i}, \tag{4.5}$$

where $\mathcal{J}$ is the squared $\mathcal{H}_2$-norm error defined in (2.7).

Proof. Define $Y_i := YS_i$, $\widehat{Q}_i := -\widehat{Q}S_i$, $X_i := -XT_i$ and $\widehat{P}_i := -\widehat{P}T_i$. Then

$$A^T Y_i + Y_i \widehat{A}_i = C^T \widehat{C}_i, \quad \widehat{A}^T \widehat{Q}_i + \widehat{Q}_i \widehat{A}_i = \widehat{C}^T \widehat{C}_i,$$

$$X_i^H A^T + \widehat{A}_i X_i^H = \widehat{B}_i^H B^T, \quad \widehat{P}_i^H \widehat{A}^T + \widehat{A}_i \widehat{P}_i^H = \widehat{B}_i^H \widehat{B}^T.$$

If $-\widehat{\lambda}_i$ is not an eigenvalue of A or $\widehat{A}$, both $(A^T - sI)^{-1}$ and $(\widehat{A}^T - sI)^{-1}$ have Taylor expansions in $(s+\widehat{\lambda}_i)$. It then follows from Lemmas 4.1 and 4.2 that

$$Y_i \phi_{\widehat{\lambda}_i}(s) = (A^T - sI)^{-1} C^T \widehat{C}_i \phi_{\widehat{\lambda}_i}(s) + O(s+\widehat{\lambda}_i)^{k_i}, \tag{4.6}$$

$$\widehat{Q}_i \phi_{\widehat{\lambda}_i}(s) = (\widehat{A}^T - sI)^{-1} \widehat{C}^T \widehat{C}_i \phi_{\widehat{\lambda}_i}(s) + O(s+\widehat{\lambda}_i)^{k_i}, \tag{4.7}$$

$$\psi_{\widehat{\lambda}_i}(s) X_i^H = \psi_{\widehat{\lambda}_i}(s)\widehat{B}_i^H B^T (A^T - sI)^{-1} + O(s+\widehat{\lambda}_i)^{k_i}, \tag{4.8}$$

$$\psi_{\widehat{\lambda}_i}(s) \widehat{P}_i^H = \psi_{\widehat{\lambda}_i}(s)\widehat{B}_i^H \widehat{B}^T (\widehat{A}^T - sI)^{-1} + O(s+\widehat{\lambda}_i)^{k_i}. \tag{4.9}$$

This then yields, by Lemma 3.2,

$$\frac{1}{2}(\nabla_{\widehat{B}}\mathcal{J})^T S_i \phi_{\widehat{\lambda}_i}(s) = (\widehat{B}^T\widehat{Q} + B^T Y) S_i \phi_{\widehat{\lambda}_i}(s) = [H^T(s) - \widehat{H}^T(s)]\widehat{C}_i \phi_{\widehat{\lambda}_i}(s) + O(s+\widehat{\lambda}_i)^{k_i},$$

$$\frac{1}{2}\psi_{\widehat{\lambda}_i}(s) T_i^H (\nabla_{\widehat{C}}\mathcal{J})^T = \psi_{\widehat{\lambda}_i}(s) T_i^H (\widehat{P}\widehat{C}^T - X^T C^T) = \psi_{\widehat{\lambda}_i}(s)\widehat{B}_i^H [H^T(s) - \widehat{H}^T(s)] + O(s+\widehat{\lambda}_i)^{k_i}.$$

□

The condition that $-\widehat{\lambda}_i$ is not a pole of $H(s)$ is satisfied when choosing stable interpolation points $\widehat{\lambda}_i$, which is typically the case in the algorithms we discuss below.

THEOREM 4.4. *With the notation and assumptions of Theorem 4.3, if $\nabla_{\widehat{B}}\mathcal{J} = 0$ and $\nabla_{\widehat{C}}\mathcal{J} = 0$, then for each Jordan block $\widehat{A}_i$ in the realization of $\widehat{H}(s)$, we have in addition the relation*

$$\frac{1}{2}\psi_{\widehat{\lambda}_i}(s) T_i^H (\nabla_{\widehat{A}}\mathcal{J})^T S_i \phi_{\widehat{\lambda}_i}(s) = \psi_{\widehat{\lambda}_i}(s)\widehat{B}_i^H [H^T(s) - \widehat{H}^T(s)]\widehat{C}_i \phi_{\widehat{\lambda}_i}(s) + O(s+\lambda)^{2k_i}. \tag{4.10}$$

Proof. Let S_i, T_i^H be as in Theorem 4.3. Define as before

$$\widehat{C}_i := \widehat{C}S_i,\ Y_i := YS_i,\ \widehat{Q}_i := -\widehat{Q}S_i,\ \widehat{B}_i^H := T_i^H \widehat{B},\ X_i^H = -T_i^H X^T,\ \widehat{P}_i^H := -T_i^H \widehat{P},$$

then we have

$$A^T Y_i + Y_i \widehat{A}_i = C^T \widehat{C}_i, \quad \widehat{A}^T \widehat{Q}_i + \widehat{Q}_i \widehat{A}_i = \widehat{C}^T \widehat{C}_i,$$

$$X_i^H A^T + \widehat{A}_i X_i^H = \widehat{B}_i^H B^T, \quad \widehat{P}_i^H \widehat{A}^T + \widehat{A}_i \widehat{P}_i^H = \widehat{B}_i^H \widehat{B}^T.$$

From Lemma 3.2 it follows that $\frac{1}{2}\nabla_{\widehat{A}}\mathcal{J} = \widehat{P}\widehat{Q} + X^T Y$. If we use Lemmas 4.1, 4.2 and (4.20) below, we then obtain

$$\frac{1}{2}\psi_{\widehat{\lambda}_i}(s) T_i^H (\nabla_{\widehat{A}}\mathcal{J})^T S_i \phi_{\widehat{\lambda}_i}(s) = \psi_{\widehat{\lambda}_i}(s) \widehat{B}_i^H [H^T(s) - \widehat{H}^T(s)] \widehat{C}_i \phi_{\widehat{\lambda}_i}(s) + O(s + \widehat{\lambda}_i)^{2k_i}.$$

□

THEOREM 4.5. *With the notation and assumptions of Theorem 4.3, if* $\nabla_{\widehat{B}}\mathcal{J} = 0$, $\nabla_{\widehat{C}}\mathcal{J} = 0$ *and* $\nabla_{\widehat{A}}\mathcal{J} = 0$, *then the following tangential interpolation conditions are satisfied for* $i = 1, \ldots, \ell$*:*

$$[H^T(s) - \widehat{H}^T(s)]\widehat{c}_i(s) = O(s + \widehat{\lambda}_i)^{k_i}, \tag{4.11}$$

$$\widehat{b}_i(s)^H [H^T(s) - \widehat{H}^T(s)] = O(s + \widehat{\lambda}_i)^{k_i}, \tag{4.12}$$

$$\widehat{b}_i(s)^H [H^T(s) - \widehat{H}^T(s)]\widehat{c}_i(s) = O(s + \widehat{\lambda}_i)^{2k_i}, \tag{4.13}$$

where $\widehat{b}_i^H(s) := \psi_{\widehat{\lambda}_i}(s)\widehat{B}_i^H$ *and* $\widehat{c}_i(s) := \widehat{C}_i \phi_{\widehat{\lambda}_i}(s)$.

Proof. Conditions (4.11) and (4.12) follow immediately from Theorem 4.3. It remains to show that (4.13) holds. We can interpret conditions (4.11)–(4.13) in terms of Taylor expansions of the error function $E(s) := H(s) - \widehat{H}(s)$. Let

$$E(s) := \sum_{j=0}^{\infty} E_i (s + \widehat{\lambda}_i)^j, \quad \widehat{c}_i(s) := \sum_{j=0}^{k_i} l_j (s + \widehat{\lambda}_i)^j, \quad \widehat{b}_i^H(s) := \sum_{j=0}^{k_i} r_j^H (s + \widehat{\lambda}_i)^j,$$

be the Taylor expansions around $s = -\widehat{\lambda}_i$ of the rational function $E(s)$ and of the polynomials $\widehat{c}_i(s)$ and $\widehat{b}_i(s)^H$. Then conditions (4.11)–(4.13) are respectively equivalent to

$$\begin{bmatrix} E_0^H & E_1^H & \ldots & E_{k_i-1}^H \\ & E_0^H & \ddots & \vdots \\ & & \ddots & E_1^H \\ & & & E_0^H \end{bmatrix} \begin{bmatrix} l_0 & l_1 & \ldots & l_{k_i-1} \\ & l_0 & \ddots & \vdots \\ & & \ddots & l_1 \\ & & & l_0 \end{bmatrix} = 0, \tag{4.14}$$

$$\begin{bmatrix} r_0^H & r_1^H & \ldots & r_{k_i-1}^H \\ & r_0^H & \ddots & \vdots \\ & & \ddots & r_1^H \\ & & & r_0^H \end{bmatrix} \begin{bmatrix} E_0^H & E_1^H & \ldots & E_{k_i-1}^H \\ & E_0^H & \ddots & \vdots \\ & & \ddots & E_1^H \\ & & & E_0^H \end{bmatrix} = 0, \tag{4.15}$$

and

$$\begin{bmatrix} r_0^H & r_1^H & \ldots & r_{2k_i-1}^H \\ & r_0^H & \ddots & \vdots \\ & & \ddots & r_1^H \\ & & & r_0^H \end{bmatrix} \begin{bmatrix} E_0^H & E_1^H & \ldots & E_{2k_i-1}^H \\ & E_0^H & \ddots & \vdots \\ & & \ddots & E_1^H \\ & & & E_0^H \end{bmatrix} \begin{bmatrix} l_0 & l_1 & \ldots & l_{2k_i-1} \\ & l_0 & \ddots & \vdots \\ & & \ddots & l_1 \\ & & & l_0 \end{bmatrix} = 0. \tag{4.16}$$

The condition that the first k_i or $2k_i$ terms of the Taylor expansion vanish is equivalent to the fact that the above partial convolutions are zero. We know that (4.14) and (4.15) hold, since (4.11) and (4.12) hold; it remains to show (4.16) to conclude the proof.

We will need the identity

$$\begin{bmatrix} E_{k_i}^H & \dots & E_{2k_i-1}^H \\ \vdots & \ddots & \vdots \\ E_1^H & \dots & E_{k_i}^H \end{bmatrix} = \mathcal{O}_{A,B}(-\lambda_i)\mathcal{C}_{A,C}(-\lambda_i) - \mathcal{O}_{\widehat{A},\widehat{B}}(-\lambda_i)\mathcal{C}_{\widehat{A},\widehat{C}}(-\lambda_i) \quad (4.17)$$

which holds since

$$E_{f+g-1}^H = B^T(A^T+\lambda_i I)^{-f}(A^T+\lambda_i I)^{-g}C^T - \widehat{B}^T(\widehat{A}^T+\lambda_i I)^{-f}(\widehat{A}^T+\lambda_i I)^{-g}\widehat{C}^T.$$

Define

$$Y_i := YS_i,\ \widehat{Q}_i := -\widehat{Q}S_i,\ X_i^H = -T_i^H X^T,\ \widehat{P}_i^H := -T_i^H\widehat{P}. \quad (4.18)$$

Using Wilson's formulas (Theorem 3.2) for the first equality, Lemmas 4.1 and 4.2 for the second one, and the identity (4.17) for the third, we have

$$\begin{aligned}
&T_i^H(\nabla_{\widehat{A}}\mathcal{J})^T S_i = \widehat{P}_i^H\widehat{Q}_i - X_i^H Y_i \\
&= \begin{bmatrix} r_0^H & r_1^H & \dots & r_{k_i-1}^H \\ & r_0^H & \ddots & \vdots \\ & & \ddots & r_1^H \\ & & & r_0^H \end{bmatrix} \mathcal{O}_{\widehat{A},\widehat{B}}(-\lambda_i)\mathcal{C}_{\widehat{A},\widehat{C}}(-\lambda_i) \begin{bmatrix} l_0 & l_1 & \dots & l_{k_i-1} \\ & l_0 & \ddots & \vdots \\ & & \ddots & l_1 \\ & & & l_0 \end{bmatrix} \\
&\quad - \begin{bmatrix} r_0^H & r_1^H & \dots & r_{k_i-1}^H \\ & r_0^H & \ddots & \vdots \\ & & \ddots & r_1^H \\ & & & r_0^H \end{bmatrix} \mathcal{O}_{A,B}(-\lambda_i)\mathcal{C}_{A,C}(-\lambda_i) \begin{bmatrix} l_0 & l_1 & \dots & l_{k_i-1} \\ & l_0 & \ddots & \vdots \\ & & \ddots & l_1 \\ & & & l_0 \end{bmatrix} \\
&= - \begin{bmatrix} r_0^H & r_1^H & \dots & r_{k_i-1}^H \\ & r_0^H & \ddots & \vdots \\ & & \ddots & r_1^H \\ & & & r_0^H \end{bmatrix} \begin{bmatrix} E_{k_i}^H & \dots & E_{2k_i-1}^H \\ \vdots & \ddots & \vdots \\ E_1^H & \dots & E_{k_i}^H \end{bmatrix} \begin{bmatrix} l_0 & l_1 & \dots & l_{k_i-1} \\ & l_0 & \ddots & \vdots \\ & & \ddots & l_1 \\ & & & l_0 \end{bmatrix}. \quad (4.19)
\end{aligned}$$

We are now ready to show (4.16). Since (4.14) and (4.15) hold, the left-hand side

of (4.16) satisfies

$$\begin{aligned}
&\begin{bmatrix} r_0^H & r_1^H & \dots & r_{2k_i-1}^H \\ & r_0^H & \ddots & \vdots \\ & & \ddots & r_1^H \\ & & & r_0^H \end{bmatrix} \begin{bmatrix} E_0^H & E_1^H & \dots & E_{2k_i-1}^H \\ & E_0^H & \ddots & \vdots \\ & & \ddots & E_1^H \\ & & & E_0^H \end{bmatrix} \begin{bmatrix} l_0 & l_1 & \dots & l_{2k_i-1} \\ & l_0 & \ddots & \vdots \\ & & \ddots & l_1 \\ & & & l_0 \end{bmatrix} \\
&= \begin{bmatrix} & \begin{bmatrix} r_0^H & \dots & r_{k_i-1}^H \\ & \ddots & \vdots \\ & & r_0^H \end{bmatrix} \begin{bmatrix} E_{k_i}^H & \dots & E_{2k_i-1}^H \\ \vdots & \ddots & \vdots \\ E_1^H & \dots & E_{k_i}^H \end{bmatrix} \begin{bmatrix} l_0 & \dots & l_{k_i-1} \\ & \ddots & \vdots \\ & & l_0 \end{bmatrix} \\ 0 & \\ & \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & -T_i^H(\nabla_{\widehat{A}}\mathcal{J})^T S_i \\ & \\ 0 & 0 \end{bmatrix},
\end{aligned} \tag{4.20}$$

where the first equality follows from a careful blockwise inspection, and the second equality uses (4.19). Since $\nabla_{\widehat{A}}\mathcal{J} = 0$, it follows that (4.16) holds, and thus (4.13) holds. □

5. Relation with tangential interpolation by projection. The gradient forms of Theorem 3.2 yield the following theorem (proved in [14]) that provides an important link to tangential interpolation by projection.

THEOREM 5.1. *At every stationary point of* $\mathcal{J}$ (2.7) *where* $\widehat{P}$ *and* $\widehat{Q}$ *are invertible, we have the following identities*

$$\widehat{A} = W^T A V, \quad \widehat{B} = W^T B, \quad \widehat{C} = CV, \quad W^T V = I_n \tag{5.1}$$

where $W := -Y\widehat{Q}^{-1}$, $V := X\widehat{P}^{-1}$ *and* X, Y, $\widehat{P}$ *and* $\widehat{Q}$ *satisfy the Sylvester equations (3.3,3.4).*

If we rewrite the above theorem as a projection problem, then we are constructing a projector $\Pi := VW^T$ (implying $W^T V = I_n$) where V and W are given by the following (transposed) Sylvester equations

$$(\widehat{Q}W^T)A + \widehat{A}^T(\widehat{Q}W^T) + \widehat{C}^T C = 0, \quad A(V\widehat{P}) + (V\widehat{P})\widehat{A}^T + B\widehat{B}^T = 0. \tag{5.2}$$

Note that $\widehat{P}$ and $\widehat{Q}$ can be interpreted as normalizations to ensure that $W^T V = I_n$.

Rewriting the Sylvester equations (5.2) as

$$W^T A + (\widehat{Q}^{-1}\widehat{A}\widehat{Q})W^T + (\widehat{C}\widehat{Q}^{-1})C = 0, \tag{5.3a}$$

$$AV + V(\widehat{P}\widehat{A}^T\widehat{P}^{-1}) + B(\widehat{B}^T\widehat{P}^{-1}) = 0, \tag{5.3b}$$

shows the relation with the tangential interpolation described in [8]. There it is shown that when solving two Sylvester equations for the unknowns $W, V \in \mathbb{R}^{N\times n}$

$$W^T A - \Sigma_\mu^T W^T + L^T C = 0, \tag{5.4}$$

$$AV - V\Sigma_\sigma + BR = 0, \tag{5.5}$$

and constructing the reduced-order model (of degree n) as follows

$$(\widehat{A}, \widehat{B}, \widehat{C}) := ((W^T V)^{-1}W^T AV, (W^T V)^{-1}W^T B, CV), \tag{5.6}$$

amounts to a tangential interpolation problem (provided the matrix W^TV is invertible). The "interpolation conditions" (Σ_σ, R) and (Σ_μ, L) (where $\Sigma_\mu, \Sigma_\sigma \in \mathbb{R}^{n\times n}$, $R \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{p\times n}$) are known to uniquely determine the projected system $(\widehat{A}, \widehat{B}, \widehat{C})$ [8]. Moreover, they reproduce exactly the conditions derived in the previous section since they can be expressed in another coordinate system by applying invertible transformations of the type $\left(Q^{-1}\Sigma_\sigma Q, RQ\right)$ and $\left(P^{-1}\Sigma_\mu P, LP\right)$ to the interpolation conditions. This yields transformed matrices VP and WQ but does not affect the transfer function of the reduced-order model $(\widehat{A}, \widehat{B}, \widehat{C})$ (see [8] for more details). The novelty of the derivation in this paper is the case of higher-order poles: the tangential interpolation conditions in Theorem 4.5 contain fewer redundant equations than those that would follow from [8].

6. Low order versus high order poles. In this section we show that $\mathcal{H}_2$-optimal reduced-order models with higher order poles can indeed occur and that in their neighborhood one can expect the tangential interpolation approach to have serious numerical difficulties. We start with a lemma that will allow us to demonstrate this.

LEMMA 6.1. *A stable n-th degree transfer function $\hat{H}(s) = \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}$ is a stationary point of the error function $\|\widehat{H}(s) - H(s)\|^2_{\mathcal{H}_2}$ if and only if $H(s)$ can be realized as follows*

$$A = \begin{bmatrix} \widehat{A} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} \widehat{B} \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} \widehat{C} & C_2 \end{bmatrix}, \tag{6.1}$$

where moreover

$$\widehat{A}\widehat{P} + \widehat{P}\widehat{A}^T + \widehat{B}\widehat{B}^T = 0, \quad A_{21}\widehat{P} + B_2\widehat{B}^T = 0, \tag{6.2}$$

$$\widehat{Q}\widehat{A} + \widehat{A}^T\widehat{Q} + \widehat{C}^T\widehat{C} = 0, \quad \widehat{Q}A_{12} + \widehat{C}^TC_2 = 0. \tag{6.3}$$

Proof. The proof follows from the stationarity conditions in Theorem 3.2. The "if" part is direct: the stationarity conditions hold with $X = \left[\begin{smallmatrix} \widehat{P} \\ 0 \end{smallmatrix}\right]$ and $Y = -\left[\begin{smallmatrix} \widehat{Q} \\ 0 \end{smallmatrix}\right]$. For the "only if" part, the assumption that $\widehat{H}(s)$ is stable and of degree n, guarantees that the matrices $\widehat{P}$ and $\widehat{Q}$ exist and are invertible. Using $Y^TX = -\widehat{P}\widehat{Q}$ one can then always choose a coordinate system for the realization of $H(s)$ in which

$$X = \begin{bmatrix} \widehat{P} \\ 0 \end{bmatrix}, \quad Y = -\begin{bmatrix} \widehat{Q} \\ 0 \end{bmatrix}$$

and hence

$$W = X\widehat{P}^{-1} = \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \quad V = -Y\widehat{Q}^{-1} = \begin{bmatrix} I_n \\ 0 \end{bmatrix}.$$

Therefore we have $A_{11} = \widehat{A},\ B_1 = \widehat{B},\ C_1 = \widehat{C}$. ∎

The special coordinate system of Lemma 6.1 can be used to construct a transfer function $H(s)$ for which a *given* $\widehat{H}(s)$ is the best $\mathcal{H}_2$ norm approximation of $H(s)$.

THEOREM 6.2. *Let $\hat{H}(s) = \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}$ be a given stable n-th degree transfer function, then there always exists a stable N-th degree transfer function $H(s) =$*

$C(sI_N - A)^{-1}B$ with $N > n$, for which $\widehat{H}(s)$ is a stationary point of the $\mathcal{H}_2$ error function.

Proof. It suffices to construct $\widehat{P}$ and $\widehat{Q}$ satisfying the Lyapunov equations in (6.2) and (6.3), and then choose $A_{21} = -B_2\widehat{B}^T\widehat{P}^{-1}$ and $A_{12} = -\widehat{Q}^{-1}\widehat{C}^T C_2$ to satisfy the conditions of Lemma 6.1. Notice that this always has a solution since $\widehat{P}$ and $\widehat{Q}$ are invertible because $\widehat{H}(s)$ is stable and minimal. In order to guarantee that $H(s)$ is also stable, one needs to choose the remaining degrees of freedom, i.e. A_{22}, B_2 and C_2 to satisfy this condition. This can be achieved in several ways, but the simplest one is to choose A_{22} stable, and the matrices B_2 and C_2 sufficiently small. The matrices $A_{21} = -B_2\widehat{B}^T\widehat{P}^{-1}$ and $A_{12} = -\widehat{Q}^{-1}\widehat{C}^T C_2$ will then also be small, and A will then be essentially block diagonal and hence stable. □

The above theorem does not show that the constructed stationary point is also a local minimum, but the following example shows that this is not too difficult to construct. Choose $\widehat{H}(s) = 1/(s-a)^3$ with $a = -1$ and a realization

$$\widehat{A} = \begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}, \widehat{B} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \widehat{C} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

then the realization

$$A = \begin{bmatrix} a & 1 & 0 & f \\ 0 & a & 1 & e \\ 0 & 0 & a & d \\ d & e & f & g \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 1 \\ h \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 & h \end{bmatrix}$$

with $d = -1/2$, $e = 3/4$, $f = -3/8$, $g = -10$, $h = 1/16$, is stable and satisfies the stationarity conditions of Lemma 6.1. Moreover, 1000 random perturbations of the stationary point $\widehat{H}(s)$ clearly indicate that this is a local minimum of the error function $\|H - \widehat{H}\|^2_{\mathcal{H}_2}$.

This example shows that if we aim for an $\mathcal{H}_2$-optimal reduced-order model $\widehat{H}(s)$ with multiple poles, the model reduction technique that restricts itself to first-order poles will not be able to produce that solution. However, what happens if we perturb $H(s)$ or $\widehat{H}(s)$? What can we say about the mapping from one to the other? This is addressed in the following theorem, which shows that if $\widehat{H}(s)$ is a stationary point of the $\mathcal{H}_2$-distance to $H(s)$, then every sufficiently nearby transfer function $\widehat{H}_\Delta(s)$ is a stationary point of a nearby system $H_\Delta(s)$.

THEOREM 6.3. *Let $\hat{H}(s) = \widehat{C}(sI_n - \widehat{A})^{-1}\widehat{B}$ and $H(s) = C(sI_N - A)^{-1}B$ be stable and minimal transfer functions such that $\widehat{H}(s)$ is a stationary point (resp., nondegenerate local minimum) of the error function $\|H(s) - \widehat{H}(s)\|^2_{\mathcal{H}_2}$. Then, for every neighborhood $\mathcal{U}$ of $H(s)$ in $\mathrm{Rat}^n_{p,m}$, there exists a neighborhood $\hat{\mathcal{U}}$ of $\widehat{H}(s)$ in $\mathrm{Rat}^N_{p,m}$ such that, for all $\widehat{H}_\Delta(s) \in \hat{\mathcal{U}}$, there exists $H_\Delta(s) \in \mathcal{U}$ for which $\widehat{H}_\Delta(s)$ is a stationary point (resp., nondegenerate local minimum) of the $\mathcal{H}_2$-distance to $H_\Delta(s)$.*

Proof. The proof consists of constructing a continuous mapping ψ from a neighborhood $\mathcal{V}$ of $\widehat{H}(s)$ in $\mathrm{Rat}^n_{p,m}$ into $\mathrm{Rat}^N_{p,m}$ such that $\widehat{H}_\Delta(s)$ is a stationary point of the $\mathcal{H}_2$-distance to $\psi(\widehat{H}_\Delta(s))$ for all $\widehat{H}_\Delta(s)$ in $\mathcal{V}$. We use Lemma 6.1 to do this. Let $(\widehat{A}_\Delta, \widehat{B}_\Delta, \widehat{C}_\Delta)$ be a nearby realization of the nearby system $\widehat{H}_\Delta(s)$. The solution $\widehat{P}_\Delta$ and $\widehat{Q}_\Delta$ of the perturbed Lyapunov equations in (6.2) and (6.3), will be close to $\widehat{P}$ and $\widehat{Q}$ by continuity of the solution of a non-singular system of equations. For the same reason we can construct nearby solutions $A_{21\Delta} = -B_2\widehat{B}^T_\Delta\widehat{P}^{-1}_\Delta$ and $A_{12\Delta} = -\widehat{Q}^{-1}_\Delta\widehat{C}^T_\Delta C_2$ to finally yield a realization

$$A_\Delta = \begin{bmatrix} \widehat{A}_\Delta & A_{12\Delta} \\ A_{21\Delta} & A_{22} \end{bmatrix}, \quad B_\Delta = \begin{bmatrix} \widehat{B}_\Delta \\ B_2 \end{bmatrix}, \quad C_\Delta = \begin{bmatrix} \widehat{C}_\Delta & C_2 \end{bmatrix},$$

for a transfer function $H_\Delta(s) =: \psi(\widehat{H}_\Delta(s))$ which is close to $H(s)$ and satisfies the conditions of Lemma 6.1. Since, in view of its expression (2.1), the $\mathcal{H}_2$-norm error function is locally smooth in terms of the coefficients of system parameters of $H(s)$ and $\widehat{H}(s)$, every stationary point that is a nondegenerate local minimum remains a local minimum for sufficiently small perturbations. The proof therefore applies to such points. □

Theorem 6.3 implies that the set of full-order models $H(s)$ that have $\mathcal{H}_2$-stationary reduced-order models with only simple poles, is open and dense in $\mathrm{Rat}_{p,m}^N$. This follows from the following reasoning. From the continuity of the mapping from $H(s)$ to $\widehat{H}(s)$ and from the fact that the set of systems with only simple poles is open, it follows that, around a system $H(s)$ with reduced-order models with only simple poles, there is a neighborhood of systems with reduced-order models with only simple poles. If $H(s)$ has a reduced-order model $\widehat{H}(s)$ with multiple poles, then, because the "reduction" map is an open map and the set of systems with only simple poles has an empty interior, it follows that any neighborhood of $H(s)$ contains a full-order model with a reduced-order model with only simple poles. One could conclude from this that one needs only consider first-order interpolation techniques, but, when one approaches a system for which the target function $\widehat{H}(s)$ has higher order poles, the first-order conditions obtained from a minimal parameterization, will become linearly dependent and they will no longer define the reduced-order model accurately.

We illustrate this with the 4th order SISO example given earlier and an iterative scheme based on a fixed point iteration. One can indeed view (3.3,3.4) and (5.1) as two coupled systems of equations

$$(X, Y, \widehat{P}, \widehat{Q}) = F(\widehat{A}, \widehat{B}, \widehat{C}) \quad \text{and} \quad (\widehat{A}, \widehat{B}, \widehat{C}) = G(X, Y, \widehat{P}, \widehat{Q})$$

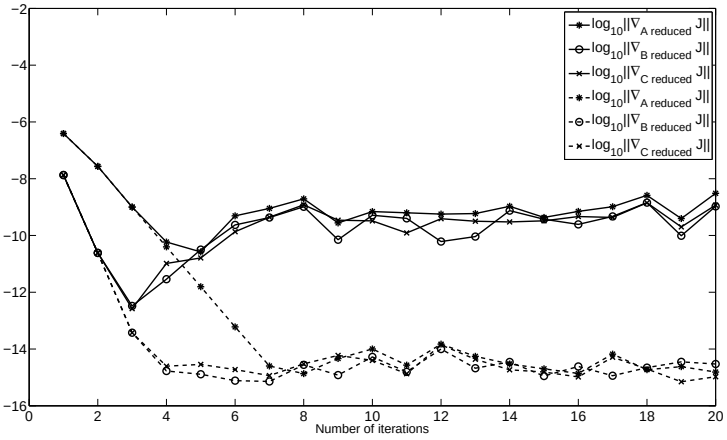
for which we have a fixed point $(\widehat{A}, \widehat{B}, \widehat{C}) = G(F(\widehat{A}, \widehat{B}, \widehat{C}))$ at every stationary point of $\mathcal{J}(\widehat{A}, \widehat{B}, \widehat{C})$. This suggests an iterative procedure of the type

$$(X, Y, \widehat{P}, \widehat{Q})_{i+1} = F(\widehat{A}, \widehat{B}, \widehat{C})_i, \quad (\widehat{A}, \widehat{B}, \widehat{C})_{i+1} = G(X, Y, \widehat{P}, \widehat{Q})_{i+1},$$

which is expected to converge to a nearby fixed point. This is the idea behind the IRKA algorithm of [10] : one first solves for $X, Y, \widehat{P}$ and $\widehat{Q}$ from the Sylvester and Lyapunov equations (3.3,3.4), and then one constructs $V = X\widehat{P}^{-1}$, $W = -Y\widehat{Q}^{-1}$ normalized to satisfy $W^TV = I_n$ and yielding $\widehat{A} = W^TAV$, $\widehat{B} = W^TB$, $\widehat{C} = CV$.

If one chooses $\widehat{A}$ to be diagonal (with diagonal elements $\widehat{\lambda}_i$), the images of the matrices V and W are then essentially spanned by $(A+\widehat{\lambda}_iI)^{-1}B$ and $(A^T+\widehat{\lambda}_iI)^{-1}C^T$, respectively (these are vectors since we are in the SISO case). This amounts to using a minimal representation for the interpolation conditions, where $\widehat{A}$ is in Jordan canonical form. But if two values $\widehat{\lambda}_i$ tend to each other, the corresponding vectors will become linearly dependent, and numerical difficulties can be expected. If, on the other hand, one solves the Sylvester equations in an over-parameterized form, such problems can be avoided. One can for example use the Schur form for $\widehat{A}$ to solve for the matrices $X, Y, \widehat{P}$ and $\widehat{Q}$ at a reasonable additional cost. This does not correspond to a minimal parameterization but allows for a smooth parameterization around defective matrices. Our numerical experiments show that the iteration using the minimal representation is significantly more sensitive, especially as one approaches the reduced order model with higher order poles, since the matrix $\widehat{A}$ tends to become defective.

FIG. 6.1. *Convergence behavior of the log norms of the gradients* $\nabla\mathcal{J}_{\widehat{A}}$, $\nabla\mathcal{J}_{\widehat{B}}$ *and* $\nabla\mathcal{J}_{\widehat{C}}$, *for the minimal parameterization (solid lines) and the Schur parameterization (dashed lines)*



In Figure 6.1 we show the log-norms of the gradients $\nabla\mathcal{J}_{\widehat{A}}$, $\nabla\mathcal{J}_{\widehat{B}}$ and $\nabla\mathcal{J}_{\widehat{C}}$ for both the *minimal* Jordan form (solid lines) and the *over-parameterized* Schur form representations (dashed lines) of the reduced order model. A reduced order model obtained from a balanced truncation is used as the initial guess, which explains why the gradients have small norms since the quality of the initial approximation is quite good. One can see that the first few steps of the fixed point iteration are comparable for both methods (the poles are not too close to each other then) but that the Jordan form iteration loses about seven figures of accuracy as one approaches the optimal solution (and is hence consistent with convergence to a pole of higher order). The error function for the Jordan approach was also worse than that of the Schur approach, but the difference was less explicit. But one should keep in mind that the error functions are significantly more costly to compute than the gradients, and hence are typically not computed for large scale systems. This loss of accuracy and erratic behavior can therefore be significant algorithmically since the gradient norm influences decisions such as termination of the iteration.

7. The discrete-time case. In this section we consider the equivalent formulation in the discrete-time case. We then have the dynamical systems

$$\begin{cases} x_{k+1} = Ax_k + Bu_k \\ y_k = Cx_k \end{cases} \quad \text{and} \quad \begin{cases} \widehat{x}_{k+1} = \widehat{A}\widehat{x}_k + \widehat{B}u \\ \widehat{y}_k = \widehat{C}\widehat{x}_k \end{cases}$$

with transfer functions

$$H(z) = C(zI - A)^{-1}B, \quad \text{and} \quad \widehat{H}(z) = \widehat{C}(zI - \widehat{A})^{-1}\widehat{B}.$$

The squared $\mathcal{H}_2$-norm of the error function $E(z) := H(z) - \widehat{H}(z)$ is then defined as

$$\mathcal{J} := \| E(z) \|_{\mathcal{H}_2}^2 := \operatorname{tr} \int_{-\infty}^{\infty} E(e^{j\omega})E(e^{j\omega})^H \frac{d\omega}{2\pi} = \operatorname{tr} \sum_{k=0}^{\infty} (C_e A_e^k B_e)(C_e A_e^k B_e)^T \quad (7.1)$$

where (A_e, B_e, C_e) defined in (2.4) is again a realization of the error transfer function $E(z)$. The $\mathcal{H}_2$-norm can now be rewritten in terms of the solutions of the Stein equations

$$A_e P_e A_e^T + B_e B_e^T = P_e, \quad A_e^T Q_e A_e + C_e^T C_e = Q_e \tag{7.2}$$

as

$$\mathcal{J} = \operatorname{tr}\left(C_e P_e C_e^T\right) = \operatorname{tr}\left(B_e^T Q_e B_e\right).$$

Partition again the solutions

$$P_e := \begin{bmatrix} P & X \\ X^T & \widehat{P} \end{bmatrix}, \quad Q_e := \begin{bmatrix} Q & Y \\ Y^T & \widehat{Q} \end{bmatrix},$$

to obtain the Stein equations in the form

$$\begin{bmatrix} A & \\ & \widehat{A} \end{bmatrix} \begin{bmatrix} P & X \\ X^T & \widehat{P} \end{bmatrix} \begin{bmatrix} A^T & \\ & \widehat{A}^T \end{bmatrix} + \begin{bmatrix} B \\ \widehat{B} \end{bmatrix} \begin{bmatrix} B^T & \widehat{B}^T \end{bmatrix} = \begin{bmatrix} P & X \\ X^T & \widehat{P} \end{bmatrix},$$
$$\begin{bmatrix} A^T & \\ & \widehat{A}^T \end{bmatrix} \begin{bmatrix} Q & Y \\ Y^T & \widehat{Q} \end{bmatrix} \begin{bmatrix} A & \\ & \widehat{A} \end{bmatrix} + \begin{bmatrix} C^T \\ -\widehat{C}^T \end{bmatrix} \begin{bmatrix} C & -\widehat{C} \end{bmatrix} = \begin{bmatrix} Q & Y \\ Y^T & \widehat{Q} \end{bmatrix}.$$

THEOREM 7.1. *The gradients $\nabla_{\widehat{A}}\mathcal{J}$, $\nabla_{\widehat{B}}\mathcal{J}$ and $\nabla_{\widehat{C}}\mathcal{J}$ of $\mathcal{J} := \|E(s)\|_{\mathcal{H}_2}^2$ are given by*

$$\nabla_{\widehat{A}}\mathcal{J} = 2(\widehat{Q}\widehat{A}\widehat{P} + Y^T A X), \quad \nabla_{\widehat{B}}\mathcal{J} = 2(\widehat{Q}\widehat{B} + Y^T B), \quad \nabla_{\widehat{C}}\mathcal{J} = 2(\widehat{C}\widehat{P} - CX), \tag{7.3}$$

where

$$A^T Y \widehat{A} - C^T \widehat{C} = Y, \quad \widehat{A}^T \widehat{Q} \widehat{A} + \widehat{C}^T \widehat{C} = \widehat{Q}, \tag{7.4}$$
$$\widehat{A} X^T A^T + \widehat{B} B^T = X^T, \quad \widehat{A}\widehat{P}\widehat{A}^T + \widehat{B}\widehat{B}^T = \widehat{P}. \tag{7.5}$$

Setting the gradient of $\mathcal{J}$ to zero yields the stationarity conditions derived in [5]. These are the discrete-time counterpart of Wilson's conditions (see [16] or Theorem 3.2). Again, at a stationary point (where all gradients are zero) we have that the projection matrices

$$W := -Y\widehat{Q}^{-1}, \quad V := X\widehat{P}^{-1}$$

satisfy $\widehat{A} = W^T A V$, $\widehat{B} = W^T B$, $\widehat{C} = CV$, $W^T V = I$ and the Sylvester equations

$$\begin{cases} \widehat{A}^T(\widehat{Q}W^T)A + \widehat{C}^T C = (\widehat{Q}W^T) \\ A(V\widehat{P})\widehat{A}^T + B\widehat{B}^T = (V\widehat{P}) \end{cases}$$

indicating that we are solving a tangential interpolation problem in the inverses of the eigenvalues of $\widehat{A}$, and this both left and right.

The tangential interpolation conditions for the discrete-time case are treated in much the same way for the continuous-time case. We give here immediately the interpolation conditions in terms of the transfer function $H_*(z) := z^{-1}H^T(z^{-1})$:

$$H_*(z) := B^T(I - zA^T)^{-1}C^T = -\sum_{i=0}^{\infty}(\lambda - z)^i B^T A^{T^i}(\lambda A^T - I)^{-i-1}C^T.$$

Since the proof is essentially the same as the one for the continuous-time case, it is omitted.

THEOREM 7.2. *Let* $\widehat{H}(z) = \sum_{i=1}^{\ell} \widehat{H}_i(z)$, $\widehat{H}_i(z) := \widehat{C}_i(zI - \widehat{A}_i)^{-1}\widehat{B}_i^H$ *where* $\{(\widehat{A}_i, \widehat{B}_i^H, \widehat{C}_i) : i = 1, \ldots, \ell\}$ *is a self-conjugate set and* $\widehat{A}_i$ *is just one Jordan block of size* k_i *associated with eigenvalue* $\widehat{\lambda}_i$*, and where* $\widehat{\lambda}_i^{-1}$ *is not a pole of* $H(z)$ *or* $\widehat{H}(z)$*. Then with*

$$\widehat{b}_i(z)^H := \begin{bmatrix} (\widehat{\lambda}_i - z)^{k_i - 1} & \ldots & (\widehat{\lambda}_i - z) & 1 \end{bmatrix} \widehat{B}_i^H,$$

$$\widehat{c}_i(z) := \widehat{C}_i \begin{bmatrix} 1 & (\widehat{\lambda}_i - z) & \ldots & (\widehat{\lambda}_i - z)^{k_i - 1} \end{bmatrix}^T,$$

we have

$$[H_*^T(z) - \widehat{H}_*^T(z)]\widehat{c}_i(z) = O(\widehat{\lambda}_i - z)^{k_i}, \tag{7.6}$$

$$\widehat{b}_i(z)^H[H_*^T(z) - \widehat{H}_*^T(z)] = O(\widehat{\lambda}_i - z)^{k_i}, \tag{7.7}$$

$$\widehat{b}_i(z)^H[H_*^T(z) - \widehat{H}_*^T(z)]\widehat{c}_i(z) = O(\widehat{\lambda}_i - z)^{2k_i}, \tag{7.8}$$

where S_i, T_i *are as defined in* (4.2).

In the case of first-order poles, the conditions reduce to the result derived in [5] in an equivalent form.

8. Conclusion. In this paper, we have characterized the stationary points of the $\mathcal{H}_2$-norm approximation error $\|H(s) - \widehat{H}(s)\|_{\mathcal{H}_2}^2$ in the MIMO case, with the reduced-order system $\widehat{H}(s)$ in Jordan canonical form. The stationarity conditions take the form of tangential interpolation conditions—whose degree depend on the size of the Jordan blocks—written in terms of the Jordan parameters of $\widehat{H}(s)$. The conditions are thus implicit, which calls for iterative algorithms. However, we have shown that the Jordan-based approach becomes ill-conditioned in the neighborhood of target transfer functions $\widehat{H}(s)$ with higher-order poles. It is therefore more robust to use the interpolation conditions in the Sylvester equation form (Theorem 5.1) since the $\mathcal{H}_2$ norm is smooth in the parameters $(\widehat{A}, \widehat{B}, \widehat{C})$ of these equations. The case of discrete-time systems has also been considered.

9. Acknowledgment. The authors would like to thank Samuel Melchior for performing the computations for the example in this paper and the anonymous referees for their constructive comments.

REFERENCES

[1] P. R. Aigrain and E. M. Williams Synthesis of n-reactance networks for desired transient response. J. Appl. Phys., 20:597-600, 1949.

[2] A. C. Antoulas. Approximation of Large-Scale Dynamical Systems. Siam Publications, Philadelphia (2005).

[3] J. Ball, I. Gohberg and L. Rodman. Interpolation of Rational Matrix Functions, Birkhauser Verlag (1990).

[4] C. Beattie and S. Gugercin. Krylov-based minimization for optimal H_2 model reduction. Proceedings 46th IEEE Conference on Decision and Control, 2007.

[5] A. Bunse-Gerstner, D. Kubalinska, G. Vossen, and D. Wilczek. $\mathcal{H}_2$-norm optimal model reduction for large-scale discrete dynamical MIMO systems. Internal Report Bremen University, 2007.
[6] C. Byrnes and P. Falb. Applications of algebraic geometry in systems theory. American Journal of Mathematics, 101(2):337-363, April 1979.
[7] K. Gallivan, A. Vandendorpe, and P. Van Dooren. Sylvester equations and projection-based model reduction. J. Comp. Appl. Math., 162:213-229, 2004.
[8] K. Gallivan, A. Vandendorpe, and P. Van Dooren. Model reduction of MIMO systems via tangential interpolation. SIAM J. Matrix Anal. Appl., 26(2):328-349, 2004.
[9] S. Gugercin. Projection methods for model reduction of large-scale linear dynamical systems. PhD Thesis, ECE Dept., Rice Univ., December 2002.
[10] S. Gugercin, A. Antoulas and C. Beattie. H_2 model reduction for large-scale linear dynamical systems. SIAM J. Matrix Anal. Appl., 30:609-638, 2008.
[11] Y. Halevi. Projection properties of the L_2 optimal reduced order model. Int. J. Control, 79(4):298-310, 2006.
[12] D. C. Hyland and D. S. Bernstein. The optimal projection equations and the relationships between the methods of Wilson, Skelton and Moore. IEEE Trans. Aut. Contr., AC30:1201-1211, 1985.
[13] L. Meier and D. Luenberger. Approximation of linear constant systems. IEEE Trans. Aut. Contr., 12:585-588, 1967.
[14] P. Van Dooren, K. Gallivan and P.-A. Absil. H_2-optimal model reduction of MIMO systems. Appl. Math. Lett., 21(12):1267-1273, 2008.
[15] P. Van Dooren, K. Gallivan and P.-A. Absil. $\mathcal{H}_2$-optimal approximation of MIMO linear dynamical systems. ArXiv paper arXiv:0807.4807v1, July 2008.
[16] D. A. Wilson, Optimum solution of model reduction problem, Proc. Inst. Elec. Eng., 117:11611165, 1970.
[17] J. H. Wilkinson, The algebraic eigenvalue problem, Clarendon Press (1965).