Zentrum für Technomathematik

Fachbereich 3 – Mathematik und Informatik

Necessary optimality conditions for $\mathcal{H}_2$-norm optimal model reduction

Georg Vossen, Angelika Bunse-Gerstner, Dorota Kubalińska, Daniel Wilczek

Report 07–05

Berichte aus der Technomathematik

Report 07–05 Oktober 2007

NECESSARY OPTIMALITY CONDITIONS FOR $\mathcal{H}_2$-NORM OPTIMAL MODEL REDUCTION

G. VOSSEN, A. BUNSE-GERSTNER, D. KUBALINSKA AND D. WILCZEK

Abstract. This paper deals with $\mathcal{H}_2$-norm optimal model reduction for linear time invariant continuous MIMO systems. We will give an overview on several representations of linear systems in state space as well as in Laplace space and discuss the $\mathcal{H}_2$-norm for continuous MIMO systems with multiple poles. On this basis, necessary optimality conditions for the $\mathcal{H}_2$-norm optimal model reduction problem are developed.

Key words. Model reduction, $\mathcal{H}_2$-norm, necessary optimality conditions

AMS subject classifications.

1. Problem formulation. Consider the following linear time invariant (LTI) descriptor system

$$\Sigma := \left(\begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right) := \left\{ \begin{array}{l} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = Cx(t), \end{array} \right\} \tag{1.1}$$

where $x \in \mathbb{C}^n$, $u \in \mathbb{C}^m$ and $y \in \mathbb{C}^p$ are called the *state variable*, the *input variable* and the *output variable*, respectively. The matrices $A \in \mathbb{C}^{n,n}$, $B \in \mathbb{C}^{n,m}$ and $C \in \mathbb{C}^{p,n}$ are constant matrices w.r.t. the time variable t. The system Σ is referred to as stable if all eigenvalues of A are in the left half complex plane (LHP), i.e., all eigenvalues λ_j of A satisfy $\mathrm{Re}(\lambda_j) < 0$. Let us now introduce the *reachability matrix* $\mathcal{R}_n$ and the *observability matrix* $\mathcal{O}_N$ defined by

$$\begin{aligned} \mathcal{R}_n(\Sigma) &= [B, AB, \dots, A^{n-1}B] \in \mathbb{C}^{n,nm}, \\ \mathcal{O}_n(\Sigma) &= [C^*, A^*C^*, \dots, (A^*)^{n-1}C^*]^* \in \mathbb{C}^{pn,n}. \end{aligned}$$

A system is called reachable, resp., observable if $\mathcal{R}_n$, resp., $\mathcal{O}_n$ has full rank n. In this paper we assume that all occuring systems are stable, reachable and observable.

The goal of model reduction is to find a reduced system

$$\hat{\Sigma}: \quad \dot{\hat{x}}(t) = \hat{A}\hat{x}(t) + \hat{B}u(t), \qquad \hat{y}(t) = \hat{C}\hat{x}(t) \tag{1.2}$$

with $\hat{x} \in \mathbb{C}^r$, $\hat{A} \in \mathbb{C}^{r,r}$, $\hat{B} \in \mathbb{C}^{r,m}$ and $\hat{C} \in \mathbb{C}^{p,r}$ with the property that a certain norm of the so-called *error system*

$$\Sigma - \hat{\Sigma} = \left(\begin{array}{cc|c} A & 0 & B \\ 0 & \hat{A} & \hat{B} \\ \hline C & -\hat{C} & 0 \end{array} \right)$$

is small. In this paper, we are interested in optimal model reduction with respect to the $\mathcal{H}_2$-norm of the system which is defined as follows. The Laplace transform

$$\mathcal{L}\{f(t)\}(s) = \int_0^\infty f(t)e^{-st}\,dt$$

applied to the system (1.1) leads to a purely algebraic system of equations in the frequency domain:

$$sX(s) - x(0) = AX(s) + BU(s), \quad Y(s) = CX(s) \tag{1.3}$$

where $X(s) = \mathcal{L}\{x(t)\}(s)$, $U(s) = \mathcal{L}\{u(t)\}(s)$ and $Y(s) = \mathcal{L}\{y(t)\}(s)$ denote the Laplace transform of the state, input and output, respectively. For simplicity, we assume $x(0) = 0$. Solving system (1.3) for Y leads to

$$Y(s) = H(s)U(s)$$

where

$$H(s) = C(sI_n - A)^{-1}B, \quad s \in \mathbb{C} \tag{1.4}$$

with I_n being the n-th order identity matrix is called the *transfer function.* The $\mathcal{H}_2$-norm of the system is Σ is defined as the $\mathcal{H}_2$-norm of its transfer function H given by

$$||H||^2_{\mathcal{H}_2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{trace}\left(H(\mathrm{i}w)^* H(\mathrm{i}w)\right) dw, \tag{1.5}$$

cf., e.g., [1] where i is the imaginary number with $\mathrm{i}^2 = -1$. Hence, the aim of $\mathcal{H}_2$-norm optimal model reduction is

$$\underset{\hat{\Sigma}}{\text{minimize}} \quad J(\hat{\Sigma}) := ||\Sigma - \hat{\Sigma}||^2_{\mathcal{H}_2}. \tag{1.6}$$

or equivalently,

$$\underset{\hat{\Sigma}}{\text{minimize}} \quad J(\hat{H}) = ||H - \hat{H}||^2_{\mathcal{H}_2}. \tag{1.7}$$

We note that representation (1.1) of the system Σ is not unique as the state basis can be transformed by $x = S\bar{x}$ with any regular matrix $S \in \mathbb{C}^{n,n}$ which yields

$$\dot{\bar{x}}(t) = \bar{A}\bar{x}(t) + \bar{B}u(t), \quad y(t) = \bar{C}\bar{x}(t) \tag{1.8}$$

with $\bar{A} = S^{-1}AS$, $\bar{B} = S^{-1}B$ and $\bar{C} = CS$ where A and $\bar{A}$ have the same eigenvalues. However, the transfer function for (1.8) is the same as for (1.1). Therefore, it is common to identify the system not with its matrices A, B and C but with its transfer function H and hence, also to speak of H as a system.

Definition 1.1 (Real system). *A system H is called real if there exist real matrices A, B and C such that $H = C(s\,\mathrm{Id} - A)^{-1}B$ holds where* Id *is the identity mapping.*

In the following, we will also use the function $\tilde{H}$ defined by

$$\tilde{H}(s) := H(s^*)^* = B^*(sI_n - A^*)^{-1}C^*, \quad s \in \mathbb{C}. \tag{1.9}$$

2. Properties of a transfer function.

2.1. SISO. We will now introduce different representations of a transer function H. On the one hand, H can be written as a quotient of two polynomials:

$$H(s) = \frac{\sum_{k=0}^{n-1} \alpha_k s^k}{\sum_{k=0}^{n} \beta_k s^k}, \quad \beta_n = 1, \tag{2.1}$$

with complex coefficients α_k, $k = 0, \ldots, n-1$ and β_k, $k = 0, \ldots, n$. Here, the zeros of the denominator are the poles of the system, resp., the eigenvalues of the matrix

A in (1.1). On the other hand, around each pole λ_j, $j = 1, \dots, n$, we can expand H into its Laurent series, i.e.,

$$H(s) = \sum_{l=-n}^{\infty} \gamma_{jl}(s - \lambda_j)^l. \tag{2.2}$$

Here, the Laurent series starts at $l = -n$ as H is a rational function with the denominator's polynom degree being equal to n. The coefficients γ_{jl} are called the Laurent coefficients of H at λ_j and $\gamma_{j,-1}$ is called the residue denoted by $\mathrm{Res}(H, \lambda_j)$. The order $-l_0(j)$ of a pole λ_j is defined as the highest index l such that $\gamma_{jl} = 0$ holds for all $l \le l_0(j)$. The case $l_0(j) = -1$ will be referred to as simple pole, for $l_0(j) < -1$ we use the notation multiple pole. We note that $-n \le l_0(j) \le -1$ holds as H is rational.

We will now give a third representation based on partial frations which will be helpful in the following proofs.

2.1.1. Simple poles. In many applications, the system has n different simple poles, i.e., $l_0(j) = -1$ holds for $j = 1, \dots, n$. Therefore, it is worth to investigate this situation in more detail. By expanding the transfer function H into partial fractions, we obtain

$$H(s) = \sum_{j=1}^{n} \frac{\phi_j}{s - \lambda_j}, \qquad \phi_j \neq 0, \quad j = 1, \dots, n, \tag{2.3}$$

where the pairwise different λ_j are the poles of the system and $\phi_j = \mathrm{Res}(H, \lambda_j)$ are the residues at λ_j, $j = 1, \dots, n$. Obviously, the matrices A, B and C defined by

$$A = \mathrm{diag}[\lambda_1, \dots, \lambda_n], \quad B = [1\ \dots\ 1]^*, \quad C = [\phi_1, \dots, \phi_n] \tag{2.4}$$

describe the system (2.3). Hence, the entries c_j of the vector C in the representation (2.4) are the residues of H at the poles λ_j.

Proposition 2.1. *The following three conditions are equivalent:*

(i) H is real,

(ii) α_k, $k = 0, \dots, n-1$, and β_k, $k = 0, \dots, n$, in (2.1) are real,

(iii) the poles λ_j, $j = 1, \dots, n$, and the residues ϕ_j, $j = 1, \dots, n$, in (2.3) appear in conjugate pairs, i.e., if λ_j is a pole with residue ϕ_j, then λ_j^* is a pole with residue ϕ_j^*.

Proof: Implication (i)⇒(ii) is obvious. For implication (ii)⇒(iii), equation (2.1) yields

$$\overline{H(\bar{s})} = \overline{\frac{\sum_{k=0}^{n-1} \alpha_k \bar{s}^k}{\sum_{k=0}^{n} \beta_k \bar{s}^k}} = \frac{\sum_{k=0}^{n-1} \bar{\alpha}_k s^k}{\sum_{k=0}^{n} \bar{\beta}_k s^k} = H(s)$$

as α_k and β_k are real. On the other hand, (2.3) implies

$$\overline{H(\bar{s})} = \overline{\sum_{j=1}^{n} \frac{\phi_j}{\bar{s} - \lambda_j}} = \sum_{j=1}^{n} \frac{\bar{\phi}_j}{s - \bar{\lambda}_j}.$$

Hence, we obtain

$$H(s) = \sum_{j=1}^{n} \frac{\phi_j}{s - \lambda_j} = \sum_{j=1}^{n} \frac{\bar{\phi}_j}{s - \bar{\lambda}_j}$$

and comparison of coefficients leads to (iii). For implication (iii)⇒(i) we will now give a representation with real matrices A, B and C if poles and residues of H appear in conjugate pairs. Without loss of generality, we consider the case $n = 2$. The proof for $n > 2$ can be done blockwise. If we have two different real poles with – due to condition (iii) – real residues, one can take representation (2.4). For nonreal poles $\lambda_1 = \lambda_2^* =: \lambda$ with residues $\phi_1 = \phi_2^* =: \phi$ the transfer function is given by

$$H(s) = \frac{\phi}{s-\lambda} + \frac{\phi^*}{s-\lambda^*}.$$

Consider the system represented by the matrices

$$A = \begin{pmatrix} \operatorname{Re}\lambda & \operatorname{Im}\lambda \\ -\operatorname{Im}\lambda & \operatorname{Re}\lambda \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad C = (\operatorname{Re}\phi - \operatorname{Im}\phi, \operatorname{Re}\phi + \operatorname{Im}\phi). \tag{2.5}$$

Its transfer function is given by

$$\begin{aligned} C(sI_2 - A)^{-1}B &= (\operatorname{Re}\phi - \operatorname{Im}\phi, \operatorname{Re}\phi + \operatorname{Im}\phi) \begin{pmatrix} s - \operatorname{Re}\lambda & -\operatorname{Im}\lambda \\ \operatorname{Im}\lambda & s - \operatorname{Re}\lambda \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= 2\,\frac{(\operatorname{Re}\phi)(s - \operatorname{Re}\lambda) - (\operatorname{Im}\phi)(\operatorname{Im}\lambda)}{(s - \operatorname{Re}\lambda)^2 + (\operatorname{Im}\lambda)^2} \\ &= \frac{\phi}{s-\lambda} + \frac{\phi^*}{s-\lambda^*} = H(s). \end{aligned}$$

This completes the proof. We note that the well-known reachable (and observable) canonical form is also a proof for impication (ii)⇒(i).

The funtion $\tilde{H}$ is given by

$$\tilde{H}(s) = H(s^*)^* = \sum_{j=1}^{n} \frac{\phi_j^*}{s - \lambda_j^*} \tag{2.6}$$

with ϕ_j and λ_j from (2.3). Due to Proposition 2.1, we obtain $\tilde{H} = H$ for real systems.

2.1.2. Multiple poles. In this case, the transfer function has the following form:

$$H(s) = \sum_{j=1}^{N} \sum_{l=1}^{n_j} \frac{\phi_{jl}}{(s-\lambda_j)^l}, \qquad \sum_{j=1}^{N} n_j = n, \quad \phi_{j,n_j} \neq 0, \;\; j = 1, \ldots, N, \tag{2.7}$$

where the pairwise different λ_j, $j = 1, \ldots, N$, are the poles, each of order n_j with corresponding coefficients ϕ_{jl}, $l = 1, \ldots, n_j$. For $1 \le j \le N$, rewrite (2.7) as

$$H(s) = \sum_{k=1}^{N} \sum_{l=1}^{n_k} \frac{\phi_{kl}}{(s-\lambda_k)^l} = \sum_{k=1, k\neq j}^{N} \sum_{l=1}^{n_k} \frac{\phi_{kl}}{(s-\lambda_k)^l} + \sum_{l=1}^{n_j} \frac{\phi_{jl}}{(s-\lambda_j)^l}.$$

The first sum is holomorphic around each λ_j. Therefore, it can be expanded into the Taylor series around λ_j which coincides with all summands of the Laurent series of $H(s)$ around λ_j with nonnegative indices. The second term corresponds to all summands of the Laurent series with negative indices, i.e., $\phi_{jl} = \gamma_{j,-l}$ for $l = 1, \ldots, n_j$. This term is often called principal part of the Laurent series around λ_j and we will therefore denote the coefficients ϕ_{jl}, $l = 1, \ldots, n_j$ as *principal coefficients*. For a closer investigation of the principal coefficients define the Jordan matrices J_j, $j = 1, \ldots, N$, as the $n_j \times n_j$-dimensional matrix with λ_j on each diagonal, ones on the super-diagonal and zeros elsewhere. Due to

$$(sI_n - J_j)^{-1} = \begin{pmatrix} (s-\lambda_j)^{-1} & \dots & (s-\lambda_j)^{-n_j} \\ & \ddots & \vdots \\ 0 & & (s-\lambda_j)^{-1} \end{pmatrix}, \tag{2.8}$$

the following matrices A, B and C describe the system (2.7):

$$\begin{aligned} &A = \operatorname{diag}[J_1, \dots, J_N], \quad B = [e_{n_1}^*, \dots, e_{n_N}^*]^*, \\ &C = [\phi_{1,n_1}, \dots, \phi_{11}, \dots, \phi_{N,n_N}, \dots, \phi_{N,n_1}]. \end{aligned} \tag{2.9}$$

Here, e_j is the j-th unity vector, i.e., $e_j = (0, \dots, 0, 1)^* \in \mathbb{R}^j$. Therefore, the entries of the matrix C in representation (2.9) are the principal coefficients ϕ_{jl} in representaion (2.7) of H.

Proposition 2.2. *The following three conditions are equivalent:*

(i) H is real,

(ii) α_k, $k = 0, \dots, n-1$, and β_k, $k = 0, \dots, n$, in (2.1) are real,

(iii) the poles λ_j, $j = 1, \dots, N$, and corresponding principal coefficients ϕ_{jl}, $j = 1, \dots, N$, $l = 1, \dots, n_j$, in (2.7) appear in conjugate pairs, i.e., if λ_j is a pole with coefficients ϕ_{jl}, then λ_j^* is a pole with coefficients ϕ_{jl}^*.

Proof: Again, (i)⇒(ii) is obvious and (ii)⇒(iii) was already proven in Proposition 2.1. For (iii)⇒(i), it is without loss of generality sufficient to find real matrices A, B and C which represent the following transfer function H with two nonreal poles $\lambda_1 = \lambda_2^* =: \lambda$ and corresponding principal coefficients $\phi_{11} = \phi_{21}^* =: \phi$, resp., $\phi_{12} = \phi_{22}^* =: \psi$:

$$H(s) = \frac{\phi}{s-\lambda} + \frac{\phi^*}{s-\lambda^*} + \frac{\psi}{(s-\lambda)^2} + \frac{\psi^*}{(s-\lambda^*)^2}.$$

Consider the system represented by

$$A = \begin{pmatrix} A_1 & I_2 \\ 0 & A_1 \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \quad C = (C_1, C_2).$$

where

$$\begin{aligned} &A_1 = \begin{pmatrix} \operatorname{Re}\lambda & \operatorname{Im}\lambda \\ -\operatorname{Im}\lambda & \operatorname{Re}\lambda \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\ &C_1 = (\operatorname{Re}\phi - \operatorname{Im}\phi, \operatorname{Re}\phi + \operatorname{Im}\phi), \quad C_2 = (\operatorname{Re}\psi - \operatorname{Im}\psi, \operatorname{Re}\psi + \operatorname{Im}\psi). \end{aligned}$$

Its transfer function is

$$\begin{aligned} C(sI_4 - A)^{-1}B &= C \begin{pmatrix} (sI_2 - A_1)^{-1} & (sI_2 - A_1)^{-2} \\ 0 & (sI_2 - A_1)^{-1} \end{pmatrix} B \\ &= C_1 (sI_2 - A_1)^{-2} B_2 + C_2 (sI_2 - A_1)^{-1} B_2 \\ &= 2\,\frac{(\operatorname{Re}\phi)(s - \operatorname{Re}\lambda) - (\operatorname{Im}\phi)(\operatorname{Im}\lambda)}{(s-\operatorname{Re}\lambda)^2 + (\operatorname{Im}\lambda)^2} \\ &\quad + 2\,\frac{(\operatorname{Re}\psi)\big((s-\operatorname{Re}\lambda)^2 - \operatorname{Im}\lambda\big) - 2(\operatorname{Im}\psi)(\operatorname{Im}\lambda)(s-\operatorname{Re}\lambda)}{\big((s-\operatorname{Re}\lambda)^2 + (\operatorname{Im}\lambda)^2\big)^2} \end{aligned}$$

which completes the proof.

The funtion $\tilde{H}$ is given by

$$\tilde{H}(s) = \sum_{j=1}^{N} \sum_{l=1}^{n_j} \frac{\phi_{jl}^*}{(s - \lambda_j^*)^l}.$$

Equivalently to the SISO case, Proposition 2.2 yields $\tilde{H} = H$ for real systems.

2.2. MIMO. Now, B and C are matrices:

$$B = [B^1 \; \dots \; B^m], \quad C = [(C^1)^* \; \dots \; (C^p)^*]^*,$$

with column vectors $B^i = (B_1^i, \dots, B_n^i)^T \in \mathbb{C}^n$ representing the i-th input for $i = 1, \dots, m$ and row vectors $C^k = (C_1^k, \dots, C_n^k) \in \mathbb{C}^n$ representing the k-th output for $k = 1, \dots, p$ of the system. Therefore, the transfer function H is a $(p \times m)$-dimensional matrix with components H_{ki}, $i = 1, \dots, m$, $k = 1, \dots, p$:

$$H(s) = \begin{pmatrix} H_{11}(s) & \dots & H_{1m}(s) \\ \dots & \ddots & \dots \\ H_{p1}(s) & \dots & H_{pm}(s) \end{pmatrix}, \quad H_{ki}(s) = C^k (sI_n - A)^{-1} B^i, \tag{2.10}$$

and each H_{ki} can be seen as a SISO transfer function with input B^i and output C^k.

2.2.1. Simple poles. Here, each H_{ki} has the form

$$H_{ki}(s) = \sum_{j=1}^{n} \frac{\phi_j^{ki}}{s - \lambda_j}, \tag{2.11}$$

where λ_j, $j = 1, \dots, n$, are the poles of H_{ki} (which are the same for each H_{ki}) and ϕ_j^{ki}, $j = 1, \dots, n$ are the residues of λ_j. We note that for $p > 1$ and $m > 1$ it is not possible to give representations A, B and C for the system as in (2.4) with B being a constant vector. Hence, the residues now depend on both B and C. Consider a representation where A is diagonalized, i.e., $A = \mathrm{diag}[\lambda_1, \dots, \lambda_n]$. By definition of H, each component H_{ki} is given by

$$H_{ki}(s) = \sum_{j=1}^{n} \frac{C_j^k B_j^i}{s - \lambda_j}$$

and comparison of coefficients yields that the residues satisfiy

$$\phi_j^{ki} = C_j^k B_j^i, \qquad i = 1, \dots, m, \quad k = 1, \dots, p, \quad j = 1, \dots, n. \tag{2.12}$$

2.2.2. Multiple poles. Here, the entries H_{ki} of H have the following representations:

$$H_{ki}(s) = \sum_{j=1}^{N} \sum_{l=1}^{n_j} \frac{\phi_{jl}^{ki}}{(s - \lambda_j)^l}, \quad \sum_{j=1}^{N} n_j = n \tag{2.13}$$

with ϕ_{jl}^{ki} being coefficients in the principal part of Laurent series of H_{ki} around the poles λ_j of order n_j, $j = 1 \dots, N$. For a better understanding of the residues we consider a Jordan representation of A, i.e. $A = \mathrm{diag}[J_1, \dots, J_N]$ (cf. the SISO case).

We will first investigate the case of one pole λ of order n. Then, we have $A = J$ and (2.8) leads to a transfer function with components

$$H_{ki}(s) = \sum_{l=1}^{n} C_l^k \sum_{q=l}^{n} \frac{B_q^i}{(s-\lambda)^{q-l+1}} = \sum_{l=1}^{n} \frac{\sum_{q=1}^{n-l+1} C_q^k B_{q+l-1}^i}{(s-\lambda)^l}. \tag{2.14}$$

For $N > 1$ poles, we divide $B = (\tilde{B}_1^*, \dots, \tilde{B}_N^*)^*$ with matrices $\tilde{B}_j \in \mathbb{C}^{n_j,m}$ and $C = (\tilde{C}_1, \dots, \tilde{C}_N)$ with matrices $\tilde{C}_j \in \mathbb{C}^{p,n_j}$ for $j = 1, \dots, N$. Due to $H(s) = \sum_{j=1}^{N} \tilde{C}_j (sI_n - J_j)^{-1} \tilde{B}_j$, equation (2.14) implies that each component of H is given by

$$H_{ki}(s) = \sum_{j=1}^{N} \sum_{l=1}^{n_j} \frac{\sum_{q=1}^{n-l+1} (\tilde{C}_j)_q^k (\tilde{B}_j)_{q+l-1}^i}{(s-\lambda_j)^l}.$$

Hence, we obtain the following representation for the residues:

$$\phi_{jl}^{ki} = \sum_{q=1}^{n-l+1} (\tilde{C}_j)_q^k (\tilde{B}_j)_{q+l-1}^i. \tag{2.15}$$

3. The $\mathcal{H}_2$-norm. We will now give explicite formulas how to derive the $\mathcal{H}_2$-norm of a given system H. Due to (1.9), the $\mathcal{H}_2$-norm becomes to

$$||H||_{\mathcal{H}_2}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{trace}\left(\tilde{H}(-\mathrm{i}w) H(\mathrm{i}w)\right) dw = \frac{1}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \operatorname{trace}\left(\tilde{H}(-z) H(z)\right) dz. \tag{3.1}$$

3.1. SISO. As in this case H is scalar, (3.1) simplifies to

$$||H||_{\mathcal{H}_2}^2 = \frac{1}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \tilde{H}(-z) H(z) dz. \tag{3.2}$$

We will calculate this complex integral via the classical Residue Theorem from complex analysis (cf., e.g. [2]) which we state here in a simple version suitable for our purpose.

THEOREM 3.1 (Residue Theorem). *Let $F : \mathbb{C} \to \mathbb{C}$ be a complex meromorphic function and γ be Jordan curve, i.e., a simple closed curve, in $\mathbb{C}$ such that F has no poles on γ. Then, the following holds:*

$$\frac{1}{2\pi\mathrm{i}} \int_\gamma F(z) dz = \sum_{\lambda \text{ pole of } F \text{ inside } \gamma} \operatorname{Res}(F, \lambda). \tag{3.3}$$

Here and in the following, a point is referred to lie "inside" a curve if its *winding number* is equal to one. For example, each point z with $|z| < 1$ is inside the anti-clockwise parametrized unit circle, i.e., $\gamma(t) := e^{2\pi\mathrm{i}t}$, $t \in [0, 1]$. We note that the sum in (3.3) is finite as the function F is meromorphic. The following result will be helpful for the computation of formula (3.2).

PROPOSITION 3.2. *Let $F : \mathbb{C} \to \mathbb{C}$ be a rational function of order $o \leq -2$, i.e.,*

$$F(z) = \frac{\sum_{k=0}^{N} \alpha_k z^k}{\sum_{k=0}^{M} \beta_k z^k}, \quad \alpha_N, \beta_M \neq 0, \quad o := N - M \leq -2,$$

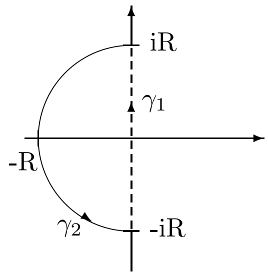

FIG. 3.1. *The curves γ_1 (dashed line) and γ_2 (solid line)*

with no poles on the imaginary axis. Then the following holds:

$$\frac{1}{2\pi \mathrm{i}} \int\limits_{-\mathrm{i}\infty}^{\mathrm{i}\infty} F(z)dz = \sum_{\lambda \text{ pole of } F \text{ in LHP}} \mathrm{Res}(F, \lambda).$$

Proof: Consider for a fixed $R > 0$ the Jordan curve γ defined by

$$\gamma(t) := \begin{cases} \gamma_1(t), \quad t \in [0,1], & \gamma_1(t) := -\mathrm{i}R + 2\mathrm{i}Rt, \quad t \in [0,1], \\ \gamma_2(t-1), \quad t \in [1,2], & \gamma_2(t) := \mathrm{i}Re^{\pi \mathrm{i}t}, \quad t \in [0,1], \end{cases}$$

for $t \in [0,2]$ as shown in Figure 3.1. Then, we have

$$\int\limits_{-\mathrm{i}R}^{\mathrm{i}R} F(z)dz = \int_{\gamma_1} F(z)dz = \int_{\gamma} F(z)dz - \int_{\gamma_2} F(z)dz. \tag{3.4}$$

The first integral can be computed via the Residue theorem:

$$\frac{1}{2\pi \mathrm{i}} \int_{\gamma} F(z)dz = \sum_{\lambda \text{ pole of } F \text{ inside } \gamma} \mathrm{Res}(F, \lambda) \underset{R\to\infty}{\longrightarrow} \sum_{\lambda \text{ pole of } F \text{ in LHP}} \mathrm{Res}(F, \lambda).$$

The second integral in (3.4) is bounded by

$$\left| \int_{\gamma_2} F(z)dz \right| \le |\gamma_2| \max_{z\in\gamma_2} |F(z)| \le \pi R \max_{|z|=R} |F(z)|.$$

For sufficiently large R, there exists a constant c such that $|F(z)| \le c|z|^{N-M}$ holds for $|z| \ge R$ which implies

$$\int_{\gamma_2} F(z)dz \le c\pi R^{N-M+1} \le \frac{c\pi}{R} \underset{R\to\infty}{\longrightarrow} 0.$$

This completes the proof.

Due to (2.1), $\tilde{H}(-z)H(z)$ is a complex rational function of order $o \ge 2$. Since for a stable system H, the function $\tilde{H}(-z)$ has no poles in LHP, the function $\tilde{H}(-z)H(z)$ has the same poles in LHP as $H(z)$ and the $\mathcal{H}_2$-norm becomes to

$$||H||^2_{\mathcal{H}_2} = \sum_{\lambda \text{ pole of } H(z) \text{ in LHP}} \mathrm{Res}\left(\tilde{H}(-z)H(z), z = \lambda\right). \tag{3.5}$$

3.1.1. Simple poles. The following result is well-known in complex analysis and can be found, e.g., in [2].

PROPOSITION 3.3. *Let F be a meromorphic complex function and λ be a simple pole of F with residue ϕ. Let furthermore G be a complex function which is holomorphic in λ. Then, we have* $\operatorname{Res}(F \cdot G, \lambda) = \operatorname{Res}(F, \lambda) G(\lambda) = \phi G(\lambda)$.

This brings us to the main result in this paragraph.

PROPOSITION 3.4. *The $\mathcal{H}_2$-norm of a stable system H with simple poles λ_j and corresponding residues ϕ_j, $j = 1, \dots, n$, is given by*

$$||H||_{\mathcal{H}_2}^2 = \sum_{j=1}^{n} \phi_j H(-\lambda_j^*)^* = \sum_{j=1}^{n} \phi_j^* H(-\lambda_j^*). \tag{3.6}$$

If H is real, we have

$$||H||_{\mathcal{H}_2}^2 = \sum_{j=1}^{n} \phi_j H(-\lambda_j). \tag{3.7}$$

Proof: Using equation (3.5), we directly obtain (3.6) where the second equality arises from the fact that $||\cdot|| = ||\cdot||^*$ holds. Furthermore, Proposition 2.1 yields (3.7).

3.1.2. Multiple poles. Due to (3.5), we have to calculate residues of $\tilde{H}(-z)H(z)$ in λ_j, $j = 1, \dots, N$, which are now poles of order n_j, respectively. We can use the following well-known result, cf., [2].

PROPOSITION 3.5. *Let F be a meromorphic complex function and λ be a pole of order k of F. Then, the residue of the function F in λ is given by*

$$\operatorname{Res}(F, \lambda) = \frac{1}{(k-1)!} \lim_{z \to \lambda} \left(\left((z - \lambda)^k F(z) \right)^{(k-1)} \right). \tag{3.8}$$

where the superscript $(k-1)$ denotes the $(k-1)$-th derivative with respect to z.

In the following, we will calculate the right hand side of equation (3.8) for $F = H$.

LEMMA 3.6. *For $j = 1, \dots, N$ and $l = 0, \dots, n_j - 1$ we have*

$$\lim_{z \to \lambda_j} \left(\left((z - \lambda_j)^{n_j} H(z) \right)^{(n_j - 1 - l)} \right) = \phi_{j,l+1} (n_j - 1 - l)! \tag{3.9}$$

Proof: The term $\left((z - \lambda_j)^{n_j} H(z) \right)^{(n_j - 1 - l)}$ can be written as

$$\left(\sum_{i=1, i \neq j}^{N} \sum_{k=1}^{n_i} \frac{\phi_{ik} (z - \lambda_j)^{n_j}}{(z - \lambda_i)^k} \right)^{(n_j - 1 - l)} + \left(\sum_{k=1}^{n_j} \phi_{jk} (z - \lambda_j)^{n_j - k} \right)^{(n_j - 1 - l)}.$$

Here, the first summand vanishes for $z = \lambda_j$ since

$$\left(\left((z - \lambda_j)^{n_j} (z - \lambda_i)^{-k} \right)^{(n_j - 1 - l)} \right) \Big|_{z = \lambda_j} = 0$$

holds, whereas for the second summand we obtain

$$\left(\sum_{k=1}^{l+1} \phi_{jk} (n_j - 1 - l)! (z - \lambda_j)^{1 + l - k} \right) \Big|_{z = \lambda_j} = \phi_{j,l+1} (n_j - 1 - l)!$$

Now we come to the main result of this paragraph.

PROPOSITION 3.7. *The $\mathcal{H}_2$-norm of a stable system H with simple poles λ_j and corresponding principal coefficients ϕ_{jl}, $j = 1,\dots,N$, $l = 1,\dots,n_j$, is given by*

$$||H||^2_{\mathcal{H}_2} = \sum_{j=1}^{N}\sum_{l=1}^{n_j} \frac{(-1)^{l-1}}{(l-1)!}\phi_{jl} H^{(l-1)}(-\lambda_j^*)^* = \sum_{j=1}^{N}\sum_{l=1}^{n_j} \frac{(-1)^{l-1}}{(l-1)!}\phi_{jl}^* H^{(l-1)}(-\lambda_j^*). \tag{3.10}$$

If H is real, we have

$$||H||^2_{\mathcal{H}_2} = \sum_{j=1}^{N}\sum_{l=1}^{n_j} \frac{(-1)^{l-1}}{(l-1)!}\phi_{jl} H^{(l-1)}(-\lambda_j). \tag{3.11}$$

Proof: Using Proposition 3.5 and Lemma 3.6, we obtain

$$\begin{aligned}
&\mathrm{Res}(\tilde{H}(-z)H(z),\lambda_j)\\
&= \frac{1}{(n_j-1)!}\lim_{z\to\lambda_j}\left(\left((z-\lambda_j)^{n_j}\tilde{H}(-z)H(z)\right)^{(n_j-1)}\right)\\
&= \frac{1}{(n_j-1)!}\lim_{z\to\lambda_j}\left(\sum_{l=0}^{n_j-1}\binom{n_j-1}{l}\left((z-\lambda_j)^{n_j}H(z)\right)^{(n_j-1-l)}\left(\tilde{H}(-z)\right)^{(l)}\right)\\
&= \frac{1}{(n_j-1)!}\sum_{l=0}^{n_j-1}\binom{n_j-1}{l}\lim_{z\to\lambda_j}\left(\left((z-\lambda_j)^{n_j}H(z)\right)^{(n_j-1-l)}\right)\left(\tilde{H}(-z)\right)^{(l)}\Big|_{z=\lambda_j}\\
&= \frac{1}{(n_j-1)!}\sum_{l=0}^{n_j-1}\binom{n_j-1}{l}\phi_{j,l+1}(n_j-1-l)!\,(-1)^l\tilde{H}^{(l)}(-\lambda_j)\\
&= \sum_{l=1}^{n_j}\frac{(-1)^{l-1}}{(l-1)!}\phi_{jl}\tilde{H}^{(l-1)}(-\lambda_j).
\end{aligned}$$

Equation (3.5) directly yields (3.10), and (3.11) can be obtained from Proposition 2.2.

3.2. MIMO. The $\mathcal{H}_2$ norm is given by

$$\begin{aligned}
||H||^2_{\mathcal{H}_2} &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{trace}\left(H(\mathrm{i}w)^*H(\mathrm{i}w)\right)dw = \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_{i,k}\left(H_{ki}(\mathrm{i}w)^*H_{ki}(\mathrm{i}w)\right)dw\\
&= \sum_{i,k}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}H_{ki}(\mathrm{i}w)^*H_{ki}(\mathrm{i}w)dw\right) = \sum_{i,k}||H_{ki}||^2_{\mathcal{H}_2}
\end{aligned} \tag{3.12}$$

and hence, the results from the previous section about SISO systems can be used to compute the $\mathcal{H}_2$-norm of a MIMO system.

3.2.1. Simple poles. Due to (3.6), for each H_{ki} the norm is given by

$$||H_{ki}||^2_{\mathcal{H}_2} = \sum_{j=1}^{n}(\phi_j^{ki})^*H_{ki}(-\lambda_j^*)$$

which, together with (3.12), implies

$$||H||^2_{\mathcal{H}_2} = \sum_{i,k} \sum_{j=1}^{n} (\phi_j^{ki})^* H_{ki}(-\lambda_j^*) = \sum_{j=1}^{n} \sum_{i,k} (\phi_j^{ki})^* H_{ki}(-\lambda_j^*) \tag{3.13}$$

3.2.2. Multiple poles. Due to (3.10) and (3.12), the norm is

$$||H||^2_{\mathcal{H}_2} = \sum_{j=1}^{N} \sum_{l=1}^{n_j} \sum_{i,k} \frac{(-1)^{l-1}}{(l-1)!} (\phi_{jl}^{ki})^* H_{ki}^{(l-1)}(-\lambda_j^*) \tag{3.14}$$

4. Necessary optimality conditions.

4.1. SISO.

4.1.1. Simple poles. As mentioned in the previous section, for simple poles it is possible to find a representation A, B and C of a system such that $B = e_n$ holds and the only free paramters are the poles of Σ on the diagonal of A and the residues in C. Hence, minimization problem (1.7) can be seen as

$$\underset{v}{\text{minimize}} \quad J(v) = ||H - \hat{H}||^2_{\mathcal{H}_2}. \tag{4.1}$$

with the optimization variable

$$v = (\hat{\lambda}_1, \ldots, \hat{\lambda}_r, \hat{\phi}_1, \ldots, \hat{\phi}_r)^T \in \mathbb{C}^{2r}. \tag{4.2}$$

If we assume that the poles of the original system are different from the poles of the reduced system, the error system $H - \hat{H}$ has $n + r$ simple poles $\lambda_1, \ldots, \lambda_n, \hat{\lambda}_1, \ldots, \hat{\lambda}_r$ with corresponding residues $\phi_1, \ldots, \phi_n, -\hat{\phi}_1, \ldots, -\hat{\phi}_r$. Formula (3.6) implies that the cost functional in (4.1) can be written as

$$J(v) = \sum_{j=1}^{n} \phi_j^* \left(H(-\lambda_j^*) - \hat{H}(-\lambda_j^*) \right) + \sum_{j=1}^{r} \hat{\phi}_j^* \left(\hat{H}(-\hat{\lambda}_j^*) - H(-\hat{\lambda}_j^*) \right)$$

which yields a nonsmooth optimization problem with the complex optimization vector $v \in \mathbb{C}^{2r}$. We consider three different types of optimization problems:

Case 1: $$\underset{v}{\text{minimize}} \quad J(v) \quad \text{subject to } v \in \mathbb{R}^{2r}, \tag{4.3}$$

i.e., the optimization vector v shall real. This is a reasonable task if the original system H is real or even has real poles and residues. In this case, due to (3.7), all $*$ superscripts can be deleted in J and we obtain

$$J(v) = \sum_{j=1}^{n} \phi_j \left(H(-\lambda_j) - \hat{H}(-\lambda_j) \right) + \sum_{j=1}^{r} \hat{\phi}_j \left(\hat{H}(-\hat{\lambda}_j) - H(-\hat{\lambda}_j) \right). \tag{4.4}$$

Furthermore, instead of adding the constraint $v \in \mathbb{R}^{2r}$, one can also directly see the problem as a real unconstrained one. Therefore, we obtain a smooth unconstraint optimization problem with the real optimization vector $v \in \mathbb{R}^{2r}$.

THEOREM 4.1. *Necessary optimality conditions for problem (4.3) are given by*

$$H(-\hat{\lambda}_j) = \hat{H}(-\hat{\lambda}_j), \quad H'(-\hat{\lambda}_j) = \hat{H}'(-\hat{\lambda}_j), \qquad j = 1, \ldots, r, \tag{4.5}$$

i.e., interpolation of $\hat{H}$ and its first derivative with H at $-\hat{\lambda}_j$, $j = 1, \ldots, r$.

This result together with a proof was already given by Meier/Luenberger [4] and Gugercin/Antoulas/Beattie [3]. However, we note that in both references the authors did not mention that their proof is only applicable in case 1, i.e., for real poles and residues, as for complex v one has to minimize the cost functional J in (4.1) which is not differentiable w.r.t. the optimization variable. We furthermore note that, nevertheless, in both publications the result is claimed to be correct and used in examples also for nonreal poles and residues. We will now investigate in which cases the result is still true for a complex v and when it should slightly be modified.

Case 2:
$$\underset{v}{\text{minimize}} \quad J(v), \tag{4.6}$$

i.e., there shall be no constraints. The typical approach for such nonsmooth optimization problems with a complex variable $v \in \mathbb{R}^{2r}$ is to consider the real variable $\tilde{v} \in \mathbb{R}^{4r}$ consisting of real and imaginary parts of the components of v as optimization vector in the general case:

$$\tilde{v} = (\operatorname{Re}\hat{\lambda}_1, \operatorname{Im}\hat{\lambda}_1, \ldots, \operatorname{Re}\hat{\lambda}_r, \operatorname{Im}\hat{\lambda}_r, \operatorname{Re}\hat{\phi}_1, \operatorname{Im}\hat{\phi}_1, \ldots, \operatorname{Re}\hat{\phi}_r, \operatorname{Im}\hat{\phi}_r)^T \in \mathbb{R}^{4r} \tag{4.7}$$

Then, J is smooth w.r.t. $\tilde{v}$ and hence, we obtain a smooth unconstrained optimization problem with optimization variable $\tilde{v} \in \mathbb{R}^{4r}$.

For the statement of the main result in case 2, the following lemma will be helpful.

Lemma 4.2.

(i) For $\mu = 1, \ldots, r$ we have:

$$\frac{\partial H(-\lambda_j^*)}{\partial \operatorname{Re}\hat{\phi}_\mu} = \frac{\partial H(-\lambda_j^*)}{\partial \operatorname{Im}\hat{\phi}_\mu} = \frac{\partial H(-\hat{\lambda}_j^*)}{\partial \operatorname{Re}\hat{\phi}_\mu} = \frac{\partial H(-\hat{\lambda}_j^*)}{\partial \operatorname{Im}\hat{\phi}_\mu} = 0,$$
$$\frac{\partial \hat{H}(-\lambda_j^*)}{\partial \operatorname{Re}\hat{\phi}_\mu} = \frac{1}{\mathrm{i}}\frac{\partial \hat{H}(-\lambda_j^*)}{\partial \operatorname{Im}\hat{\phi}_\mu} = \frac{1}{-\lambda_j^* - \hat{\lambda}_\mu}, \quad \frac{\partial \hat{H}(-\hat{\lambda}_j^*)}{\partial \operatorname{Re}\hat{\phi}_\mu} = \frac{1}{\mathrm{i}}\frac{\partial \hat{H}(-\hat{\lambda}_j^*)}{\partial \operatorname{Im}\hat{\phi}_\mu} = \frac{1}{-\hat{\lambda}_j^* - \hat{\lambda}_\mu}.$$

(ii) For $\mu = 1, \ldots, r$ we have:

$$\frac{\partial H(-\lambda_j^*)}{\partial \operatorname{Re}\hat{\lambda}_\mu} = \frac{\partial H(-\lambda_j^*)}{\partial \operatorname{Im}\hat{\lambda}_\mu} = 0, \quad \frac{\partial \hat{H}(-\lambda_j^*)}{\partial \operatorname{Re}\hat{\lambda}_\mu} = \frac{1}{\mathrm{i}}\frac{\partial \hat{H}(-\lambda_j^*)}{\partial \operatorname{Im}\hat{\lambda}_\mu} = \frac{\hat{\phi}_\mu}{(\lambda_j^* + \hat{\lambda}_\mu)^2},$$
$$\frac{\partial H(-\hat{\lambda}_j^*)}{\partial \operatorname{Re}\hat{\lambda}_\mu} = -\frac{1}{\mathrm{i}}\frac{\partial H(-\hat{\lambda}_j^*)}{\partial \operatorname{Im}\hat{\lambda}_\mu} = \begin{cases} 0, & j \neq \mu, \\ -H'(-\hat{\lambda}_j^*), & j = \mu, \end{cases}$$
$$\frac{\partial \hat{H}(-\hat{\lambda}_j^*)}{\partial \operatorname{Re}\hat{\lambda}_\mu} = \begin{cases} \frac{\hat{\phi}_\mu}{(\hat{\lambda}_j^* + \hat{\lambda}_\mu)^2}, & j \neq \mu, \\ -\hat{H}'(-\hat{\lambda}_j^*) + \frac{\hat{\phi}_\mu}{(\hat{\lambda}_j^* + \hat{\lambda}_\mu)^2}, & j = \mu, \end{cases}$$
$$\frac{\partial \hat{H}(-\hat{\lambda}_j^*)}{\partial \operatorname{Im}\hat{\lambda}_\mu} = \begin{cases} \mathrm{i}\frac{\hat{\phi}_\mu}{(\hat{\lambda}_j^* + \hat{\lambda}_\mu)^2}, & j \neq \mu, \\ \mathrm{i}\hat{H}'(-\hat{\lambda}_j^*) - \frac{\hat{\phi}_\mu}{(\hat{\lambda}_j^* + \hat{\lambda}_\mu)^2}, & j = \mu. \end{cases}$$

Theorem 4.3. *Necessary optimality conditions for problem (4.6) are given by*

$$H(-\hat{\lambda}_j^*) = \hat{H}(-\hat{\lambda}_j^*), \quad H'(-\hat{\lambda}_j^*) = \hat{H}'(-\hat{\lambda}_j^*), \qquad j = 1, \ldots, r, \tag{4.8}$$

i.e., interpolation of $\hat{H}$ and its first derivative with H at $-\hat{\lambda}_j^*$, $j = 1, \ldots, r$.

Note that $-\hat{\lambda}_j^*$, $j=1,\ldots,r$, are called *mirror images.*

Proof: In view of Lemma 4.2, differentiating J with respect to $\mathrm{Re}\,\hat{\phi}_\mu$ yields

$$\begin{aligned}\frac{\partial J}{\partial \mathrm{Re}\,\hat{\phi}_\mu} &= -\sum_{j=1}^{n}\phi_j^*\frac{1}{-\lambda_j^*-\hat{\lambda}_\mu} + \hat{H}(-\hat{\lambda}_\mu^*) + \sum_{j=1}^{r}\hat{\phi}_j^*\frac{1}{-\hat{\lambda}_j^*-\hat{\lambda}_\mu} - H(-\hat{\lambda}_\mu^*)\\ &= -\tilde{H}(-\hat{\lambda}_\mu) + \hat{H}(-\hat{\lambda}_\mu^*) + \tilde{\hat{H}}(-\hat{\lambda}_\mu) - H(-\hat{\lambda}_\mu^*)\\ &= -H(-\hat{\lambda}_\mu^*)^* + \hat{H}(-\hat{\lambda}_\mu^*) + \hat{H}(-\hat{\lambda}_\mu^*)^* - H(-\hat{\lambda}_\mu^*)\\ &= 2\mathrm{Re}\left(\hat{H}(-\hat{\lambda}_\mu^*) - H(-\hat{\lambda}_\mu^*)\right).\end{aligned} \tag{4.9}$$

Analogously, we obtain

$$\begin{aligned}\frac{\partial J}{\partial \mathrm{Im}\,\hat{\phi}_\mu} &= -\mathrm{i}H(-\hat{\lambda}_\mu^*)^* - \mathrm{i}\hat{H}(-\hat{\lambda}_\mu^*) + \mathrm{i}\hat{H}(-\hat{\lambda}_\mu^*)^* + \mathrm{i}H(-\hat{\lambda}_\mu^*)\\ &= 2\mathrm{Im}\left(\hat{H}(-\hat{\lambda}_\mu^*) - H(-\hat{\lambda}_\mu^*)\right).\end{aligned} \tag{4.10}$$

In view of equations (4.9) and (4.10), the necessary optimality condition $\partial J/\partial\tilde{z}=0$ implies the first condition in (4.8). Differentiation of J with respect to real, resp., imaginary part of $\hat{\lambda}_\mu$ yields

$$\begin{aligned}\frac{\partial J}{\partial \mathrm{Re}\,\hat{\lambda}_\mu} &= -\sum_{j=1}^{n}\phi_j^*\frac{\hat{\phi}_\mu}{(\lambda_j^*+\hat{\lambda}_\mu)^2} + \sum_{j=1,j\neq\mu}^{r}\hat{\phi}_j^*\frac{\hat{\phi}_\mu}{(\hat{\lambda}_j^*+\hat{\lambda}_\mu)^2}\\ &\quad + \left(-\hat{H}'(-\hat{\lambda}_\mu^*) + \frac{\hat{\phi}_\mu}{(\hat{\lambda}_\mu^*+\hat{\lambda}_\mu)^2}\right)\hat{\phi}_\mu^* + \hat{\phi}_\mu^* H'(-\hat{\lambda}_\mu^*)\\ &= \hat{\phi}_\mu\tilde{H}'(-\hat{\lambda}_\mu) - \hat{\phi}_\mu\tilde{\hat{H}}'(-\hat{\lambda}_\mu) - \hat{\phi}_\mu^*\hat{H}'(-\hat{\lambda}_\mu^*) + \hat{\phi}_\mu^* H'(-\hat{\lambda}_\mu^*)\\ &= 2\mathrm{Re}\left(\hat{\phi}_\mu^*\left(H'(-\hat{\lambda}_\mu^*) - \hat{H}'(-\hat{\lambda}_\mu^*)\right)\right),\\ \frac{\partial J}{\partial \mathrm{Im}\,\hat{\lambda}_\mu} &= -\mathrm{i}\sum_{j=1}^{n}\phi_j^*\frac{\hat{\phi}_\mu}{(\lambda_j^*+\hat{\lambda}_\mu)^2} + \mathrm{i}\sum_{j=1,j\neq\mu}^{r}\hat{\phi}_j^*\frac{\hat{\phi}_\mu}{(\hat{\lambda}_j^*+\hat{\lambda}_\mu)^2}\\ &\quad + \mathrm{i}\left(\hat{H}'(-\hat{\lambda}_\mu^*) + \frac{\hat{\phi}_\mu}{(\hat{\lambda}_\mu^*+\hat{\lambda}_\mu)^2}\right)\hat{\phi}_\mu^* - \hat{\phi}_\mu^* H'(-\hat{\lambda}_\mu^*)\\ &= 2\mathrm{Im}\left(\hat{\phi}_\mu^*\left(H'(-\hat{\lambda}_\mu^*) - \hat{H}'(-\hat{\lambda}_\mu^*)\right)\right).\end{aligned} \tag{4.11}$$

In view (4.11), $\partial J/\partial\tilde{z}=0$ implies the second condition in (4.8) which completes the proof.

REMARK 4.4. *If the resulting system in problem (4.6) for case 2 turns out to be real, corresponding optimality conditions (4.8) are equivalent to conditions (4.5) for case 1.*

Case 3:

$$\underset{v}{\text{minimize}}\quad J(v)\quad\text{subject to }\hat{H}\text{ is real.} \tag{4.12}$$

This case is reasonable if already the original system H is real what we will assume in the following. Now, v can be complex but all poles and residues appear in conjugate pairs as shown in Proposition 2.1. For simplicity of notation, we assume that all components of v are purely imaginary and hence, that $r = 2R$, $R \in \mathbb{N}$, is even. For all real components one may use the results of case 1. Then, (4.12) is equivalent to

$$\underset{v}{\text{minimize}} \quad J(\tilde{v}) \quad \text{subject to} \quad \hat{\lambda}_j = \hat{\lambda}^*_{R+j}, \quad \hat{\phi}_j = \hat{\phi}^*_{R+j}, \quad j = 1, \ldots, R. \tag{4.13}$$

This can be seen as a constrained smooth optimization problem using again $\tilde{v}$ from (4.7) as optimization vector.

Theorem 4.5. *Necessary optimality conditions for problem (4.12) are given by*

$$H(-\hat{\lambda}_j) = \hat{H}(-\hat{\lambda}_j), \quad H'(-\hat{\lambda}_j) = \hat{H}'(-\hat{\lambda}_j), \qquad j = 1, \ldots, r.$$

Proof: The Lagrangian function in normal form for (4.13) is given by

$$\begin{aligned} L(\psi, \chi, v) = J(v) &+ \sum_{j=1}^{R} \operatorname{Re} \psi_j (\operatorname{Re} \hat{\lambda}_j - \operatorname{Re} \hat{\lambda}^*_{R+j}) + \sum_{j=1}^{R} \operatorname{Im} \psi_j (\operatorname{Im} \hat{\lambda}_j + \operatorname{Im} \hat{\lambda}^*_{R+j}) \\ &+ \sum_{j=1}^{R} \operatorname{Re} \chi_j (\operatorname{Re} \hat{\phi}_j - \operatorname{Re} \hat{\phi}^*_{R+j}) + \sum_{j=1}^{R} \operatorname{Im} \chi_j (\operatorname{Im} \hat{\phi}_j + \operatorname{Im} \hat{\phi}^*_{R+j}) \end{aligned}$$

with the complex Lagrange multipliers $\psi = (\psi_1, \ldots, \psi_R)$, $\chi = (\chi_1, \ldots, \chi_R) \in \mathbb{C}^R$. In view of (4.9), for $\mu = 1, \ldots, R$ we obtain

$$\frac{\partial L(\psi, \chi, v)}{\partial \operatorname{Re} \hat{\phi}_\mu} = 2\operatorname{Re} \left(\hat{H}(-\hat{\lambda}^*_\mu) - H(-\hat{\lambda}^*_\mu) \right) + \operatorname{Re} \chi_\mu \tag{4.14}$$

and

$$\begin{aligned} \frac{\partial L(\psi, \chi, v)}{\partial \operatorname{Re} \hat{\phi}_{R+\mu}} &= 2\operatorname{Re} \left(\hat{H}(-\hat{\lambda}^*_{R+\mu}) - H(-\hat{\lambda}^*_{R+\mu}) \right) - \operatorname{Re} \chi_\mu \\ &= 2\operatorname{Re} \left(\hat{H}(-\hat{\lambda}^*_\mu) - H(-\hat{\lambda}^*_\mu) \right) - \operatorname{Re} \chi_\mu \end{aligned} \tag{4.15}$$

since $\hat{\lambda}_j = \hat{\lambda}^*_{R+j}$ holds and H and $\hat{H}$ are real. Necessary optimality conditions for (4.13) imply that $\partial L(\psi, \chi, v)/\partial \tilde{v}$ vanishes. Setting equations (4.14) and (4.15) to zero and subtracting them, we obtain $\operatorname{Re} \chi_\mu = 0$ for $j = 1, \ldots, R$. Analogously, one can see that $\operatorname{Im} \chi_\mu = \operatorname{Re} \psi_\mu = \operatorname{Im} \psi_\mu = 0$ holds which means that all Lagrange multipliers vanish and we obtain the same necessary conditions as in (4.8), resp., (4.5) since H and $\hat{H}$ are real.

4.1.2. Multiple poles. Due to (2.9), for any system H it is always possible to find a representation A, B and C with the same matrix B and only the poles and the principal coefficients as free variables. Hence, the optimization variable for problem (1.7) in the case of multiple poles is given by

$$v = (\hat{\lambda}_1, \ldots, \hat{\lambda}_R, \hat{\phi}_{11}, \ldots, \hat{\phi}_{1,r_1}, \ldots, \hat{\phi}_{R1}, \ldots, \hat{\phi}_{R,r_R})^T \in \mathbb{C}^{R+r} \tag{4.16}$$

where R is the given and fixed number of poles in the reduced system, each of given and fixed order r_j, with corresponding principal coefficients, ϕ_{jl}, $j = 1, \ldots, R$, $l = 1, \ldots, r_R$, $r_1 + \ldots + r_R = r$. Assuming again that all poles of the original system are different from the poles of the reduced system, the error system $H - \hat{H}$ has $n + R$ simple poles $\lambda_1, \ldots, \lambda_n, \hat{\lambda}_1, \ldots, \hat{\lambda}_R$ with corresponding coefficients

$\phi_{11},\ldots,\phi_{N,n_N},-\hat{\phi}_{11},\ldots,-\hat{\phi}_{R,r_R}$. Due to (3.10), we have the following cost functional:

$$J(v)=\sum_{j=1}^{N}\sum_{l=1}^{n_j}\frac{(-1)^{l-1}}{(l-1)!}\phi_{jl}^*\left(H^{(l-1)}(-\lambda_j^*)-\hat{H}^{(l-1)}(-\lambda_j^*)\right)$$
$$+\sum_{j=1}^{R}\sum_{l=1}^{r_j}\frac{(-1)^{l-1}}{(l-1)!}\hat{\phi}_{jl}^*\left(\hat{H}^{(l-1)}(-\lambda_j^*)-H^{(l-1)}(-\lambda_j^*)\right).$$

As for simple poles, we will consider the following three cases.

Case 1: $\quad \underset{v}{\text{minimize}} \quad J(v) \quad \text{subject to } v\in\mathbb{R}^{2r}, \qquad (4.17)$

Case 2: $\quad \underset{v}{\text{minimize}} \quad J(v), \qquad (4.18)$

Case 3: $\quad \underset{v}{\text{minimize}} \quad J(v) \quad \text{subject to } \hat{H} \text{ is real.} \qquad (4.19)$

We begin with the general case 2 and consider again the real optimization variable $\tilde{v}\in\mathbb{R}^{2(R+r)}$ consisting of real and imaginary parts of the components of v. The following result can be proved analogously to Lemma 4.2 in the case of simple poles.

Lemma 4.6.

(i) For $\mu=1,\ldots,R$, $\nu=1,\ldots,r_\mu$ we have:

$$\frac{\partial H^{(l-1)}(-\lambda_j^*)}{\partial \mathrm{Re}\,\hat{\phi}_{\mu\nu}}=\frac{\partial H^{(l-1)}(-\lambda_j^*)}{\partial \mathrm{Im}\,\hat{\phi}_{\mu\nu}}=\frac{\partial H^{(l-1)}(-\hat{\lambda}_j^*)}{\partial \mathrm{Re}\,\hat{\phi}_{\mu\nu}}=\frac{\partial H^{(l-1)}(-\hat{\lambda}_j^*)}{\partial \mathrm{Im}\,\hat{\phi}_{\mu\nu}}=0,$$
$$\frac{\partial \hat{H}^{(l-1)}(-\lambda_j^*)}{\partial \mathrm{Re}\,\hat{\phi}_{\mu\nu}}=\frac{1}{\mathrm{i}}\frac{\partial \hat{H}^{(l-1)}(-\lambda_j)}{\partial \mathrm{Im}\,\hat{\phi}_{\mu\nu}}=\frac{(-1)^\nu(\nu+k-2)!}{(\nu-1)!(\lambda_j^*+\hat{\lambda}_\mu)^{\nu+l-1}},$$
$$\frac{\partial \hat{H}^{(l-1)}(-\hat{\lambda}_j^*)}{\partial \mathrm{Re}\,\hat{\phi}_{\mu\nu}}=\frac{1}{\mathrm{i}}\frac{\partial \hat{H}^{(l-1)}(-\hat{\lambda}_j)}{\partial \mathrm{Im}\,\hat{\phi}_{\mu\nu}}=\frac{(-1)^\nu(\nu+k-2)!}{(\nu-1)!(\hat{\lambda}_j^*+\hat{\lambda}_\mu)^{\nu+l-1}}.$$

(ii) For $\mu=1,\ldots,R$ we have:

$$\frac{\partial H^{(l-1)}(-\lambda_j^*)}{\partial \mathrm{Re}\,\hat{\lambda}_\mu}=\frac{\partial H^{(l-1)}(-\lambda_j^*)}{\partial \mathrm{Im}\,\hat{\lambda}_\mu}=0,$$
$$\frac{\partial \hat{H}^{(l-1)}(-\lambda_j^*)}{\partial \mathrm{Re}\,\hat{\lambda}_\mu}=\frac{1}{\mathrm{i}}\frac{\partial \hat{H}^{(l-1)}(-\lambda_j)}{\partial \mathrm{Im}\,\hat{\lambda}_\mu}=\sum_{k=1}^{r_\mu}\frac{(-1)^{k-1}(k+l-1)!\hat{\phi}_{\mu k}}{(k-1)!(\lambda_j^*+\hat{\lambda}_\mu)^{k+l}},$$
$$\frac{\partial H^{(l-1)}(-\hat{\lambda}_j^*)}{\partial \mathrm{Re}\,\hat{\lambda}_\mu}=-\frac{1}{\mathrm{i}}\frac{\partial H^{(l-1)}(-\hat{\lambda}_j)}{\partial \mathrm{Im}\,\hat{\lambda}_\mu}=\begin{cases}0, & j\neq\mu,\\ -H^{(k)}(-\hat{\lambda}_j^*), & j=\mu,\end{cases}$$
$$\frac{\partial \hat{H}^{(l-1)}(-\hat{\lambda}_j^*)}{\partial \mathrm{Re}\,\hat{\lambda}_\mu}=\begin{cases}\sum_{k=1}^{r_\mu}\frac{(-1)^{k-1}(k+l-1)!\hat{\phi}_{\mu k}}{(k-1)!(\hat{\lambda}_j^*+\hat{\lambda}_\mu)^{k+l}}, & j\neq\mu,\\ -\hat{H}^{(l)}(-\hat{\lambda}_j^*)+\sum_{k=1}^{r_\mu}\frac{(-1)^{k-1}(k+l-1)!\hat{\phi}_{\mu k}}{(k-1)!(\hat{\lambda}_j^*+\hat{\lambda}_\mu)^{k+l}}, & j=\mu.\end{cases}$$
$$\frac{\partial \hat{H}^{(l-1)}(-\hat{\lambda}_j^*)}{\partial \mathrm{Im}\,\hat{\lambda}_\mu}=\begin{cases}\sum_{k=1}^{r_\mu}\frac{(-1)^{k-1}(k+l-1)!\hat{\phi}_{\mu k}}{(k-1)!(\hat{\lambda}_j^*+\hat{\lambda}_\mu)^{k+l}}, & j\neq\mu,\\ \mathrm{i}\hat{H}^{(l)}(-\hat{\lambda}_j^*)-\sum_{k=1}^{r_\mu}\frac{(-1)^{k-1}(k+l-1)!\hat{\phi}_{\mu k}}{(k-1)!(\hat{\lambda}_j^*+\hat{\lambda}_\mu)^{k+l}}, & j=\mu.\end{cases}$$

Theorem 4.7. *Necessary optimality conditions for problem (4.18) are given by*

$$H^{(l)}(-\hat{\lambda}_j^*) = \hat{H}^{(l)}(-\hat{\lambda}_j^*), \qquad j = 1, \dots, R, \quad l = 0, \dots, r_j \tag{4.20}$$

i.e., interpolation of $\hat{H}$ and its first r_j derivative with H at $-\hat{\lambda}_j^*$, $j = 1, \dots, R$.

Proof: Similar to the case of simple poles and using Lemma 4.6, differentiating J with respect to $\tilde{v}$ yields

$$\frac{\partial J}{\partial \mathrm{Re}\, \hat{\phi}_{\mu\nu}} = 2\frac{(-1)^\nu}{(\nu-1)!}\mathrm{Re}\left(\hat{H}^{(\nu-1)}(-\hat{\lambda}_\mu^*) - H^{(\nu-1)}(-\hat{\lambda}_\mu^*)\right), \tag{4.21}$$

$$\frac{\partial J}{\partial \mathrm{Im}\, \hat{\phi}_{\mu\nu}} = 2\frac{(-1)^\nu}{(\nu-1)!}\mathrm{Im}\left(\hat{H}^{(\nu-1)}(-\hat{\lambda}_\mu^*) - H^{(\nu-1)}(-\hat{\lambda}_\mu^*)\right) \tag{4.22}$$

for $\mu = 1, \dots, R$ and $\nu = 1, \dots, r_\mu$ and

$$\begin{aligned}\frac{\partial J}{\partial \mathrm{Re}\, \hat{\lambda}_\mu} &= 2\sum_{l=1}^{r_\mu}\frac{(-1)^{l-1}}{(l-1)!}\mathrm{Re}\left(\hat{\phi}_{\mu l}^*\left(H^{(l)}(-\hat{\lambda}_\mu^*) - \hat{H}^{(l)}(-\hat{\lambda}_\mu^*)\right)\right),\\ \frac{\partial J}{\partial \mathrm{Im}\, \hat{\lambda}_\mu} &= 2\sum_{l=1}^{r_\mu}\frac{(-1)^{l-1}}{(l-1)!}\mathrm{Im}\left(\hat{\phi}_{\mu l}^*\left(H^{(l)}(-\hat{\lambda}_\mu^*) - \hat{H}^{(l)}(-\hat{\lambda}_\mu^*)\right)\right)\end{aligned} \tag{4.23}$$

for $\mu = 1, \dots, R$. Equations (4.21) and (4.21) gives interpolation of the first $r_j - 1$ derivatives which, together with (4.23), implies interpolation of r_j-th derivatives in each $-\hat{\lambda}_j^*$, $j = 1, \dots, R$.

For case 1, one can use the previous proofs. We obtain all derivative formulas w.r.t. v by omitting the "Re" notation in all preceding formulas for derivatives w.r.t. the real parts of the components of $\tilde{v}$. This directly yields the optimality conditions (4.20). Using Proposition 2.2, we obtain the following result.

Theorem 4.8. *Necessary optimality conditions for problem (4.17) are given by*

$$H^{(l)}(-\hat{\lambda}_j) = \hat{H}^{(l)}(-\hat{\lambda}_j), \qquad j = 1, \dots, R, \quad l = 0, \dots, r_j$$

i.e., interpolation of $\hat{H}$ and its first r_j derivative with H at $-\hat{\lambda}_j$, $j = 1, \dots, R$.

Case 3 can be handled as for simple poles. Again, we obtain that the Lagrangian multipliers of problem (4.19) vanish and we obtain the following result.

Theorem 4.9. *Necessary optimality conditions for problem (4.19) are given by*

$$H^{(l)}(-\hat{\lambda}_j) = \hat{H}^{(l)}(-\hat{\lambda}_j), \qquad j = 1, \dots, R, \quad l = 0, \dots, r_j$$

i.e., interpolation of $\hat{H}$ and its first r_j derivative with H at $-\hat{\lambda}_j$, $j = 1, \dots, R$.

4.2. MIMO. For MIMO systems, we will only consider the general case (i.e., case 2 for SISO) of optimization problem (1.7) without constraints. Contrary to the SISO case, the residues of a transfer function now depend on both B and C if A is given in eigenvalue decomposition form. Hence, both $\hat{B}$ and $\hat{C}$ act as optimization variables additionally to the eigenvalues of $\hat{A}$.

4.2.1. Simple poles. The optimization vector is now given by

$$v = (\hat{\lambda}_1, \dots, \hat{\lambda}_r, \hat{B}_1^1, \dots, \hat{B}_r^m, \hat{C}_1^1, \dots, \hat{C}_r^p)^T \in \mathbb{C}^{r+mr+pr}, \tag{4.24}$$

resp., $\tilde{v} = (\mathrm{Re}\, v^T, \mathrm{Im}\, v^T)^T \in \mathbb{R}^{2(r+mr+pr)}$. Similarly to the SISO case we consider the error system $H - \hat{H}$ which has the $n+r$ simple poles $\lambda_1, \dots, \lambda_n, \hat{\lambda}_1, \dots, \hat{\lambda}_r$ with corresponding residues $\phi_1^{ki}, \dots, \phi_n^{ki}, -\hat{\phi}_1^{ki}, \dots, -\hat{\phi}_r^{ki}$ in the ki-th component of the error system. In view of (3.13) and (2.12), the optimization problem can be formulated as

$$\begin{aligned} \underset{v}{\text{minimize}} \quad J(\tilde{v}) = & \sum_{j=1}^{n} \sum_{i,k} (\phi_j^{ki})^* \left(H_{ki}(-\lambda_j^*) - \hat{H}_{ki}(-\lambda_j^*) \right) \\ & + \sum_{j=1}^{r} \sum_{i,k} (\hat{\phi}_j^{ki})^* \left(\hat{H}_{ki}(-\hat{\lambda}_j^*) - H_{ki}(-\hat{\lambda}_j^*) \right) \end{aligned} \tag{4.25}$$

which is smooth with respect to $\tilde{v}$ and unconstrained.

THEOREM 4.10. *Necessary optimality conditions for problem (4.25) are given by*

$$\left.\begin{aligned} H(-\hat{\lambda}_j^*)\hat{B}_j^* &= \hat{H}(-\hat{\lambda}_j^*)\hat{B}_j^*, \\ \hat{C}_j^* H(-\hat{\lambda}_j^*) &= \hat{C}_j^* \hat{H}(-\hat{\lambda}_j^*), \\ \hat{C}_j^* H'(-\hat{\lambda}_j^*)\hat{B}_j^* &= \hat{C}_j^* \hat{H}'(-\hat{\lambda}_j^*)\hat{B}_j^*, \end{aligned}\right\} \quad j = 1, \dots, r, \tag{4.26}$$

where

$$\hat{C}_j := j\text{-th column of } \hat{C}, \quad \hat{B}_j := j\text{-th row of } \hat{B}.$$

In other words, the optimal reduced transfer function $\hat{H}$ is a left sided tangential interpoland into the directions C_j^*, a right sided tangential interpoland into the directions B_j^* and a both sided tangential interpoland into the directions C_j^*, resp., B_j^* of the original transfer function H.

Proof: Since J can be written as $J = \sum_{i,k} J_{ki}$ with

$$J_{ki} = \sum_{j=1}^{n} (\phi_j^{ki})^* \left(H_{ki}(-\lambda_j^*) - \hat{H}_{ki}(-\lambda_j^*) \right) + \sum_{j=1}^{r} (\hat{\phi}_j^{ki})^* \left(\hat{H}_{ki}(-\hat{\lambda}_j^*) - H_{ki}(-\hat{\lambda}_j^*) \right)$$

we can use the derivative formulas from the SISO case for each J_{ik} to calculate the derivatives of J. Define the vectors

$$\hat{\phi}^{ki} := (\hat{\phi}_1^{ki}, \dots, \hat{\phi}_r^{ki})^T \in \mathbb{C}^r, \quad \hat{\phi} := (\hat{\phi}^{11}, \dots, \hat{\phi}^{pm}) \in \mathbb{C}^{rpm}.$$

Then, we have

$$\frac{\partial J}{\mathrm{Re}\, \hat{\phi}} = \left(\frac{\partial J_{11}}{\partial \mathrm{Re}\, \hat{\phi}^{11}}, \dots, \frac{\partial J_{pm}}{\partial \mathrm{Re}\, \hat{\phi}^{pm}} \right), \quad \frac{\partial J}{\mathrm{Im}\, \hat{\phi}} = \left(\frac{\partial J_{11}}{\partial \mathrm{Im}\, \hat{\phi}^{11}}, \dots, \frac{\partial J_{pm}}{\partial \mathrm{Im}\, \hat{\phi}^{pm}} \right)$$

with entries as in the SISO case given in (4.9) and (4.10). Due to (2.12), the residues satisfy $\hat{\phi}_j^{ki} = \hat{C}_j^k \hat{B}_j^i$, resp.,

$$\mathrm{Re}\, \hat{\phi}_j^{ki} = \mathrm{Re}\, \hat{C}_j^k \mathrm{Re}\, \hat{B}_j^i - \mathrm{Im}\, \hat{C}_j^k \mathrm{Im}\, \hat{B}_j^i, \quad \mathrm{Im}\, \hat{\phi}_j^{ki} = \mathrm{Re}\, \hat{C}_j^k \mathrm{Im}\, \hat{B}_j^i + \mathrm{Im}\, \hat{C}_j^k \mathrm{Re}\, \hat{B}_j^i.$$

Hence, for $i = 1, \dots, m$ we obtain

$$\frac{\partial \mathrm{Re}\, \hat{\phi}_j^{ki}}{\partial \mathrm{Re}\, \hat{C}_\mu^\kappa} = \mathrm{Re}\, \hat{B}_\mu^i, \quad \frac{\partial \mathrm{Im}\, \hat{\phi}_j^{ki}}{\partial \mathrm{Re}\, \hat{C}_\mu^\kappa} = \mathrm{Im}\, \hat{B}_\mu^i, \quad \frac{\partial \mathrm{Re}\, \hat{\phi}_j^{ki}}{\partial \mathrm{Im}\, \hat{C}_\mu^\kappa} = -\mathrm{Im}\, \hat{B}_\mu^i, \quad \frac{\partial \mathrm{Im}\, \hat{\phi}_j^{ki}}{\partial \mathrm{Im}\, \hat{C}_\mu^\kappa} = \mathrm{Re}\, \hat{B}_\mu^i$$

if $\kappa = k$ and $\mu = j$. For $\kappa \neq k$ or $\mu \neq j$, the derivatives are zero. Then, the chain rule yields

$$\begin{aligned}
\frac{\partial J}{\partial \mathrm{Re}\,\hat{C}^\kappa_\mu} &= \frac{\partial J}{\partial \mathrm{Re}\,\hat{\phi}} \frac{\partial \mathrm{Re}\,\phi}{\partial \mathrm{Re}\,\hat{C}^\kappa_\mu} + \frac{\partial J}{\partial \mathrm{Im}\,\hat{\phi}} \frac{\partial \mathrm{Im}\,\phi}{\partial \mathrm{Re}\,\hat{C}^\kappa_\mu} \\
&= \sum_{i,k} \frac{\partial J_{ki}}{\partial \mathrm{Re}\,\hat{\phi}^{ki}} \frac{\partial \mathrm{Re}\,\phi^{ki}}{\partial \mathrm{Re}\,\hat{C}^\kappa_\mu} + \sum_{i,k} \frac{\partial J_{ki}}{\partial \mathrm{Im}\,\hat{\phi}^{ki}} \frac{\partial \mathrm{Im}\,\phi^{ki}}{\partial \mathrm{Re}\,\hat{C}^\kappa_\mu} \\
&= 2\sum_{i=1}^m \mathrm{Re}\,\Big(\hat{H}_{\kappa i}(-\hat{\lambda}^*_\mu) - H_{\kappa i}(-\hat{\lambda}^*_\mu)\Big)\,\mathrm{Re}\,\hat{B}^i_\mu \\
&\quad + 2\sum_{i=1}^m \mathrm{Im}\,\Big(\hat{H}_{\kappa i}(-\hat{\lambda}^*_\mu) - H_{\kappa i}(-\hat{\lambda}^*_\mu)\Big)\,\mathrm{Im}\,\hat{B}^i_\mu \\
&= 2\sum_{i=1}^m \mathrm{Re}\,\Big[\Big(\hat{H}_{\kappa i}(-\hat{\lambda}^*_\mu) - H_{\kappa i}(-\hat{\lambda}^*_\mu)\Big)\big(\hat{B}^i_\mu\big)^*\Big].
\end{aligned}$$

Analogously, we obtain

$$\begin{aligned}
\frac{\partial J}{\partial \mathrm{Im}\,\hat{C}^\kappa_\mu} &= 2\sum_{i=1}^m \mathrm{Im}\,\Big[\Big(\hat{H}_{\kappa i}(-\hat{\lambda}^*_\mu) - H_{\kappa i}(-\hat{\lambda}^*_\mu)\Big)\big(\hat{B}^i_\mu\big)^*\Big], \\
\frac{\partial J}{\partial \mathrm{Re}\,\hat{B}^\iota_\mu} &= 2\sum_{k=1}^p \mathrm{Re}\,\Big[\Big(\hat{H}_{k\iota}(-\hat{\lambda}^*_\mu) - H_{k\iota}(-\hat{\lambda}^*_\mu)\Big)\big(\hat{C}^k_\mu\big)^*\Big], \\
\frac{\partial J}{\partial \mathrm{Im}\,\hat{B}^\iota_\mu} &= 2\sum_{k=1}^p \mathrm{Im}\,\Big[\Big(\hat{H}_{k\iota}(-\hat{\lambda}^*_\mu) - H_{k\iota}(-\hat{\lambda}^*_\mu)\Big)\big(\hat{C}^k_\mu\big)^*\Big].
\end{aligned}$$

Hence, the necessary optimality condition $\partial J/\partial \tilde{z} = 0$ imply

$$\begin{aligned}
\sum_{i=1}^m \hat{H}_{\kappa i}(-\hat{\lambda}^*_\mu)\big(\hat{B}^i_\mu\big)^* &= \sum_{i=1}^m H_{\kappa i}(-\hat{\lambda}^*_\mu)\big(\hat{B}^i_\mu\big)^*, \quad \kappa = 1,\ldots,p, \\
\sum_{k=1}^p \big(\hat{C}^k_\mu\big)^* \hat{H}_{k\iota}(-\hat{\lambda}^*_\mu) &= \sum_{k=1}^p \big(\hat{C}^k_\mu\big)^* H_{k\iota}(-\hat{\lambda}^*_\mu), \quad \iota = 1,\ldots,m,
\end{aligned}$$

for $\mu = 1,\ldots,r$ which is equivalent to the first two conditions in (4.26). Differentiation of J with respect to real, resp., imaginary part of $\hat{\lambda}_\mu$ via formulas (4.11) for J_{ik} directly yields

$$\begin{aligned}
\frac{\partial J}{\partial \mathrm{Re}\,\hat{\lambda}_\mu} &= 2\sum_{i,k} \mathrm{Re}\,\Big(\big(\hat{\phi}^{ik}_\mu\big)^*\Big(H'(-\hat{\lambda}^*_\mu) - \hat{H}'(-\hat{\lambda}^*_\mu)\Big)\Big), \\
\frac{\partial J}{\partial \mathrm{Im}\,\hat{\lambda}_\mu} &= 2\sum_{i,k} \mathrm{Im}\,\Big(\big(\hat{\phi}^{ik}_\mu\big)^*\Big(H'(-\hat{\lambda}^*_\mu) - \hat{H}'(-\hat{\lambda}^*_\mu)\Big)\Big)
\end{aligned}$$

for $\mu = 1,\ldots,r$. Due to necessary conditions and $\phi^{ki}_j = C^k_j B^i_j$, we obtain

$$\sum_{i,k} \big(\hat{C}^k_\mu\big)^* \hat{H}'_{ki}(-\hat{\lambda}^*_\mu)\big(\hat{B}^i_\mu\big)^* = \sum_{i,k} \big(\hat{C}^k_\mu\big)^* H'_{ki}(-\hat{\lambda}^*_\mu)\big(\hat{B}^i_\mu\big)^*, \quad \mu = 1,\ldots,r.$$

This is equivalent to the third condition in (4.26) and hence, completes the proof.

4.2.2. Multiple poles. Now, the optimization problem is given by

$$\begin{aligned}\underset{v}{\text{minimize}} \quad J(\tilde v) = &\sum_{j=1}^{N}\sum_{l=1}^{n_j}\sum_{i,k}\frac{(-1)^{l-1}}{(l-1)!}(\phi_{jl}^{ki})^*\left(H_{ki}^{(l-1)}(-\lambda_j^*) - \hat H_{ki}^{(l-1)}(-\lambda_j^*)\right)\\ &+\sum_{j=1}^{R}\sum_{l=1}^{r_j}\sum_{i,k}\frac{(-1)^{l-1}}{(l-1)!}(\hat\phi_{jl}^{ki})^*\left(\hat H_{ki}^{(l-1)}(-\hat\lambda_j^*) - H_{ki}^{(l-1)}(-\hat\lambda_j^*)\right)\end{aligned} \tag{4.27}$$

with the optimization vector

$$v = (\hat\lambda_1,\dots,\hat\lambda_R,(\hat{\hat B}_1)_1^1,\dots,(\hat{\hat B}_N)_{n_N}^m,(\hat{\hat C}_1)_1^1,\dots,(\hat{\hat C}_N)_{n_N}^p)^T \in \mathbb{C}^{R+mr+pr}, \tag{4.28}$$

resp., $\tilde v = (\mathrm{Re}\, v^T, \mathrm{Im}\, v^T)^T \in \mathbb{R}^{2(R+mr+pr)}$ and the principal coefficients ϕ_{jl}^{ki} and $\hat\phi_{jl}^{ki}$ as in (2.15).

Theorem 4.11. *Necessary optimality conditions for problem (4.27) are given by*

$$\begin{aligned}\sum_{q=0}^{r_j-l}\frac{(-1)^q}{q!}H^{(q)}(-\hat\lambda_j^*)\big((\hat{\hat B}_j)_{l+q}\big)^* &= \sum_{q=0}^{r_j-l}\frac{(-1)^q}{q!}\hat H^{(q)}(-\hat\lambda_j^*)\big((\hat{\hat B}_j)_{l+q}\big)^*,\\ \sum_{q=0}^{l-1}\frac{(-1)^q}{q!}\big((\hat{\hat C}_j)_{l-q}\big)^*H^{(q)}(-\hat\lambda_j^*) &= \sum_{q=0}^{l}\frac{(-1)^q}{q!}\big((\hat{\hat C}_j)_{l-q}\big)^*\hat H^{(q)}(-\hat\lambda_j^*),\end{aligned} \tag{4.29}$$

for $j = 1,\dots,r$, $l = 1,\dots,r_j$ and

$$\begin{aligned}&\sum_{l=1}^{r_j}\frac{(-1)^{l-1}}{(l-1)!}\sum_{q=0}^{l-1}\big((\hat{\hat C}_j)_q\big)^*H^{(l)}(-\hat\lambda_j^*)\big((\hat{\hat B}_j)_{q+l-1}\big)^*\\ =&\sum_{l=1}^{r_j}\frac{(-1)^{l-1}}{(l-1)!}\sum_{q=0}^{l-1}\big((\hat{\hat C}_j)_q\big)^*\hat H^{(l)}(-\hat\lambda_j^*)\big((\hat{\hat B}_j)_{q+l-1}\big)^*\end{aligned} \tag{4.30}$$

for $j = 1,\dots,r$ where

$$(\hat{\hat C}_j)_q := q\text{-th column of the matrix } \hat{\hat C}_j, \quad (\hat{\hat B}_j)_q := q\text{-th row of the matrix } \hat{\hat B}_j.$$

Proof: Again, we write $J = \sum_{i,k} J_{ik}$ and use the formulas from the SISO case. Due to the principal coefficients representation (2.15), for $i = 1,\dots,m$ we have

$$\frac{\partial \mathrm{Re}\,\phi_{jl}^{ki}}{\partial \mathrm{Re}\,(\hat{\hat C}_\mu)_\nu^\kappa} = \mathrm{Re}\,(\hat{\hat B}_\mu)_{\nu+l-1}^i$$

if $\kappa = k$, $\mu = j$, $1 \le \nu \le r_j - l + 1$, resp., for $k = 1,\dots,p$ we obtain

$$\frac{\partial \mathrm{Re}\,\phi_{jl}^{ki}}{\partial \mathrm{Re}\,(\hat{\hat B}_\mu)_\nu^\iota} = \mathrm{Re}\,(\hat{\hat C}_\mu)_{\nu-l+1}^k$$

if $\iota = i$, $\mu = j$, $1 \le \nu - l + 1 \le r_j - l + 1$, whereas derivatives for other indices vanish. Similar formulas hold for derivatives of imaginary parts of $\hat{\phi}$ and also for derivatives with respect to imaginary parts of $\hat{B}$ and $\hat{C}$. Therefore, the chain rule yields

$$\frac{\partial J}{\partial \mathrm{Re}\,(\hat{\hat{C}}_\mu)^\kappa_\nu} = 2\sum_{i=1}^{m}\sum_{l=1}^{r_\mu-\nu+1}\frac{(-1)^{l-1}}{(l-1)!}\mathrm{Re}\left[\left(\hat{H}^{(l-1)}_{\kappa i}(-\hat{\lambda}^*_\mu) - H^{(l-1)}_{\kappa i}(-\hat{\lambda}^*_\mu)\right)\left((\hat{\hat{B}}_\mu)^i_{\nu+l-1}\right)^*\right],$$

$$\frac{\partial J}{\partial \mathrm{Im}\,(\hat{\hat{C}}_\mu)^\kappa_\nu} = 2\sum_{i=1}^{m}\sum_{l=1}^{r_\mu-\nu+1}\frac{(-1)^{l-1}}{(l-1)!}\mathrm{Im}\left[\left(\hat{H}^{(l-1)}_{\kappa i}(-\hat{\lambda}^*_\mu) - H^{(l-1)}_{\kappa i}(-\hat{\lambda}^*_\mu)\right)\left((\hat{\hat{B}}_\mu)^i_{\nu+l-1}\right)^*\right],$$

$$\frac{\partial J}{\partial \mathrm{Re}\,(\hat{\hat{B}}_\mu)^\iota_\nu} = 2\sum_{k=1}^{p}\sum_{l=1}^{\nu}\frac{(-1)^{l-1}}{(l-1)!}\mathrm{Re}\left[\left(\hat{H}^{(l-1)}_{k\iota}(-\hat{\lambda}^*_\mu) - H^{(l-1)}_{k\iota}(-\hat{\lambda}^*_\mu)\right)\left((\hat{\hat{C}}_\mu)^k_{\nu-l+1}\right)^*\right],$$

$$\frac{\partial J}{\partial \mathrm{Im}\,(\hat{\hat{B}}_\mu)^\iota_\nu} = 2\sum_{k=1}^{p}\sum_{l=1}^{\nu}\frac{(-1)^{l-1}}{(l-1)!}\mathrm{Im}\left[\left(\hat{H}^{(l-1)}_{k\iota}(-\hat{\lambda}^*_\mu) - H^{(l-1)}_{k\iota}(-\hat{\lambda}^*_\mu)\right)\left((\hat{\hat{C}}_\mu)^k_{\nu-l+1}\right)^*\right].$$

As all derivatives have to vanish, we obtain the two conditions in (4.29). Furthermore, from (4.23) we have

$$\frac{\partial J}{\partial \mathrm{Re}\,\hat{\lambda}_\mu} = 2\sum_{i,k}\sum_{l=1}^{r_\mu}\frac{(-1)^{l-1}}{(l-1)!}\mathrm{Re}\left((\hat{\phi}^{ki}_{\mu l})^*\left(H^{(l)}_{ki}(-\hat{\lambda}^*_\mu) - \hat{H}^{(l)}_{ki}(-\hat{\lambda}^*_\mu)\right)\right),$$

$$\frac{\partial J}{\partial \mathrm{Im}\,\hat{\lambda}_\mu} = 2\sum_{i,k}\sum_{l=1}^{r_\mu}\frac{(-1)^{l-1}}{(l-1)!}\mathrm{Im}\left((\hat{\phi}^{ki}_{\mu l})^*\left(H^{(l)}_{ki}(-\hat{\lambda}^*_\mu) - \hat{H}^{(l)}_{ki}(-\hat{\lambda}^*_\mu)\right)\right)$$

which, together with the representation (2.15) for the principal coefficients, yields condition (4.30).

REMARK 4.12. *Note that in case of multiple poles we do not have tangential interpolation as necessary conditions but the sum of certain directions multiplied with certain derivatives of H and $\hat{H}$ must coincide. This furthermore implies that contrary to the SISO case, we cannot use the first two conditions (4.29) to obtain a simple third condition which involves only r_j-th derivatives of the transfer functions.*

REFERENCES

[1] A.C. Antoulas, *Approximation of large-scale dynamical systems*, Advances in Design and Control **6**, SIAM, Philadelphia, 2005.

[2] J.B. Conway, *Functions of one complex variable*, Springer, 1978.

[3] S. Gugercin, C. Beattie and A.C. Antoulas, *Rational Krylov Methods for Optimal $\mathcal{H}_2$ Model Reduction*, ICAM Technical Report, Virginia Tech, 2006 and submitted to SIAM Journal on Matrix Analysis and Applications.

[4] L. Meier and D.G. Luenberger, *Approximation of Linear Constant Systems*, IEEE Trans. Automatic Control. **12** (1967) 585-588.