Optimal H2 model approximation based on multiple input/output delays systems

I. Pontes Duff, Charles Poussot-Vassal, Cédric Seren

Optimal $\mathcal{H}_2$ model approximation based on multiple input/output delays systems

I. Pontes Duff[a], C. Poussot-Vassal[b], C. Seren[b]

[a] ISAE SUPAERO & ONERA
[b] ONERA - The French Aerospace Lab, F-31055 Toulouse, France

Abstract

In this paper, the $\mathcal{H}_2$ optimal approximation of a $n_y \times n_u$ transfer function $\mathbf{G}(s)$ by a finite dimensional system $\hat{\mathbf{H}}_d(s)$ including input/output delays, is addressed. The underlying $\mathcal{H}_2$ optimality conditions of the approximation problem are firstly derived and established in the case of a poles/residues decomposition. These latter form an extension of the tangential interpolatory conditions, presented in [1, 2] for the delay-free case, which is the main contribution of this paper. Secondly, a two stage algorithm is proposed in order to practically obtain such an approximation.

Keywords: Model reduction, time-delay systems, large-scale systems, linear systems.

1. Introduction

Model approximation plays a pivotal role in many simulation based optimization, control, analysis procedures. Indeed, due to memory and computational burden limitations working with a reduced order model in place of the original one, potentially large-scale, might be a real advantage. To this aim, most of the results presented in the literature address the linear dynamical systems approximation problem in the delay-free case[1]. More specifically, this problem has been widely studied using either Lyapunov-based methods [3, 4, 5], interpolation-based algorithm [6, 1, 2, 7], or matching moments approaches [8, 9], leading to a variety of solutions and applications. Recent surveys are available in [10, 11, 12]. The presence of input/output delays in the approximation model was tackled in [13] (exploiting both Lyapunov equations and grammians properties derived in [4] for the free-delay case). The bottleneck of this approach is that it requires to solve Lyapunov equations which might be costly in the large-scale context. From the moment matching side, [14] proposed a problem formulation that enables the construction of an approximation which contains very rich delay structure (including state delay), but where the delays and the interpolation points are supposed to be a priori known. From the Loewner framework side, [15] and after [16] generalizes the Loewner framework from [17] to the state delay case enabling data-driven interpolation. However, as for the moment matching case, the delays and the interpolation points are supposed to be a priori known.

In this paper, the problem of approximating a given large-scale model by a low order one including (a priori unknown) I/O delays using the interpolatory framework, is addressed. An alternative "poles/residues"-based approach is developed, which enables to reach the $\mathcal{H}_2$ optimality

[1] "Delay-free case" means that the approximation model is a dynamical model without any input/output/state delays.

conditions, treated as interpolation ones. Then, the main contribution of this paper consists in extending the interpolation results of [1] to the case of approximate models with an extended structure, namely, including non-zero input(s)/output(s) delays. Last but not least, $\mathcal{H}_2$ optimality conditions for such cases are also elegantly derived together with a single numerical procedure.

The paper is organized as follows: after introducing the notations and the mathematical problem statement in Section 2, Section 3 recalls some necessary preliminary results related to the computational aspects of the $\mathcal{H}_2$ inner product and $\mathcal{H}_2$ norm when the calculations are based on the poles/residues decomposition of a transfer function. Section 4 establishes the $\mathcal{H}_2$ optimality conditions solving the input/output delay dynamical model approximation problem. It also proposes an algorithm which permits to practically compute such an approximation. Section 5 details the results obtained after treating an academic example. Conclusions and prospects end this article in Section 6.

2. Notations and problem statement

Notations. Let us consider a stable **M**ultiple-**I**nput/**M**ultiple-**O**utput (**MIMO**) linear dynamical system, denoted by $\mathbf{G}$ in the sequel, with n_u (*resp.* n_y) $\in \mathbb{N}^*$ input(s) (*resp.* output(s)), represented by its transfer function $\mathbf{G}(s) \in \mathbb{C}^{n_y \times n_u}$. Let $\mathcal{H}_2^{n_y \times n_u}$ be the Hilbert space of holomorphic functions $\mathbf{F} : \mathbb{C} \to \mathbb{C}^{n_y \times n_u}$ which are analytic in the open right-half plane and for which $\int_{-\infty}^{+\infty} \mathsf{trace}\left(\overline{\mathbf{F}(i\omega)}\mathbf{F}^T(i\omega)\right) d\omega < +\infty$. For given $\mathbf{G}, \mathbf{H} \in \mathcal{H}_2^{n_y \times n_u}$, the associated inner-product reads:

$$\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathsf{trace}\left(\overline{\mathbf{G}(i\omega)}\mathbf{H}^T(i\omega)\right) d\omega, \tag{1}$$

and the $\mathcal{H}_2^{n_y \times n_u}$ induced norm can be explained:

$$\|\mathbf{G}\|_{\mathcal{H}_2} = \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \|\mathbf{G}(i\omega)\|_{\mathsf{F}}^2 d\omega\right)^{1/2} = \langle \mathbf{G}, \mathbf{G} \rangle_{\mathcal{H}_2}, \tag{2}$$

where $\|\mathbf{G}\|_{\mathsf{F}}^2 = \langle \mathbf{G}, \mathbf{G} \rangle_{\mathsf{F}}$ and $\langle \mathbf{G}, \mathbf{H} \rangle_{\mathsf{F}} = \mathsf{trace}(\overline{\mathbf{G}}\mathbf{H}^T)$ are the *Frobenius norm* and *inner-product*, respectively. Dynamical system $\mathbf{H}$ will be said *real iff.* $\forall s \in \mathbb{C}$, $\overline{\mathbf{H}(s)} = \mathbf{H}(\overline{s})$. It is noteworthy that if $\mathbf{G}(s), \mathbf{H}(s) \in \mathcal{H}_2^{n_y \times n_u}$ are real, then $\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} = \langle \mathbf{H}, \mathbf{G} \rangle_{\mathcal{H}_2} \in \mathbb{R}_+$.

Besides, any dynamical matrix $\mathbf{\Delta}(s)$ will belong to $\mathcal{H}_\infty^{n_y \times n_u}$ *iff.* $\sup\{\sigma_{max}(\mathbf{\Delta}(i\omega))/\omega \in \mathbb{R}\} < +\infty$. $\sigma_{max}(\mathbf{\Delta}(i\omega))$ refers to the largest singular value of matrix $\mathbf{\Delta}(i\omega)$.

Followingly, let $\hat{\mathbf{H}}_d$ be a multiple-input/output delays **MIMO** system *s.t.* $\hat{\mathbf{H}}_d(s) \in \mathcal{H}_2^{n_y \times n_u}$ and represented by:

$$\hat{\mathbf{H}}_d : \left\{ \begin{array}{rcl} \hat{\mathbf{E}}\dot{\hat{\mathbf{x}}}(t) &=& \hat{\mathbf{A}}\hat{\mathbf{x}}(t) + \hat{\mathbf{B}}\mathbf{\Delta}_i(\mathbf{u}(t)) \\ \hat{\mathbf{y}}(t) &=& \mathbf{\Delta}_o(\hat{\mathbf{C}}\hat{\mathbf{x}}(t)) \end{array} \right. , \tag{3}$$

where $\hat{\mathbf{E}}, \hat{\mathbf{A}} \in \mathbb{R}^{n \times n}$ (with state dimension $n \in \mathbb{N}^*$), $\hat{\mathbf{B}} \in \mathbb{R}^{n \times n_u}$, $\hat{\mathbf{C}} \in \mathbb{R}^{n_y \times n}$ and $\mathbf{\Delta}_i$ and $\mathbf{\Delta}_o$ are delay operators. The matrix transfer functions $\hat{\mathbf{\Delta}}_i(s)$ and $\hat{\mathbf{\Delta}}_o(s)$ defined in (5) represent the frequency behavior of the delays operators $\mathbf{\Delta}_i$ and $\mathbf{\Delta}_o$, receptively. The transfer function of the underlying system (3) from input $\hat{\mathbf{u}}(t)$ to output $\hat{\mathbf{y}}(t)$ vectors is given by:

$$\hat{\mathbf{H}}_d(s) = \hat{\mathbf{\Delta}}_o(s)\hat{\mathbf{H}}(s)\hat{\mathbf{\Delta}}_i(s) \in \mathcal{H}_2^{n_y \times n_u}, \tag{4}$$

where:

$$\left\{\begin{array}{lcl}\hat{\mathbf{H}}(s) & = & \hat{\mathbf{C}}(\hat{\mathbf{E}}s-\hat{\mathbf{A}})^{-1}\hat{\mathbf{B}} \in \mathcal{H}_2^{n_y\times n_u} \\ \hat{\mathbf{\Delta}}_i(s) & = & \mathsf{diag}(e^{-s\hat{\tau}_1}\ \ldots\ e^{-s\hat{\tau}_{n_u}}) \in \mathcal{H}_\infty^{n_u\times n_u} \\ \hat{\mathbf{\Delta}}_o(s) & = & \mathsf{diag}(e^{-s\hat{\gamma}_1}\ \ldots\ e^{-s\hat{\gamma}_{n_y}}) \in \mathcal{H}_\infty^{n_y\times n_y}.\end{array}\right. \tag{5}$$

From this point, we will denote by $\hat{\mathbf{H}}_d = (\hat{\mathbf{E}}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{\Delta}}_i, \hat{\mathbf{\Delta}}_o)$ a **MIMO** input/output delayed system of the form (4). $\hat{\mathbf{H}}_d$ will also be said to have order $n \ll N$ (where N is the original model order).

Problem statement. The main objective addressed in this paper is to solve the following $\mathcal{H}_2$ approximation problem:

Problem 2.1. (Delay model $\mathcal{H}_2$-optimal approximation) *Given a stable N^{th} order system $\mathbf{G} \in \mathcal{H}_2^{n_y\times n_u}$, find a reduced n^{th} order (*s.t. *$n \ll N$) multiple-input/output delays model $\hat{\mathbf{H}}_d^\star = (\hat{\mathbf{E}}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{\Delta}}_i, \hat{\mathbf{\Delta}}_o)$ s.t.:*

$$\hat{\mathbf{H}}_d^\star = \underset{\substack{\hat{\mathbf{H}}_d \in \mathcal{H}_2^{n_y\times n_u} \\ \dim(\hat{\mathbf{H}}_d)\leq n}}{\operatorname{argmin}} ||\mathbf{G}-\hat{\mathbf{H}}_d||_{\mathcal{H}_2},$$

where $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o\hat{\mathbf{H}}\hat{\mathbf{\Delta}}_i$ as in, (4).

This search for an optimal solution will be carried out assuming that both $\mathbf{G}$ and $\hat{\mathbf{H}}$ from Eq. (5) have semi-simple poles *i.e., s.t.* their respective transfer function matrix can be decomposed as follows:

$$\mathbf{G}(s) = \sum_{j=1}^{N} \frac{\mathbf{l}_j\mathbf{r}_j^T}{s-\mu_j} \text{ and } \hat{\mathbf{H}}(s) = \sum_{k=1}^{n} \frac{\hat{\mathbf{c}}_k\hat{\mathbf{b}}_k^T}{s-\hat{\lambda}_k}, \tag{6}$$

where $\forall j = 1\ldots N,\ \forall k = 1\ldots n,\ \mathbf{r}_j, \hat{\mathbf{b}}_k \in \mathbb{C}^{n_u\times 1}$ and $\mathbf{l}_j, \hat{\mathbf{c}}_k \in \mathbb{C}^{n_y\times 1}$. The poles $\mu_j, \hat{\lambda}_k$ are elements of $\mathbb{C}_-$ so that $\mathbf{G}$ and $\hat{\mathbf{H}}$ belong to $\mathcal{H}_2^{n_y\times n_u}$.

3. Preliminary results

In this section, some elementary but important, results, which will be useful along this paper, are recalled and generalized.

First of all, a fundamental result dealing with the $\mathcal{H}_2$ norm invariance in case of input/output delayed systems is presented.

Proposition 3.1. **($\mathcal{H}_2$ norm invariance)** *Let $\hat{\mathbf{H}} \in \mathcal{H}_2^{n_y\times n_u}$ be a stable dynamical system and $\mathbf{M} \in \mathcal{H}_\infty^{n_u\times n_u}$, $\mathbf{N} \in \mathcal{H}_\infty^{n_y\times n_y}$ s.t.:*

$$\forall\omega \in \mathbb{R},\ \overline{\mathbf{M}(i\omega)}\mathbf{M}(i\omega)^T = \mathbf{I}_{n_u},\ \mathbf{N}(i\omega)^T\overline{\mathbf{N}(i\omega)} = \mathbf{I}_{n_y}. \tag{7}$$

If $\hat{\mathbf{H}}_d = \mathbf{N}\hat{\mathbf{H}}\mathbf{M}$ then $||\hat{\mathbf{H}}_d||_{\mathcal{H}_2} = ||\hat{\mathbf{H}}||_{\mathcal{H}_2}$.

Proof. If $\hat{\mathbf{H}}_d = \mathbf{N}\hat{\mathbf{H}}\mathbf{M}$, the scaled term $2\pi\|\hat{\mathbf{H}}_d\|^2_{\mathcal{H}_2}$ will then read by definition:

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \text{trace}\left(\overline{\mathbf{N}(i\omega)}\overline{\hat{\mathbf{H}}(i\omega)}\overline{\mathbf{M}(i\omega)}\mathbf{M}^T(i\omega)\hat{\mathbf{H}}^T(i\omega)\mathbf{N}^T(i\omega)\right)d\omega \\
&= \int_{-\infty}^{+\infty} \text{trace}\left(\overline{\mathbf{N}(i\omega)}\overline{\hat{\mathbf{H}}(i\omega)}\hat{\mathbf{H}}^T(i\omega)\mathbf{N}^T(i\omega)\right)d\omega \\
&= \int_{-\infty}^{+\infty} \text{trace}\left(\overline{\hat{\mathbf{H}}(i\omega)}\hat{\mathbf{H}}^T(i\omega)\mathbf{N}^T(i\omega)\overline{\mathbf{N}(i\omega)}\right)d\omega \\
&= \int_{-\infty}^{+\infty} \text{trace}\left(\overline{\hat{\mathbf{H}}(i\omega)}\,\hat{\mathbf{H}}(i\omega)^T\right)d\omega = 2\pi\|\hat{\mathbf{H}}\|^2_{\mathcal{H}_2}. \qquad \square
\end{aligned}$$

One can easily check that condition (7) appearing in Proposition 3.1 is satisfied by the delays matrices of the two last lines of (5) when $\mathbf{M} = \hat{\mathbf{\Delta}}_i$ and $\mathbf{N} = \hat{\mathbf{\Delta}}_o$. In other words, the $\mathcal{H}_2$ norm does not depend on the input, nor output delays. The following proposition makes now explicit the calculation of the $\mathcal{H}_2$ norm associated with the dynamical mismatch gap $\mathbf{G} - \hat{\mathbf{H}}_d$, which conditions Problem 2.1 criterion.

Proposition 3.2. *Let* $\mathbf{G},\ \hat{\mathbf{H}}_d \in \mathcal{H}_2^{n_y\times n_u}$ *s.t.* $\hat{\mathbf{H}}_d$ *is given by Eq.* (4)*. The* $\mathcal{H}_2$ *norm of the approximation gap (or mismatch error), denoted by* $\mathcal{J}$*, can be expressed as:*

$$\begin{aligned}
\mathcal{J} &= \|\mathbf{G} - \hat{\mathbf{\Delta}}_o\hat{\mathbf{H}}\hat{\mathbf{\Delta}}_i\|^2_{\mathcal{H}_2} \\
&= \|\mathbf{G}\|^2_{\mathcal{H}_2} - 2\langle \mathbf{G}, \hat{\mathbf{\Delta}}_o\hat{\mathbf{H}}\hat{\mathbf{\Delta}}_i\rangle_{\mathcal{H}_2} + \|\hat{\mathbf{H}}\|^2_{\mathcal{H}_2}.
\end{aligned} \tag{8}$$

Proof. Simply develop the $\mathcal{H}_2$ norm using the inner product definition and exploit the previous result $\|\hat{\mathbf{\Delta}}_o\hat{\mathbf{H}}\hat{\mathbf{\Delta}}_i\|_{\mathcal{H}_2} = \|\hat{\mathbf{H}}\|_{\mathcal{H}_2}$. $\square$

Obviously, regarding Eq. (8), minimizing $\mathcal{J}$ is equivalent to minimize $-2\langle \mathbf{G}, \hat{\mathbf{\Delta}}_o\hat{\mathbf{H}}\hat{\mathbf{\Delta}}_i\rangle_{\mathcal{H}_2} + \|\hat{\mathbf{H}}\|^2_{\mathcal{H}_2}$ and thus to look for the optimal values of the decision variables contained in both the realization $\hat{\mathbf{H}} \in \mathcal{H}_2^{n_y\times n_u}$ and the delay blocks $\hat{\mathbf{\Delta}}_i,\ \hat{\mathbf{\Delta}}_o \in \mathcal{H}_\infty^{n_y\times n_u}$. At this point, it could be profitable to derive suitable analytical expressions for the inner-product and the $\mathcal{H}_2$ norm of $\hat{\mathbf{H}}$ in order to define more precisely the aforementioned $\mathcal{H}_2$ gap between the two transfer functions. To this aim, the previous assumption made for both $\mathbf{G}$ and $\hat{\mathbf{H}}$ systems (see Eq. (6)) will be essential to obtain the following results.

Proposition 3.3. **($\mathcal{H}_2$ inner product computation with input/output delays)** *Let* $\mathbf{G},\ \hat{\mathbf{H}}$ *be two systems* $\in \mathcal{H}_2^{n_y\times n_u}$ *whose respective transfer functions* $\mathbf{G}(s)$ *and* $\hat{\mathbf{H}}(s)$ *can be expressed as in* (6)*. Let* $\hat{\mathbf{\Delta}}_i,\ \hat{\mathbf{\Delta}}_o$ *be real,* $\mathcal{H}_\infty^{n_u\times n_u}$ *and* $\mathcal{H}_\infty^{n_y\times n_y}$ *respectively, models satisfying* $\sup\{\|\hat{\mathbf{\Delta}}_o(s), \|\hat{\mathbf{\Delta}}_i(s)\|/s \in \mathbb{C}_-\} = M < +\infty$*. By denoting* $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o\hat{\mathbf{H}}\hat{\mathbf{\Delta}}_i$*, the inner product* $\langle \hat{\mathbf{H}}_d, \mathbf{G}\rangle_{\mathcal{H}_2}$ *reads:*

$$\begin{aligned}
\langle \hat{\mathbf{H}}_d, \mathbf{G}\rangle_{\mathcal{H}_2} &= \sum_{j=1}^{N} \mathit{trace}\left(\mathrm{Res}\left[\hat{\mathbf{H}}_d(-s)\mathbf{G}^T(s), \mu_j\right]\right) \\
&= \sum_{j=1}^{N} \mathbf{l}_j^T\hat{\mathbf{\Delta}}_o(-\mu_j)\hat{\mathbf{H}}(-\mu_j)\hat{\mathbf{\Delta}}_i(-\mu_j)\mathbf{r}_j.
\end{aligned} \tag{9}$$

Proof. Observing that the poles of the complex function $\hat{\mathbf{H}}_d(-s)\mathbf{G}(s)$ are $\mu_1,\ \mu_2,\ \ldots,\ \mu_N \in \mathbb{C}_-$ and $-\hat{\lambda}_1,\ -\hat{\lambda}_2,\ \ldots,\ -\hat{\lambda}_n \in \mathbb{C}_+$, let us consider the following semi-circular contour Γ_C located in the left half plane *s.t.*:

$$\Gamma_C = \Gamma_I \cup \Gamma_R$$

with: $\begin{cases} \Gamma_I = \{s \in \mathbb{C}/s = i\omega \text{ and } \omega \in [-R; R],\ R \in \mathbb{R}_+\} \\ \Gamma_R = \{s \in \mathbb{C}/s = Re^{i\theta} \text{ where } \theta \in [\pi/2; 3\pi/2]\} \end{cases}$.

Thus, for a sufficient large radius value R, the Γ_C contour will contain all the poles of the transfer function $\mathbf{G}(s)$ *i.e.,* $\mu_1,\ \mu_2,\ \ldots,\ \mu_N$. Thus, by applying the residues theorem, it follows that:

$$\begin{aligned} \langle \hat{\mathbf{H}}_d, \mathbf{G} \rangle_{\mathcal{H}_2} &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \text{trace}\left(\overline{\hat{\mathbf{H}}_d(i\omega)} \mathbf{G}^T(i\omega) \right) d\omega \\ &= \lim_{R \to +\infty} \frac{1}{2i\pi} \int_{\Gamma_C} \hat{\mathbf{H}}_d(-s)\mathbf{G}(s) ds \\ &= \sum_{j=1}^{N} \text{trace}\left(\text{Res}\left[\hat{\mathbf{H}}_d(-s)\mathbf{G}^T(s), \mu_j \right] \right). \end{aligned}$$

where Res(.) denotes the residue operator. The second equality line holds true since:

$$\int_{\Gamma_R} \left\| \hat{\mathbf{H}}_d(-s)\mathbf{G}(s) ds \right\| \leq M^2 \int_{\Gamma_R} \left\| \hat{\mathbf{H}}(-s)\mathbf{G}(s) ds \right\| \to 0^+,$$

when $R \to +\infty$. □

One may note that Proposition 3.3 is a generalization of Lemma 3.5 appearing in [1] in the case of **MIMO** systems with multiple-input/output delays. It is noteworthy that the $\hat{\mathbf{\Delta}}_i$, $\hat{\mathbf{\Delta}}_o$ matrices defined by (5) clearly verifies the hypothesis Proposition 3.3.

Remark 3.1. (Delay-free case "symmetry") *An equivalent proposition was derived in the delay-free case [1]. It can be recovered from Proposition 3.3 by taking* $\hat{\mathbf{\Delta}}_i = \mathbf{I}_{n_u}$ *and* $\hat{\mathbf{\Delta}}_o = \mathbf{I}_{n_y}$. *The result corresponds to the symmetric expression of the inner product i.e., the evaluation of* $\mathbf{G}$ *in the poles of* $\hat{\mathbf{H}}$ *and its associated residues* $\hat{\mathbf{c}}_k$ *and* $\hat{\mathbf{b}}_k$ *s.t.:*

$$\langle \mathbf{G}, \hat{\mathbf{H}} \rangle_{\mathcal{H}_2} = \sum_{k=1}^{n} \hat{\mathbf{c}}_k^T \hat{\mathbf{G}}(-\hat{\lambda}_k) \hat{\mathbf{b}}_k = \sum_{j=1}^{N} \mathbf{l}_j^T \hat{\mathbf{H}}(-\mu_j) \mathbf{r}_j = \langle \hat{\mathbf{H}}, \mathbf{G} \rangle_{\mathcal{H}_2}.$$

In the presence of input/output delays, since the $\mathcal{H}_2$ *norm cannot be approximated using one contour containing the poles of* $\hat{\mathbf{H}}_d$ *only, this result is no longer true. Indeed, it can be easily shown that in this case, the integral on* Γ_R *will depend on a positive exponential argument which will not converge to* 0^+ *when* $R \to +\infty$. *This justifies the assumption that* $\sup\{\|\hat{\mathbf{\Delta}}_o(s), \|\hat{\mathbf{\Delta}}_i(s)\|\|/s \in \mathbb{C}_-\} = M < +\infty$ *and relevance of Proposition 3.3.*

Finally, let us recall the pole(s)/residue(s) $\mathcal{H}_2$ norm formula.

Corollary 3.1. (Poles/residues $\mathcal{H}_2$ norm [1]) *Assume that* $\hat{\mathbf{H}}_d(s)$, $\hat{\mathbf{H}}(s)$ *belong to* $\mathcal{H}_2^{n_y \times n_u}$ *and that* $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o \hat{\mathbf{H}} \hat{\mathbf{\Delta}}_i$. *Besides, suppose that* $\hat{\mathbf{H}}$ *can be expressed such as in* (6)*, then,*

$$\left\| \hat{\mathbf{H}}_d \right\|_{\mathcal{H}_2}^2 = \sum_{k=1}^{n} \hat{\mathbf{c}}_k^T \hat{\mathbf{H}}(-\hat{\lambda}_k) \hat{\mathbf{b}}_k.$$

Proof. See [1]. □

In the next section, the main result, namely $\mathcal{H}_2$ optimality conditions related to Problem 2.1, are firstly established and an interpolation-based algorithm is proposed to numerically compute the approximation $\hat{\mathbf{H}}_d$.

4. Approximation by multiple I/O delays MIMO systems: $\mathcal{H}_2$ optimality conditions

Considering the mathematical formulation of Problem 2.1 and the reduced order system structure $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o \hat{\mathbf{H}} \hat{\mathbf{\Delta}}_i$, where $\hat{\mathbf{H}}(s)$ is given as in (6), the underlying optimization issue that must be solved is parameterized by ($k = 1, \ldots, n$): *(i)* the n pole(s) $\hat{\lambda}_k \in \mathbb{C}_-$; *(ii)* the n bi-tangential directions $(\hat{\mathbf{b}}_k, \hat{\mathbf{c}}_k) \in \mathbb{C}^{n_u \times 1} \times \mathbb{C}^{n_y \times 1}$; and *(iii)* the $n_u + n_y$ delay values $(\hat{\tau}_l, \hat{\gamma}_m)$, $l = 1 \ldots n_u$, $m = 1 \ldots n_y$. Our primary objective consists in rewriting the expression of the $\mathcal{H}_2$ gap $\mathcal{J}$ as a function of these latter parameters which will subsequently facilitate the derivation of the $\mathcal{H}_2$ optimality conditions for Problem 2.1. This forms the topic of the three following propositions and of Theorem 4.1, which stands as the main result of the paper.

Proposition 4.1. *From the preliminary results, the mismatch $\mathcal{H}_2$ gap defined previously in Proposition 3.2 can be equivalently rewritten as:*

$$\begin{aligned}\mathcal{J} = \|\mathbf{G}\|_{\mathcal{H}_2}^2 + &\sum_{k=1}^{n} \hat{\mathbf{c}}_k^T \hat{\mathbf{H}}(-\hat{\lambda}_k) \hat{\mathbf{b}}_k \ldots \\ &-2 \sum_{j=1}^{N} \mathbf{l}_j^T \hat{\mathbf{\Delta}}_o(-\mu_j) \hat{\mathbf{H}}(-\mu_j) \hat{\mathbf{\Delta}}_i(-\mu_j) \mathbf{r}_j.\end{aligned} \tag{10}$$

Proof. The result is immediate. To be established, it requires to develop the $\mathcal{H}_2$ norm expression showing the inner product and then to use both Proposition 3.3 and Corollary 3.1 results. □

From the previous equation (10), the first-order optimality conditions related to the minimization of $\mathcal{J}$ can be analytically computed. The gradient expressions of the $\mathcal{H}_2$ gap *w.r.t.* each parameters (delays, tangential directions and poles) are detailed in the two following propositions. Starting with the simplest calculations, we first derive the gradient of $\mathcal{J}$ *w.r.t.* the delays since the second term of the right-hand side part of (10) is delay-dependent, only.

Proposition 4.2. *The gradients of the $\mathcal{H}_2$ gap $\mathcal{J}$ with respect to the delays read $\forall l = 1 \ldots n_u$, $\forall m = 1 \ldots n_y$:*

$$\left\{\begin{array}{rcl} \nabla_{\hat{\tau}_l} \mathcal{J} & = & -2 \dfrac{\partial \langle \hat{\mathbf{H}}_d, \mathbf{G} \rangle_{\mathcal{H}_2}}{\partial \hat{\tau}_l} \\ & = & -2 \displaystyle\sum_{j=1}^{N} \mu_j \mathbf{l}_j^T \hat{\mathbf{\Delta}}_o(-\mu_j) \hat{\mathbf{H}}(-\mu_j) \mathbf{D}_l \hat{\mathbf{\Delta}}_i(-\mu_j) \mathbf{r}_j, \\ \nabla_{\hat{\gamma}_m} \mathcal{J} & = & -2 \dfrac{\partial \langle \hat{\mathbf{H}}_d, \mathbf{G} \rangle_{\mathcal{H}_2}}{\partial \hat{\gamma}_m} \\ & = & -2 \displaystyle\sum_{j=1}^{N} \mu_j \mathbf{l}_j^T \mathbf{D}_m \hat{\mathbf{\Delta}}_o(-\mu_j) \hat{\mathbf{H}}(-\mu_j) \hat{\mathbf{\Delta}}_i(-\mu_j) \mathbf{r}_j, \end{array}\right.$$

where elements of $\mathbf{D}_l \in \mathbb{R}^{n_u \times n_u}$, $\mathbf{D}_m \in \mathbb{R}^{n_y \times n_y}$, are defined as:

$$[\mathbf{D}_k]_{ij} = \delta_{ijk} = \begin{cases} 1 & \text{if } i = j = k \\ 0 & \text{otherwise} \end{cases}.$$

Proof. The proof is straightforward to establish since both $\hat{\mathbf{\Delta}}_i$ and $\hat{\mathbf{\Delta}}_o$ terms are diagonal matrices and the exponential derivative function is obvious. □

Proposition 4.3. *The gradients of the $\mathcal{H}_2$ gap $\mathcal{J}$ with respect to parameters $\hat{\mathbf{c}}_k$, $\hat{\mathbf{b}}_k$ and $\hat{\lambda}_k$, $\forall k = 1 \dots n$ read:*

$$\begin{cases} \nabla_{\hat{\mathbf{c}}_k}\mathcal{J} &= -2\dfrac{\partial\langle \hat{\mathbf{H}}_d, \mathbf{G}\rangle_{\mathcal{H}_2}}{\partial \hat{\mathbf{c}}_k} + \dfrac{\partial\|\hat{\mathbf{H}}\|^2_{\mathcal{H}_2}}{\partial \hat{\mathbf{c}}_k} \\ &= -2\hat{\mathbf{b}}_k^T\left(\tilde{\mathbf{G}}(-\hat{\lambda}_k) - \hat{\mathbf{H}}(-\hat{\lambda}_k)\right)^T, \\ \nabla_{\hat{\mathbf{b}}_k}\mathcal{J} &= -2\hat{\mathbf{c}}_k^T\left(\tilde{\mathbf{G}}(-\hat{\lambda}_k) - \hat{\mathbf{H}}(-\hat{\lambda}_k)\right), \\ \nabla_{\hat{\lambda}_k}\mathcal{J} &= 2\hat{\mathbf{c}}_k^T\left(\tilde{\mathbf{G}}'(-\hat{\lambda}_k) - \hat{\mathbf{H}}'(-\hat{\lambda}_k)\right)\hat{\mathbf{b}}_k, \end{cases}$$

where:

$$\tilde{\mathbf{G}}(s) = \sum_{j=1}^{N} \hat{\mathbf{\Delta}}_o(-\mu_j)\frac{\mathbf{l}_j^T\mathbf{r}_j}{s-\mu_j}\hat{\mathbf{\Delta}}_i(-\mu_j). \tag{11}$$

and where $\tilde{\mathbf{G}}'$ and $\hat{\mathbf{H}}'$ are the Laplace derivative of $\tilde{\mathbf{G}}$ and $\hat{\mathbf{H}}$, respectively.

Proof. By defining $\tilde{\mathbf{r}}_j = \hat{\mathbf{\Delta}}_i(-\mu_j)\mathbf{r}_j$ and $\tilde{\mathbf{l}}_j^T = \mathbf{l}_j^T\hat{\mathbf{\Delta}}_o(-\mu_j)$ with $j = 1 \dots N$, the $\mathcal{H}_2$ gap can be written as:

$$\begin{aligned} \mathcal{J} &= \|\mathbf{G}\|^2_{\mathcal{H}_2} - 2\sum_{j=1}^{N}\tilde{\mathbf{l}}_j^T\Big(\sum_{m=1}^{n}\frac{\hat{\mathbf{c}}_m\hat{\mathbf{b}}_m^T}{-\mu_j-\hat{\lambda}_m}\Big)\tilde{\mathbf{r}}_j \\ &+ \sum_{k=1}^{n}\hat{\mathbf{c}}_k^T\Big(\sum_{m=1}^{n}\frac{\hat{\mathbf{c}}_m\hat{\mathbf{b}}_m^T}{-\hat{\lambda}_k-\hat{\lambda}_m}\Big)\hat{\mathbf{b}}_k. \end{aligned}$$

Then, calculating the gradients *w.r.t.* $\hat{\mathbf{b}}_l$, $\hat{\mathbf{c}}_l$ and $\hat{\lambda}_l$ ($l = 1 \dots n$) gives:

$$\nabla_{\hat{\mathbf{b}}_l}\mathcal{J} = -2\frac{\partial\langle \hat{\mathbf{H}}_d, \mathbf{G}\rangle_{\mathcal{H}_2}}{\partial \hat{\mathbf{b}}_l} + \frac{\partial\|\hat{\mathbf{H}}\|^2_{\mathcal{H}_2}}{\partial \hat{\mathbf{b}}_l}$$

Thus, by computing both terms on this expression

$$\begin{aligned} \frac{\partial\|\hat{\mathbf{H}}\|^2_{\mathcal{H}_2}}{\partial \hat{\mathbf{b}}_l} &= \sum_{k=1}^{n}\sum_{m=1}^{n}\frac{(\hat{\mathbf{c}}_k^T\hat{\mathbf{c}}_m)}{-\hat{\lambda}_k-\hat{\lambda}_m}\frac{\partial}{\partial\hat{\mathbf{b}}_l}\left(\hat{\mathbf{b}}_m^T\hat{\mathbf{b}}_k\right) \\ &= 2\sum_{k=1}^{n}\frac{\hat{\mathbf{c}}_l^T\hat{\mathbf{c}}_k\hat{\mathbf{b}}_k^T}{-\hat{\lambda}_k-\hat{\lambda}_l} = 2\hat{\mathbf{c}}_l^T\hat{\mathbf{H}}(-\hat{\lambda}_l) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial\langle \hat{\mathbf{H}}_d, \mathbf{G}\rangle_{\mathcal{H}_2}}{\partial \hat{\mathbf{b}}_l} &= \sum_{j=1}^{N}\sum_{m=1}^{n}\frac{(\tilde{\mathbf{l}}_j^T\hat{\mathbf{c}}_m)\tilde{\mathbf{r}}_j^T}{-\mu_j-\hat{\lambda}_m}\nabla_{\hat{\mathbf{b}}_l}\hat{\mathbf{b}}_m \\ &= \hat{\mathbf{c}}_l^T\sum_{j=1}^{N}\frac{\tilde{\mathbf{l}}_j\tilde{\mathbf{r}}_j^T}{-\mu_j-\hat{\lambda}_l} = \hat{\mathbf{c}}_l^T\tilde{\mathbf{G}}(-\hat{\lambda}_l). \end{aligned}$$

one obtains the gradient.

It is noteworthy that $\nabla_{\hat{\mathbf{c}}_l}\mathcal{J}$ can be obtained in the same way as $\nabla_{\hat{\mathbf{b}}_l}\mathcal{J}$. The calculation of $\nabla_{\hat{\lambda}_l}\mathcal{J}$

is straightforwardly derived as follows:

$$\begin{aligned}\nabla_{\hat{\lambda}_l}\mathcal{J} &= -2\sum_{j=1}^{N}\frac{\tilde{\mathbf{l}}_j^T\hat{\mathbf{c}}_l\hat{\mathbf{b}}_l^T\tilde{\mathbf{r}}_j}{(-\hat{\lambda}_l-\mu_j)^2} - \hat{\mathbf{c}}_l^T\hat{\mathbf{H}}'(-\hat{\lambda}_l)\hat{\mathbf{b}}_l \dots \\ &\quad + \sum_{k=1}^{n}\frac{\hat{\mathbf{c}}_k^T\hat{\mathbf{c}}_l\hat{\mathbf{b}}_l^T\hat{\mathbf{b}}_k}{(-\hat{\lambda}_l-\hat{\lambda}_k)^2} \\ &= 2\hat{\mathbf{c}}_l^T\left(\tilde{\mathbf{G}}'(-\hat{\lambda}_l)-\hat{\mathbf{H}}'(-\hat{\lambda}_l)\right)\hat{\mathbf{b}}_l. \qquad \square\end{aligned}$$

Theorem 4.1 gathers all the first-order optimality conditions related to Problem 2.1 and stands as the main result of the paper.

Theorem 4.1. (Delay model approximation first-order $\mathcal{H}_2$ optimality conditions) *Let us consider $\mathbf{G} \in \mathcal{H}_2^{n_y\times n_u}$ whose transfer function is $\mathbf{G}(s) \in \mathbb{C}^{n_y\times n_u}$. Let $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o\hat{\mathbf{H}}\hat{\mathbf{\Delta}}_i$ be a local optimum of Problem 2.1. It is assumed that $\hat{\mathbf{H}} \in \mathcal{H}_2^{n_y\times n_u}$ corresponds to a model with semi-simple poles only and whose transfer function is denoted by $\hat{\mathbf{H}}(s) = \hat{\mathbf{C}}(s\hat{\mathbf{E}}-\hat{\mathbf{A}})^{-1}\hat{\mathbf{B}} \in \mathbb{C}^{n_y\times n_u}$. Let $\hat{\mathbf{\Delta}}_i$, $\hat{\mathbf{\Delta}}_o$ be elements of $\mathcal{H}_\infty^{n_u\times n_u}$ and $\mathcal{H}_\infty^{n_y\times n_y}$, respectively, s.t. Propositions 3.1 and 3.3 are verified. Then, the following equalities hold:*

$$\left\{\begin{array}{l}\hat{\mathbf{H}}(-\hat{\lambda}_k)\hat{\mathbf{b}}_k = \tilde{\mathbf{G}}(-\hat{\lambda}_k)\hat{\mathbf{b}}_k,\\ \hat{\mathbf{c}}_k^T\hat{\mathbf{H}}(-\hat{\lambda}_k) = \hat{\mathbf{c}}_k^T\tilde{\mathbf{G}}(-\hat{\lambda}_k),\\ \hat{\mathbf{c}}_k^T\hat{\mathbf{H}}'(-\hat{\lambda}_k)\hat{\mathbf{b}}_k = \hat{\mathbf{c}}_k^T\tilde{\mathbf{G}}'(-\hat{\lambda}_k)\hat{\mathbf{b}}_k,\end{array}\right. \tag{12}$$

$$\left\{\begin{array}{l}\displaystyle\sum_{j=1}^{N}\mu_j\mathbf{l}_j^T\hat{\mathbf{\Delta}}_o(-\mu_j)\hat{\mathbf{H}}(-\mu_j)\mathbf{D}_l\hat{\mathbf{\Delta}}_i(-\mu_j)\mathbf{r}_j = 0,\\ \displaystyle\sum_{j=1}^{N}\mu_j\mathbf{l}_k^T\mathbf{D}_m\hat{\mathbf{\Delta}}_o(-\mu_j)\hat{\mathbf{H}}(-\mu_j)\hat{\mathbf{\Delta}}_i(-\mu_j)\mathbf{r}_j = 0,\end{array}\right. \tag{13}$$

for all $k = 1\dots n$, $l = 1\dots n_u$ and $m = 1\dots n_y$ where $\tilde{\mathbf{G}}(s)$ is given by (11).

Proof. The interpolation conditions gathered in (12) are deduced by taking $\nabla_{\hat{\mathbf{c}}_l}\mathcal{J} = 0$, $\nabla_{\hat{\mathbf{b}}_l}\mathcal{J} = 0$ and $\nabla_{\hat{\lambda}_l}\mathcal{J} = 0$. Conditions (13) are obtained similarly by taking $\nabla_{\hat{\tau}_l}\mathcal{J} = 0$ and $\nabla_{\hat{\gamma}_m}\mathcal{J} = 0$. $\square$

Theorem 4.1 asserts that any solution of the $\mathcal{H}_2$ model approximation Problem 2.1, denoted by $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o\hat{\mathbf{H}}\hat{\mathbf{\Delta}}_i$ is *s.t.* $\hat{\mathbf{H}}$ satisfies, at the same time, a set of $3n$ bi-tangential interpolation conditions detailed in (12) and another set of $n_u + n_y$ relations on the delays contained in the $\hat{\mathbf{\Delta}}_i$ and $\hat{\mathbf{\Delta}}_o$ diagonal matrices (13).

Remark 4.1. ($\mathcal{H}_2$ optimality conditions in the SISO case) *In the SISO case, all the conditions provided in Theorem 4.1 appear much simpler and can be stated as follows. Considering:*

$$\mathbf{G}(s) = \sum_{j=1}^{N}\frac{\psi_j}{s-\mu_j},\ \hat{\mathbf{H}}_d(s) = \sum_{k=1}^{n}\frac{\phi_k e^{-\tau s}}{s-\hat{\lambda}_k},$$

s.t. $\hat{\mathbf{H}}_d$ is a local optimum of Problem 2.1, then the following conditions hold:

$$\left\{\begin{array}{l}\hat{\mathbf{H}}(-\hat{\lambda}_k) = \tilde{\mathbf{G}}(-\hat{\lambda}_k),\\ \hat{\mathbf{H}}'(-\hat{\lambda}_k) = \tilde{\mathbf{G}}'(-\hat{\lambda}_k),\end{array}\right. \tag{14}$$

$$\sum_{j=1}^{N} \mu_j \psi_j \left(\sum_{k=1}^{n} \frac{\phi_k}{\mu_j + \hat{\lambda}_k} \right) e^{\tau \mu_j} = 0. \tag{15}$$

for all $k = 1 \ldots n$, and where $\tilde{\mathbf{G}}$ is as in (11)*:*

$$\tilde{\mathbf{G}}(s) = \sum_{j=1}^{N} \frac{\psi_j}{s - \mu_j} e^{\tau \mu_j}.$$

Remark 4.2 (Impulse response of $\tilde{\mathbf{G}}(s)$ and advance effect)**.** *The $\mathcal{H}_2$-optimality conditions given in Theorem 4.1 involves a model $\tilde{\mathbf{G}}(s)$ which has a pole-residue decomposition defined by* (11)*. For simplicity, let us consider the SISO case where $\mathbf{G}$ and $\tilde{\mathbf{G}}$ is given by*

$$\mathbf{G}(s) = \sum_{j=1}^{N} \frac{\psi_j}{s - \mu_j} \quad \tilde{\mathbf{G}}(s) = \sum_{j=1}^{N} \frac{\psi_j}{s - \mu_j} e^{\mu_j \tau}.$$

Thus, the the impulse response of $\tilde{\mathbf{G}}(s)$ is

$$\begin{aligned} \tilde{\mathbf{g}}(t) = \sum_{j=1}^{N} \psi_j e^{\mu_j t} e^{\mu_j \tau} \mathbf{1}(t) &= \sum_{j=1}^{N} \psi_j e^{\mu_j (t+\tau)} \mathbf{1}(t) \\ &= \mathbf{g}(t+\tau) \mathbf{1}(t), \ \ t \in \mathbb{R} \end{aligned}$$

where $\mathbf{1}(t)$ corresponds to the Heaviside step function and $\mathbf{g}(t)$ is the impulse response of model $\mathbf{G}(s)$. Therefore, $\tilde{\mathbf{G}}(s)$ behaves as a time advance of $\mathbf{G}(s)$ and correspond to the "causal part" of the model $\mathbf{G}(s)e^{s\tau}$.

4.1. Practical considerations

In this subsection, three considerations about Problem 2.1 and Theorem 4.1 are discussed. These latter are relevant to sketch an algorithm which enables the computation of model $\hat{\mathbf{\Delta}}_o \hat{\mathbf{H}} \hat{\mathbf{\Delta}}_i$ satisfying the optimality conditions of Theorem 4.1. Let us consider that $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o \hat{\mathbf{H}} \hat{\mathbf{\Delta}}_i$ is a local minimum of the $\mathcal{H}_2$ optimization Problem 2.1 where $\hat{\mathbf{H}}$ is given by (6), then:

- **Consideration ❶.** If the matrices $\hat{\mathbf{\Delta}}_o$, $\hat{\mathbf{\Delta}}_i$ and the reduced order model poles $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_n$ are assumed to be known, Problem 2.1 is reduced to a much simpler problem that can be solved, for example, by using the well-known Loewner framework such as in [17];
- **Consideration ❷.** If the delay matrices $\hat{\mathbf{\Delta}}_o$, $\hat{\mathbf{\Delta}}_i$ are known, then Problem 2.1 can be solved by finding a model realization $\hat{\mathbf{H}}$ which satisfies the interpolation conditions (12) of Theorem 4.1, only. This can be done using, for instance, a very efficient iterative algorithm, *e.g.,* **IRKA** (see [1]);
- **Consideration ❸.** Assume that the system realization $\hat{\mathbf{H}}$ has already been determined. It follows that Problem 2.1 is equivalent to look for optimal delays matrices $(\hat{\mathbf{\Delta}}_o^\star, \hat{\mathbf{\Delta}}_i^\star) \in \mathcal{H}_\infty^{n_y \times n_y} \times \mathcal{H}_\infty^{n_u \times n_u}$ *s.t.*:
$$(\hat{\mathbf{\Delta}}_o^\star, \hat{\mathbf{\Delta}}_i^\star) = \underset{(\hat{\mathbf{\Delta}}_o, \hat{\mathbf{\Delta}}_i)}{\operatorname{argmax}} \langle \hat{\mathbf{\Delta}}_o \hat{\mathbf{H}} \hat{\mathbf{\Delta}}_i, \mathbf{G} \rangle_{\mathcal{H}_2}. \tag{16}$$
Interestingly, since $\langle \hat{\mathbf{\Delta}}_o \hat{\mathbf{H}} \hat{\mathbf{\Delta}}_i, \mathbf{G} \rangle_{\mathcal{H}_2} \to 0$ when the delays go to infinity, this problem can be restricted to a compact set and thus a global solution exists.

4.2. *Computational considerations*

An algorithm which allows to numerically compute a model $\hat{\mathbf{H}}_d$ satisfying the previous $\mathcal{H}_2$ optimality conditions is proposed in this subsection. It relies on the considerations above discussed (Section 4.1). Therefore, the proposed approach corresponds to an iterative algorithm in which each iteration can be decomposed in two steps. The first one aims at computing a realization $\hat{\mathbf{H}}$ which satisfies the interpolation conditions (12) while fixing the matrices $\hat{\mathbf{\Delta}}_o$, $\hat{\mathbf{\Delta}}_i$ at their values obtained from the previous iteration. This can be done using, for instance, the **IRKA** algorithm (*Step 4*). In the second step, the resulting $\hat{\mathbf{H}}$ is then exploited to determine the $n_u + n_y$ optimal values for the $\hat{\mathbf{\Delta}}_o$, $\hat{\mathbf{\Delta}}_i$ matrices elements (*Step 5*). This step is achieved by solving the nonlinear optimization problem defined in (16) using an appropriate solver. Then, the whole process is repeated and these two steps performed again until the convergence[2]. At the end of the procedure, the model built will satisfy the $\mathcal{H}_2$ optimality conditions on which Theorem 4.1 relies. This sequential procedure can be summarized such as in Algorithm 1, and referred to as **MIMO IO-dIRKA**.

Algorithm 1 MIMO IO-dIRKA (MIMO Input Output delay IRKA)

Require: A N^{th}-order model $\mathbf{G} \in \mathcal{H}_2^{n_y \times n_u}$, dimension $n \in \mathbb{N}^*$ ($n \ll N$) and initial guesses for both $\hat{\mathbf{\Delta}}_i^{\text{it}=0}, \hat{\mathbf{\Delta}}_o^{\text{it}=0}$.

1: **while** not converged **do**
2: Set $\text{it} \leftarrow \text{it} + 1$
3: Build $\tilde{\mathbf{G}}^{\text{it}}$ as in (11)
4: Build $\hat{\mathbf{H}}^{\text{it}}$ satisfying the bi-tangential interpolation conditions (12) using **IRKA** [1] on $\tilde{\mathbf{G}}^{\text{it}}$
5: Determine $(\hat{\mathbf{\Delta}}_i^\star, \hat{\mathbf{\Delta}}_o^\star)$ which solve (16) using $\hat{\mathbf{H}}^{\text{it}}$
6: Set $\hat{\mathbf{\Delta}}_i^{\text{it}} \leftarrow \hat{\mathbf{\Delta}}_i^\star, \hat{\mathbf{\Delta}}_o^{\text{it}} \leftarrow \hat{\mathbf{\Delta}}_o^\star$
7: **end while**
8: Construct $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o^{\text{it}} \hat{\mathbf{H}}^{\text{it}} \hat{\mathbf{\Delta}}_i^{\text{it}}$

Ensure: $\hat{\mathbf{H}}_d$ satisfies the interpolation conditions of Theorem 4.1.

4.3. *Structured input/output delays*

All the previous results are left unchanged in the case of structured input/output delays *i.e.,* if, for example, delays does not apply on given input(s) and/or output(s) of $\hat{\mathbf{H}}_d$. The results can be derived in a straightforward way, without any loss of generality, just by considering the following ordered delays matrices (where delays are present on the first $n_{d1} < n_u$ inputs and $n_{d2} < n_y$ outputs):

$$\begin{cases} \hat{\mathbf{\Delta}}_i(s) = \mathsf{diag}(e^{-s\hat{\tau}_1},\ e^{-s\hat{\tau}_2},\ \ldots,\ e^{-s\hat{\tau}_{n_{d1}}},\ 1,\ \ldots,\ 1) \\ \hat{\mathbf{\Delta}}_o(s) = \mathsf{diag}(e^{-s\hat{\gamma}_1},\ e^{-s\hat{\gamma}_2},\ \ldots,\ e^{-s\hat{\gamma}_{n_{d2}}},\ 1,\ \ldots,\ 1). \end{cases}$$

One can easily note that the preliminary results from Sections 3 and 4 still remain true when introducing these matrices. The main result stated in Theorem 4.1 thus remains unchanged.

[2] In practice, different stopping criteria might be considered, *e.g.* *(i)* the variation of the interpolation points materialized by $\hat{\lambda}_k$ ($k = 1, \ldots, n$), as in [1], *(ii)* the interpolation conditions check (Theorem 4.1) or *(iii)* the mismatch $\mathcal{H}_2$ error check (if the order N of the original system is reasonably low).

5. Numerical application

This section is dedicated to the application of the results obtained in Sections 4, namely, the input/output-delay optimal $\mathcal{H}_2$ model approximation and its first -order optimality conditions. We will emphasize the potential benefit and effectiveness of the proposed approach.

Let us consider a model $\mathbf{G}$ of order $N = 20$, given by the following transfer function

$$\mathbf{G}(s) = \prod_{j=1}^{N} \frac{\mu_j}{s - \mu_j}, \tag{17}$$

where $\mu_j \in \mathbb{R}_-$ $(j = 1, \ldots, N)$ are linearly spaced between $[-2\ \ -1]$. The impulse response of $\mathbf{G}$ is given by the solid dotted blue line in Figure 1. Interestingly, it behaves like a system with an input delay. In order to fit the framework proposed in this paper, input-delay $\mathcal{H}_2$ optimal model $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o \hat{\mathbf{H}} \hat{\mathbf{\Delta}}_i$ of order $n = 2$ (solid red) was obtained by applying Theorem 4.1 and **IO-dIRKA**, as described in Section 4. The obtained delay model is compared with delay-free approximations of order $n = \{2, 3, 4\}$, obtained with **IRKA**[3]. All the results are reported on Figure 1.

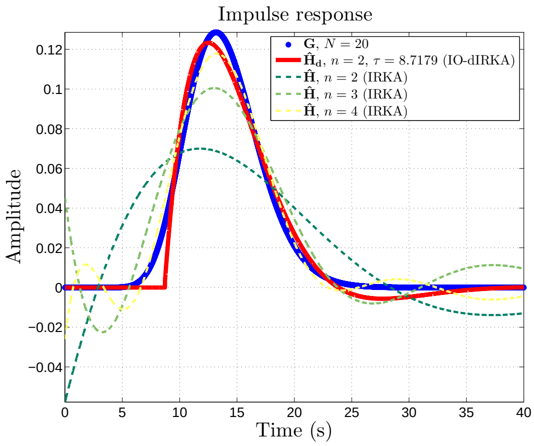

Figure 1: Impulse response of the original model $\mathbf{H}$ of order $N = 20$ (solid dotted blue line), the input-delay $\mathcal{H}_2$-optimal model $\hat{\mathbf{H}}_d$ of order $n = 2$ (solid red line) and the delay-free $\mathcal{H}_2$-optimal models $\hat{\mathbf{H}}$ of order $n = \{2, 3, 4\}$ (dashed dark green, light green and yellow lines).

As clearly shown on Figure 1, the proposed methodology allows to obtain an input-delay $\mathcal{H}_2$ approximation of model $\mathbf{G}$ that clearly provides a better matching than the delay-free cases, even for higher orders (here, **IRKA** with $n = 4$ still have a bad matching and exhibits difficulties in accurately catching the delay and main dynamics). Indeed, the delay-free cases exhibits an oscillatory behaviour during the first seconds while the input-delay model $\hat{\mathbf{H}}_d$ takes benefit of the delay structure to focus on the main dynamical effect. Moreover, the approximation model of $\hat{\mathbf{H}}_d$ satisfies the conditions given in Theorem 4.1.

Remark 5.1 (Numerical results (SISO case, $n = 2$))**.** *For sake of completeness, the optimal numerical values obtained with **MIMO IO-dIRKA** are:* $\hat{\lambda}_{1,2} = -2.0320 \times 10^{-1} \pm i\, 2.0700 \times 10^{-1}$,

[3] Using the implementation available in the **MORE toolbox** [18], `http://w3.onera.fr/more/`.

$\hat{\phi}_{1,2} = 1.5713 \times 10^{-3} \pm i\ 1.8691 \times 10^{-1}$ *and the optimal delay* $\tau = 8.7179$. *The interpolation conditions can then easily be checked:*

- *Condition* (14) *leads to* $\hat{\mathbf{H}}(-\hat{\lambda}_{1,2}) = \tilde{\mathbf{G}}(-\hat{\lambda}_{1,2}) = 2.3567 \times 10^{-1} - i\ 2.3614 \times 10^{-1}$ *and* $\hat{\mathbf{H}}'(-\hat{\lambda}_{1,2}) = \tilde{\mathbf{G}}'(-\hat{\lambda}_{1,2}) = 5.6466 \times 10^{-1} \pm i\ 1.1465$.
- *When evaluating* $\sum_{j=1}^{N} \mu_j \psi_j \left(\sum_{k=1}^{n} \frac{\phi_k}{\mu_j + \hat{\lambda}_k} \right) e^{\tau \mu_j}$, *one obtains* 9.7284×10^{-5}, *which is close to zero, as stated by condition* (15).

With reference to Figure 2, similar results are obtained in the case of an input delay-dependent approximation of order $n = 4$ (using **IO-dIRKA**) and delay-free approximation of order $n = \{4, 5, 6\}$ (using **IRKA**). Then, Figure 3 shows the impulse response mismatch error for these different configurations. For each reduced order models, the mean square absolute error ε of the impulse response are computed. The main observation that can be made is that the mismatch error obtained for $\hat{\mathbf{H}}_d$ of order $n = 4$ is lower that the one obtained by a delay-free model $\hat{\mathbf{H}}$ of order $n = 6$ (a better result is obtained for a delay-free model with an order $n = 7$). This motivates the use of the specific approximation model delay structure.

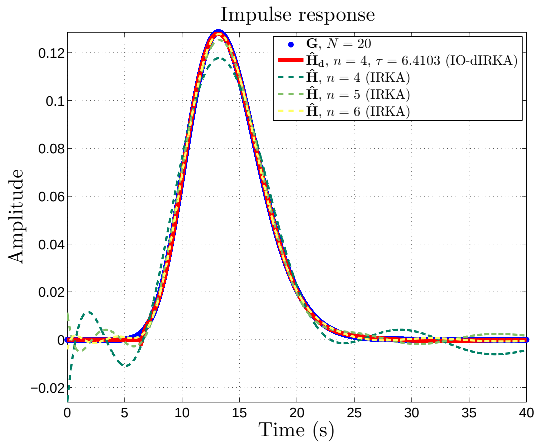

Figure 2: Impulse response of the original model $\mathbf{H}$ of order $N = 20$ (solid dotted blue line), the input-delay $\mathcal{H}_2$-optimal model $\hat{\mathbf{H}}_d$ of order $n = 4$ (solid red line) and the delay-free $\mathcal{H}_2$-optimal models $\hat{\mathbf{H}}$ of order $n = \{4, 5, 6\}$ (dashed dark green, light green and yellow lines).

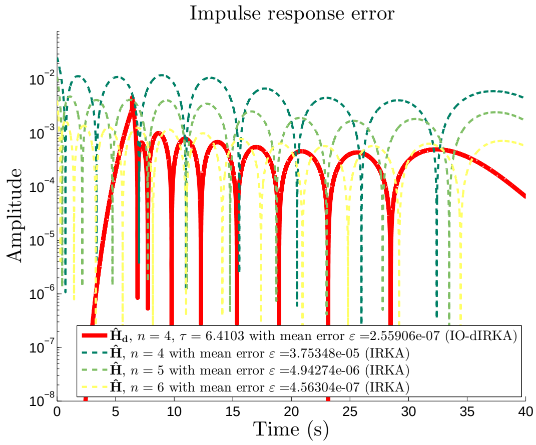

Figure 3: Impulse response error between the original model $\mathbf{H}$ of order $N = 20$ and the input-delay $\mathcal{H}_2$-optimal model $\hat{\mathbf{H}}_d$ of order $n = 4$ (solid red line) and the delay-free $\mathcal{H}_2$-optimal models $\hat{\mathbf{H}}$ of order $n = \{4, 5, 6\}$ (dashed dark green, light green and yellow lines).

6. Conclusion

The main contribution of this paper is the derivation of the first-order $\mathcal{H}_2$ optimality conditions for Problem 2.1. It forms a direct extension of the bi-tangential interpolation conditions of the delay-free case derived in [1, 2]. Theorem 4.1 establishes that if $\hat{\mathbf{H}}_d = \hat{\mathbf{\Delta}}_o \hat{\mathbf{H}} \hat{\mathbf{\Delta}}_i$ is a local optimum, then the parameters of this latter verify an extended set of matricial equalities. These ones are of two types: first, *(i)* a subset of interpolation conditions (12) satisfied by the rational part $\hat{\mathbf{H}}$ of $\hat{\mathbf{H}}_d$, which generalizes the delay-free case; secondly, *(ii)* a subset of matricial relationships (13) focussing on the input/output delay blocks $\hat{\mathbf{\Delta}}_o$, $\hat{\mathbf{\Delta}}_i$. These conditions all are dependent on the reduced order model parametrization described by $\hat{\mathbf{b}}_k$, $\hat{\mathbf{c}}_k$, $\hat{\lambda}_k$, $\hat{\tau}_l$ and $\hat{\gamma}_m$, and solving Problem 2.1 requires to tackle a non-convex optmization problem. Nevertheless, an algorithm referred to as **IO-dIRKA**, has been proposed to practically address this issue. This latter decorrelates the decision variables between them by solving, firstly for given $\hat{\mathbf{\Delta}}_i$, $\hat{\mathbf{\Delta}}_o$ matrices, an optimal $\mathcal{H}_2$ approximation problem, and then, in a second stage, a nonlinear maximization problem (16) to determine the optimal values of the delays. Both optimizations rely on descent methods, taking benefits from the analytical expressions of the gradients of the $\mathcal{H}_2$ mismatch gap $\nabla\mathcal{J}$. Numerical experiment have also been presented, illustrating the benefit of the proposed approximation delay structure with respect to standard delay-free approximation methods.

References

[1] S. Gugercin, A. C. Antoulas, C. Beattie, $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems, SIAM Journal on matrix analysis and applications 30 (2) (2008) 609–638.

[2] P. Van Dooren, K. A. Gallivan, P.-A. Absil, $\mathcal{H}_2$-optimal model reduction of MIMO systems, Applied Mathematics Letters 21 (12) (2008) 1267–1273.

[3] J. T. Spanos, M. H. Milman, D. L. Mingori, A new algorithm for l_2 optimal model reduction, Automatica 28 (5) (1992) 897–909.

[4] D. C. Hyland, D. S. Bernstein, The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore, IEEE Transactions on Automatic Control 30 (12) (1985) 1201–1211.

[5] D. A. Wilson, Optimum solution of model-reduction problem, in: Proceedings of the Institution of Electrical Engineers, Vol. 117, IET, 1970, pp. 1161–1165.
[6] L. Meier III, D. G. Luenberger, Approximation of linear constant systems, IEEE Transactions on Automatic Control 12 (5) (1967) 585–588.
[7] C. A. Beattie, S. Gugercin, A trust region method for optimal $\mathcal{H}_2$ model reduction, in: Proceedings of the 48th Conference on Decision and Control, IEEE, 2009, pp. 5370–5375.
[8] E. J. Grimme, Krylov projection methods for model reduction, Ph.D. thesis, Ph.D. thesis, University of Illinois, Urbana-Champaign, Urbana, IL (1997).
[9] A. Astolfi, Model reduction by moment matching for linear and nonlinear systems, IEEE Transactions on Automatic Control 55 (10) (2010) 2321–2336.
[10] A. C. Antoulas, D. C. Sorensen, S. Gugercin, A survey of model reduction methods for large-scale systems, Contemporary mathematics 280 (2001) 193–220.
[11] A. C. Antoulas, Approximation of large-scale dynamical systems, Vol. 6, SIAM, 2005.
[12] M. Opmeer, Optimal model reduction for non-rational functions, in: 14th annual European Control Conference July 15th, Vol. to appear, 2015.
[13] Y. Halevi, Reduced-order models with delay, International journal of control 64 (4) (1996) 733–744.
[14] G. Scarciotti, A. Astolfi, Model reduction by moment matching for linear time-delay systems, in: 19th IFAC World Congress, Cape Town, South Africa, August 24, Vol. 29, 2014.
[15] I. Pontes Duff, C. Poussot-Vassal, C. Seren, Realization independent single time-delay dynamical model interpolation and $\mathcal{H}_2$-optimal approximation, in: Proceedings of the 54st IEEE Conference on Decision and Control, 2015.
[16] P. Schulze, B. Unger, Data-driven interpolation of dynamical systems with delay, Preprint-Reihe des Instituts fur Mathematik, Technische Universitat Berlin.
[17] A. J. Mayo, A. C. Antoulas, A framework for the solution of the generalized realization problem, Linear algebra and its applications 425 (2) (2007) 634–662.
[18] C. Poussot-Vassal, P. Vuillemin, Introduction to MORE: a MOdel REduction Toolbox, in: Proceedings of the IEEE Multi-conference on Systems and Control, Dubrovnik, Croatia, 2012, pp. 776–781.