Poles Residues Descent Algorithm for Optimal Frequency-Limited $\mathcal{H}_2$ Model Approximation

Pierre Vuillemin, Charles Poussot-Vassal and Daniel Alazard

***Abstract*— Model approximation of multiple-inputs/multiple-outputs (MIMO) linear dynamical systems over a bounded frequency range can be expressed as an optimization problem in terms of the frequency-limited $\mathcal{H}_2$-norm. In this paper, a new formulation of the frequency-limited $\mathcal{H}_2$ model approximation error is presented and its gradient derived. It is then used in a descent algorithm which does not require to solve any Lyapunov equations but one eigenvalue problem for the full-order model. The efficiency of the method is illustrated through numerical benchmarks.**

I. INTRODUCTION

A. Context and problem statement

The $\mathcal{H}_2$ optimal model approximation problem has been extensively addressed and several effective algorithms have been developed [18], [7], [5], [12] and successfully applied to industrial use case (see for instance [13]). The underlying framework for these approaches is the interpolation of complex functions by means of particular Krylov subspaces construction and Petrov-Galerkin conditions enforcement.

Whenever the entire frequency response of the model is not needed or not accurately known, approximating the full-order model such that the reduced-order one accurately reproduces its behaviour over a bounded frequency range *only* can be more adequate [20]. Here, an optimization procedure based on the poles/residues representation of the system is developed to achieve this objective. The idea is similar to what have been done in [3] for the $\mathcal{H}_2$ case.

Problem 1: Let us consider an asymptotically stable MIMO LTI continuous dynamical systems of order n described by its transfer function $H \in \mathcal{H}_\infty$[1],

$$H(s) = C(sI_n - A)^{-1}B + D \in \mathcal{H}_\infty, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n_u}$, $C \in \mathbb{R}^{n_y\times n}$ and $D \in \mathbb{R}^{n_y\times n_u}$. The goal is to find a reduced order model of order $r \ll n$, described by

$$\hat{H}(s) = \hat{C}(sI_r - \hat{A})^{-1}\hat{B} + \hat{D} \in \mathcal{H}_\infty, \tag{2}$$

where $\hat{A} \in \mathbb{R}^{r\times r}$, $\hat{B} \in \mathbb{R}^{r\times n_u}$, $\hat{C} \in \mathbb{R}^{n_y\times r}$ and $\hat{D} \in \mathbb{R}^{n_y\times n_u}$ such that $\hat{H}(s)$ minimizes the frequency-limited $\mathcal{H}_2$-norm, denoted $\mathcal{H}_{2,\omega}$-norm (see Definition 1), of the error system, *i.e.*

$$\|H - \hat{H}\|^2_{\mathcal{H}_{2,\omega}} = \min_{G\in\mathcal{H}_\infty} \|H - G\|^2_{\mathcal{H}_{2,\omega}} = \min_{G\in\mathcal{H}_\infty} \mathcal{J}_{\mathcal{H}_{2,\omega}}. \tag{3}$$

Pierre Vuillemin and Daniel Alazard are with Université de Toulouse and Onera - The French Aerospace Lab, F-31055 Toulouse, France, pierre.vuillemin@onera.fr.

Charles Poussot-Vassal is with Onera - The French Aerospace Lab, F-31055 Toulouse, France.

[1] $\mathcal{H}_\infty$ denotes the set of $n_y \times n_u$ matrix valued functions which entries are analytic in the open right half plane and bounded on the imaginary axis.

Problem 1 has recently been addressed in [11] through the gramian formulation of the $\mathcal{H}_{2,\omega}$-norm. Grounded on this formulation, the authors have derived a gradient which can be used in an optimization scheme. In this paper, the recently proposed spectral formulation of the $\mathcal{H}_{2,\omega}$-norm [19] is used to express the *approximation error* and to derive a *gradient*. Both are then used in a *descent algorithm*. Instead of Lyapunov equations and Fréchet derivatives, this formulation requires to solve one eigenvalue problem for the full-order model.

B. Notations and paper structure

The paper is divided as follows: Section II recalls preliminary results related to the $\mathcal{H}_{2,\omega}$-norm. Then, in Section III, new expressions for the $\mathcal{H}_{2,\omega}$ approximation error and its gradient are presented, as well as a complex optimization algorithm scheme for optimal $\mathcal{H}_{2,\omega}$ model approximation. In Section IV, the proposed approach is applied on standard benchmark models to illustrate its efficiency. Section V finally concludes this paper.

The notations used throughout this paper are the following: elements with $\tilde{.}$ are related to the error model, elements with $\hat{.}$ refer to the reduced order model, λ_i and ϕ_i are the i-th pole and residue of the system, respectively. j denotes the complex variable, *i.e.* $j^2 = -1$, $\mathbf{Re}(z)$ denotes the real part of z and $\mathbf{atan}\,(z)$ is the complex arctangent of z (see [8], [19] for further details). A^T is the transpose of A, A^H its conjugate transpose and A^* its conjugate. $\mathbf{tr}\,(A)$ and $\mathbf{diag}(A)$ denote the trace and diagonal of A respectively, the Hadamard product of matrices is denoted $\odot$, the element of A located at the i-th row and the j-th column is denoted $[A]_{ij}$.

II. PRELIMINARY RESULTS ON THE $\mathcal{H}_{2,\omega}$-NORM

A. Generalities

The $\mathcal{H}_{2,\omega}$-norm has been suggested in [1] as the restriction of the $\mathcal{H}_2$-norm over a bounded frequency range (see Definition 1). It has recently been used in robust analysis [10] and in model reduction [11] and is of great interest in many applicative problem such as aircraft model approximation and control.

Definition 1 ($\mathcal{H}_{2,\omega}$-norm): Given a continuous MIMO LTI dynamical system $H(s)$ described as in (1), the frequency-limited $\mathcal{H}_2$-norm of $H(s)$, denoted $\|H\|_{\mathcal{H}_{2,\omega}}$, is the restriction of its $\mathcal{H}_2$-norm over a bounded frequency range $[0\ \omega]$, $\omega \in \mathbb{R}^+$, and is given as

$$\|H\|^2_{\mathcal{H}_{2,\omega}} := \frac{1}{2\pi}\int_{-\omega}^{\omega} \mathbf{tr}\left(H(j\nu)H(-j\nu)^T\right) d\nu. \tag{4}$$

$$\begin{aligned}\mathcal{J}_{\mathcal{H}_{2,\omega}} &= \sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\mathbf{tr}\left(\phi_i\phi_k^T\right)}{\lambda_i+\lambda_k}\mathbf{a}_{\omega,i}+\sum_{i=1}^{r}\sum_{k=1}^{r}\frac{\mathbf{tr}\left(\hat{\phi}_i\hat{\phi}_k^T\right)}{\hat{\lambda}_i+\hat{\lambda}_k}\hat{\mathbf{a}}_{\omega,i}-\sum_{i=1}^{n}\sum_{k=1}^{r}\frac{\mathbf{tr}\left(\phi_i\hat{\phi}_k^T\right)}{\lambda_i+\hat{\lambda}_k}\left(\mathbf{a}_{\omega,i}+\hat{\mathbf{a}}_{\omega,k}\right)\dots\\ &\quad+\sum_{i=1}^{r}\mathbf{tr}\left(\hat{\phi}_i\tilde{D}^T\right)\hat{\mathbf{a}}_{\omega,i}-\sum_{i=1}^{n}\mathbf{tr}\left(\phi_i\tilde{D}^T\right)\mathbf{a}_{\omega,i}+\frac{\omega}{\pi}\mathbf{tr}\left(\tilde{D}\tilde{D}^T\right)\end{aligned} \tag{18}$$

The $\mathcal{H}_{2,\omega}$-norm is usually computed by means of the frequency-limited gramians [4], but another formulation based on the eigenvalues and residues of the model has been recently proposed by the authors in [19]. For stable systems with simple poles only, $\|H\|^2_{\mathcal{H}_{2,\omega}}$ can indeed be written as

$$\|H\|^2_{\mathcal{H}_{2,\omega}}=\mathbf{tr}\left(\sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\phi_i\phi_k^T}{\lambda_i+\lambda_k}\mathbf{a}_{\omega,i}+\frac{\omega}{\pi}DD^T\dots-\sum_{i=1}^{n}\phi_iD^T\mathbf{a}_{\omega,i}\right) \tag{5}$$

where $\mathbf{a}_{\omega,i}=\frac{2}{\pi}\mathbf{atan}\left(\frac{\omega}{\lambda_i}\right)$ and $\phi_i=\lim_{s\to\lambda_i}(s-\lambda_i)H(s)$.

Remark 1: (Strictly proper case) If $H(s)\in\mathbb{C}^{n_y\times n_u}$ is strictly proper, then (5) becomes

$$\|H\|^2_{\mathcal{H}_{2,\omega}}=\mathbf{tr}\left(\sum_{i=1}^{n}-\mathbf{a}_{\omega,i}\phi_iH^T(-\lambda_i)\right), \tag{6}$$

which is very similar to the $\mathcal{H}_2$ expression [2, chap. 5],

$$\|H(s)\|^2_{\mathcal{H}_2}=\mathbf{tr}\left(\sum_{i=1}^{n}\phi_iH^T(-\lambda_i)\right), \tag{7}$$

excepted from the weighting coefficients $\mathbf{a}_{\omega,i}$, $i=1,\dots,n$, specific to the $\mathcal{H}_{2,\omega}$-norm.

Remark 2 (Frequency interval): The $\mathcal{H}_2$-norm can also be restricted to the interval $\Omega=[\omega_1\ \omega_2]$. Indeed

$$\|H\|^2_{\mathcal{H}_{2,\Omega}}=\|H\|^2_{\mathcal{H}_{2,[\omega_1,\omega_2]}}=\|H\|^2_{\mathcal{H}_{2,\omega_2}}-\|H\|^2_{\mathcal{H}_{2,\omega_1}}. \tag{8}$$

As a consequence, all the results presented in this paper are directly extensible to more complex frequency intervals.

Note that if $H(s)$ is stable and strictly proper then,

$$\lim_{\omega\to\infty}\|H\|_{\mathcal{H}_{2,\omega}}=\|H\|_{\mathcal{H}_2}. \tag{9}$$

Thus model approximation in the $\mathcal{H}_{2,\omega}$-norm generalises the $\mathcal{H}_2$-norm case [19].

B. About the $\mathcal{H}_{2,\omega}$-norm computation

To efficiently compute the $\mathcal{H}_{2,\omega}$-norm (5), the residues ϕ_i should not be explicitly computed. Indeed, by noting $X\in\mathbb{C}^{n\times n}$ the matrix which columns are the right eigenvectors of A, $Y=X^{-1}$, and e_i the canonical basis column vector, it turns out that

$$\begin{aligned}\mathbf{tr}\left(\phi_i\phi_k^T\right) &= \mathbf{tr}\left(CXe_ie_i^TYB(CXe_ke_k^TYB)^T\right)\\ &= e_k^T\underbrace{(CX)^TCX}_{M_1}e_ie_i^T\underbrace{YB(YB)^T}_{M_2}e_k.\end{aligned} \tag{10}$$

Since M_1 and M_2 are symmetric,

$$\mathbf{tr}\left(\phi_i\phi_k^T\right)=[M_1]_{k,_i}[M_2]_{i,k}=[M_1]_{i,k}[M_2]_{i,k}. \tag{11}$$

Consequently, by denoting L the matrix defined by

$$[L]_{i,k}=\frac{1}{\lambda_i+\lambda_k}, \tag{12}$$

it comes that

$$\sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\mathbf{tr}\left(\phi_i\phi_k^T\right)}{\lambda_i+\lambda_k}\mathbf{a}_{\omega,i}=\mathbb{1}^T\left(M_1\odot M_2\odot L\right)\begin{bmatrix}\mathbf{a}_{\omega,1}\\ \vdots\\ \mathbf{a}_{\omega,n}\end{bmatrix}, \tag{13}$$

where $\mathbb{1}$ denotes the column vector filled with ones and $\odot$ the Hadamard product. Similarly,

$$\sum_{i=1}^{n}\mathbf{tr}\left(\phi_iD^T\right)\mathbf{a}_{\omega,i}=\mathbf{diag}\left(YBD^TCX\right)\begin{bmatrix}\mathbf{a}_{\omega,1}\\ \vdots\\ \mathbf{a}_{\omega,n}\end{bmatrix}. \tag{14}$$

III. MAIN RESULTS

In order to derive an efficient optimization scheme, the $\mathcal{H}_{2,\omega}$ mismatch error and its associated gradient are firstly derived, then used to construct a descent procedure. From now on, Problem 1 is restricted to models with simple poles only. Under this assumption, $H(s)$ and $\hat{H}(s)$ can be written as,

$$H(s)=\sum_{i=1}^{n}\frac{\phi_i}{s-\lambda_i}+D \tag{15}$$

where $\mathbf{Re}(\lambda_i)<0$, $i=1,\dots,n$, and

$$\hat{H}(s)=\sum_{i=1}^{r}\frac{\hat{\phi}_i}{s-\hat{\lambda}_i}+\hat{D}, \tag{16}$$

where $\mathbf{Re}(\hat{\lambda}_i)<0$, $i=1,\dots,r$. Formulation (5) of the $\mathcal{H}_{2,\omega}$-norm can then be used to express the $\mathcal{H}_{2,\omega}$ mismatch error.

A. Expression of the $\mathcal{H}_{2,\omega}$ mismatch error, $\mathcal{J}_{\mathcal{H}_{2,\omega}}$

The following Theorem expresses the $\mathcal{H}_{2,\omega}$ mismatch error between the full-order and reduced order models.

Theorem 1: Given a n-th order model $H(s)$ and a r-th order model $\hat{H}(s)$, described by (15) and (16), respectively, the $\mathcal{H}_{2,\omega}$-norm of the approximation error

$$\mathcal{J}_{\mathcal{H}_{2,\omega}}=\|H-\hat{H}\|^2_{\mathcal{H}_{2,\omega}}, \tag{17}$$

is given by relation (18), where $\tilde{D}=D-\hat{D}$, $\mathbf{a}_{\omega,i}=\frac{2}{\pi}\mathbf{atan}\left(\frac{\omega}{\lambda_i}\right)$ and $\hat{\mathbf{a}}_{\omega,i}=\frac{2}{\pi}\mathbf{atan}\left(\frac{\omega}{\hat{\lambda}_i}\right)$.

Proof: See Appendix. ■

for $m = 1, \ldots, r$,

$$\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial \hat{\lambda}_m^*} = -\sum_{i=1}^{r} \hat{c}_m^* \hat{c}_i^T \hat{b}_i \hat{b}_m^H \left(\frac{\hat{\mathbf{a}}_{\omega,i} + \hat{\mathbf{a}}_{\omega,m}^*}{(\hat{\lambda}_i + \hat{\lambda}_m^*)^2} + \frac{2}{\pi} \frac{\omega}{(\hat{\lambda}_m^{*2} + \omega^2)(\hat{\lambda}_i + \hat{\lambda}_m^*)} \right) \ldots + \sum_{i=1}^{n} \hat{c}_m^* c_i^T b_i \hat{b}_m^H \left(\frac{\mathbf{a}_{\omega,i} + \hat{\mathbf{a}}_{\omega,m}^*}{(\lambda_i + \hat{\lambda}_m^*)^2} + \frac{2}{\pi} \frac{\omega}{(\hat{\lambda}_m^{*2} + \omega^2)(\lambda_i + \hat{\lambda}_m^*)} \right) + \frac{2}{\pi} \hat{b}_m^* \tilde{D}^T \hat{c}_m^H \frac{\omega}{\hat{\lambda}_m^{*2} + \omega^2} \tag{19}$$

$$\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial \hat{b}_m^*} = \sum_{i=1}^{r} \frac{\hat{b}_i^T \hat{c}_i \hat{c}_m^H}{\hat{\lambda}_i + \hat{\lambda}_m^*} \left(\hat{\mathbf{a}}_{\omega,i} + \hat{\mathbf{a}}_{\omega,m}^* \right) - \sum_{i=1}^{n} \frac{b_i^T c_i \hat{c}_m^H}{\lambda_i + \hat{\lambda}_m^*} \left(\mathbf{a}_{\omega,i} + \hat{\mathbf{a}}_{\omega,m}^* \right) + \tilde{D}^T \hat{c}_m^H \hat{\mathbf{a}}_{\omega,m}^* \tag{20}$$

$$\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial \hat{c}_m^*} = \sum_{i=1}^{r} \frac{\hat{c}_i^T \hat{b}_i \hat{b}_m^H}{\hat{\lambda}_i + \hat{\lambda}_m^*} \left(\hat{\mathbf{a}}_{\omega,i} + \hat{\mathbf{a}}_{\omega,m}^* \right) - \sum_{i=1}^{n} \frac{c_i^T b_i \hat{b}_m^H}{\lambda_i + \hat{\lambda}_m^*} \left(\mathbf{a}_{\omega,i} + \hat{\mathbf{a}}_{\omega,m}^* \right) + \tilde{D} \hat{b}_m^H \hat{\mathbf{a}}_{\omega,m}^* \tag{21}$$

$$\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial \hat{D}} = -\sum_{i=1}^{r} \hat{c}_i^T \hat{b}_i \hat{\mathbf{a}}_{\omega,i} + \sum_{i=1}^{n} c_i^T b_i \mathbf{a}_{\omega,i} - \frac{2}{\pi} \omega \tilde{D}. \tag{22}$$

With reference to relation (18), one can notice that, as for the $\mathcal{H}_2$ case, the approximation error is composed of the $\mathcal{H}_{2,\omega}$-norm of the full-order model (first term), the $\mathcal{H}_{2,\omega}$-norm of the reduced-order model (second term) and some additional cross terms.

It is important to note that formulation (16) of $\hat{H}(s)$ implicitly imposes that the residues $\hat{\phi}_i$, $i = 1, \ldots, r$, are of rank one. It does not have any impact on the expression of the error, but for the optimization procedure, in order to force them to be of rank one, residues must be written as $\hat{\phi}_i = \hat{c}_i^T \hat{b}_i$ where $\hat{c}_i \in \mathbb{C}^{1 \times n_y}$ and $\hat{b}_i \in \mathbb{C}^{1 \times n_u}$. Problem 1 is thus restricted to the determination of $\hat{c}_i$, $\hat{b}_i$, $\hat{\lambda}_i$, $i = 1, \ldots, r$ and $\hat{D}$ which minimize $\mathcal{J}_{\mathcal{H}_{2,\omega}}$. For notation consistency, the residues ϕ_i of the full-order model $H(s)$ will also be written as the product $\phi_i = c_i^T b_i$ where $c_i \in \mathbb{C}^{1 \times n_y}$ and $b_i \in \mathbb{C}^{1 \times n_u}$, $i = 1, \ldots, n$.

$\mathcal{J}_{\mathcal{H}_{2,\omega}}$ and $\hat{H}$ are thus functions of $r(n_y + n_u + 1) + n_u n_y$ parameters in the general case and $r(n_y + n_u + 1)$ in the strictly proper case. In the $\mathcal{H}_2$ framework described in [3], additional constraints were added to the norms of $\hat{c}_i$ and $\hat{b}_i$ thus reducing the number of decision variables to $r(n_y + n_u)$. Yet, to fit in the complex optimization framework stated in [17], here, no constraint is added.

B. Expression of the conjugate gradient, $\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial z^*}$

$\mathcal{J}_{\mathcal{H}_{2,\omega}}$ is a real function of complex variables $z = \left\{ \hat{\lambda}_i, \hat{c}_i, \hat{b}_i, \hat{D} \right\}$, their complex conjugates z^*. As recalled in [17], such a function is non analytic in z. The common way of dealing with such kind of functions consists in writing them as functions of the real and imaginary parts of z. Yet it is more convenient to keep $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ as a function of complex variables, that is why the complex optimization framework, presented in [17], has been preferred. In this framework, $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ can be minimized indifferently in the complex or real spaces by constructing optimization schemes based on the scaled conjugate cogradient $2\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial z^*}$. That is why, in what follows, $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ (18) is derived with respect to $\hat{\lambda}_i^*$, $\hat{c}_i^*$, $\hat{b}_i^*$, and $\hat{D}$ without differentiating the real or complex parts. The following Theorem holds.

Theorem 2: Given a n-th order model $H(s)$ and a r-th order model $\hat{H}(s)$ described by (15) and (16), respectively, the complex derivatives of the approximation error $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ (18) with respect to $\hat{\lambda}_m^*$, $\hat{b}_m^*$, $\hat{c}_m^*$, and $\hat{D}$ are given by the equations (19), (20), (21) and (22), respectively.

Proof: The gradient is obtained through straight derivation of the error expression (18). ■

As for the $\mathcal{H}_{2,\omega}$-norm, the conjugate gradient of $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ can be computed efficiently through matrix/vectors products (see Section II-B).

C. Descent Algorithm for Residues and Poles Optimization (DARPO)

The error $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ and the gradient $\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial z^*}$ presented in Theorems 1 and 2 are now used in a complex gradient descent algorithm, called **DARPO**. The basic optimization scheme is described in Algorithm 1. The following remarks can be made :

- Step 2 : the initialization is done by selecting a perturbed subset of the full-order model parameters which corresponds to the highest residues in the considered frequency bands. The initialization can also be done with another model approximation algorithm, **DARPO** would then be a refining part.
- Step 4 : several choices exist for the converge criterion. Here, **DARPO** stops when $\left| p_k^H \frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial z^*} |_{z=z_k} \right|$, the derivative of $\mathcal{J}_{\mathcal{H}_{2,\omega}}(z_k + \alpha_k p_k)$ with respect to α_k when $\alpha_k = 0$, falls below a tolerance $\epsilon > 0$.
- Step 6 : the descent direction is symbolically chosen here as the opposite of the conjugate gradient, but in practice, a quasi-newton procedure, like BFGS, is implemented.
- Step 11 : by constructing arbitrary eigenvectors X_r associated to the eigenvalues of the reduced-order model, its state-space representation can be obtained as,

$$\begin{aligned} \hat{A} &= X_r \mathbf{diag}(\hat{\lambda}_1^{(k)}, \ldots, \hat{\lambda}_r^{(k)}) X_r^{-1} \\ \hat{B} &= X_r \begin{bmatrix} \hat{b}_1^{(k)T} & \ldots & \hat{b}_r^{(k)T} \end{bmatrix}^T \\ \hat{C} &= \begin{bmatrix} \hat{c}_1^{(k)T} & \ldots & \hat{c}_r^{(k)T} \end{bmatrix} X_r^{-1}. \end{aligned} \tag{23}$$

- The approximation is done either over a simple frequency band $\Omega = [0\ \omega]$ or over an union of K frequency bands $\Omega = \bigcup_{k=1}^{K} \left[\omega_1^{(k)}\ \omega_2^{(k)}\right]$, $\omega_1^{(k)} < \omega_2^{(k)} < \omega_1^{(k+1)}$ where $\omega_2^{(K)}$ can be infinite if and only if the system is strictly proper.
- Problem 1 is non-linear and non-convex and DARPO is a local optimisation algorithm, hence the result will strongly depends on the initial point.

Algorithm 1 Optimization scheme of **DARPO**

Require: $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n_u}$, $C \in \mathbb{R}^{n_y\times n}$, $D \in \mathbb{R}^{n_y\times n_u}$, $[0\ \omega]$ with $\omega > 0$ and $r \in \mathbb{N}^*$.

1: Compute the eigenvalues of A and their associated right and left eigenvectors X and Y, respectively.
2: Choose an initial point z_0 composed of $\hat{\lambda}_i^{(0)}$, $\hat{c}_i^{(0)}$, $\hat{b}_i^{(0)}$, $i = 1, \ldots, r$, and $\hat{D}^{(0)}$.
3: k $\leftarrow$ 0.
4: **while not converged do**
5: Compute $\mathcal{J}_{\mathcal{H}_{2,\omega}}(z_k)$ and $\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial z^*}|_{z=z_k}$.
6: Set $p_k = -2\frac{\partial \mathcal{J}_{\mathcal{H}_{2,\omega}}}{\partial z^*}|_{z=z_k}$.
7: Choose α_k such that $\mathcal{J}_{\mathcal{H}_{2,\omega}}(z_k + \alpha_k p_k)$ satisfies the complex strong Wolfe conditions [17].
8: Set $z_{k+1} = z_k + \alpha_k p_k$.
9: k $\leftarrow$ k+1.
10: **end while**
11: Use $\hat{\lambda}_i^{(k)}$, $\hat{c}_i^{(k)}$, $\hat{b}_i^{(k)}$, $i = 1, \ldots, r$ to construct $\left\{\hat{A}, \hat{B}, \hat{C}\right\}$.

Note that, grounded on the optimization scheme described in Algorithm 1, **DARPO** can be used in two main ways:

- given an order r, it will find a r-th order model $\hat{H}(s)$ which minimizes $\mathcal{J}_{\mathcal{H}_{2,\omega}}$,
- given a relative $\mathcal{H}_{2,\omega}$ error $E_{\mathcal{H}_{2,\omega}}$, it will find a reduced-order model $\hat{H}(s)$ which satisfies

$$100\frac{\sqrt{\mathcal{J}_{\mathcal{H}_{2,\omega}}}}{\|H\|_{\mathcal{H}_{2,\omega}}} = 100\frac{\|H-\hat{H}\|_{\mathcal{H}_{2,\omega}}}{\|H\|_{\mathcal{H}_{2,\omega}}} \leq E_{\mathcal{H}_{2,\omega}}, \quad (24)$$

by iteratively increasing the order r of $\hat{H}(s)$ and performing an optimization for each r.

Remark 3: The **DARPO** is made available in the MORE Toolbox [14] available at http://w3.onera.fr/more.

Remark 4: Once the eigenvalues and associated eigenvectors of the initial model are computed (step 1), steps 4 to 10 in Algorithm 1 are cheap since they involve matrix/vector and Hadamard products only.

IV. NUMERICAL APPLICATIONS

In this section, the efficiency of **DARPO**, is illustrated through several examples. The method is compared *(i)* to the Frequency-Limited Balanced Truncation (**FL-BT** [4]) and to the Frequency-Limited Iterative SVD-Tangential Interpolation Algorithm (**FL-ISTIA** [20]), in the $\mathcal{H}_{2,\omega}$ case and *(ii)* to the Iterative Tangential Interpolation Algorithm (**ITIA** or MIMO **IRKA** [18], [7]), to the Iterative SVD-Tangential Interpolation Algorithm (ISTIA [5], [12]) and to the Balanced Truncation (**BT**), when $\Omega = [0, \infty]$.

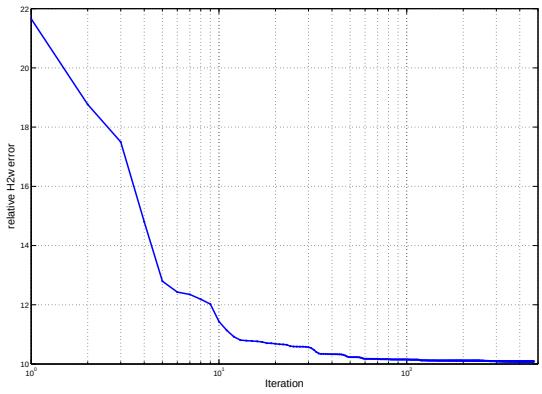

Fig. 1. Evolution of $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ against the iteration using **DARPO** (ISS model, $r = 16$, $\Omega_4 = [0\ \infty]$).

Interval Ω	DARPO	FL-BT or BT	FL-ISTIA	ITIA
[0 3]	**0.3847**	0.7378	0.3871	-
[0 12]	1.4526	**1.2529**	1.4714	-
[12 ∞]	**8.8753**	17.7489	13.9143	-
[0 ∞]	**10.0857**	10.0935	11.9822	15.7460

TABLE I

Relative $\mathcal{H}_{2,\omega}$ errors (in %) obtained after reduction of the ISS model to a 16-th order model over different frequency-bands.

A. Fixed approximation order r

For the first application, the International Space Station (ISS) model [9] is considered. This 270-th order MIMO (3 inputs and 3 outputs) model is reduced to a 16-th order model over the frequency intervals *(i)* $\Omega_1 = [0\ 3]$, *(ii)* $\Omega_2 = [0\ 12]$, *(iii)* $\Omega_3 = [12\ \infty]$ and *(iv)* $\Omega_4 = [0\ \infty]$. The relative errors are reported in Table I.

On this example, the proposed approach leads to slightly better results than other algorithms except when the approximation is done over Ω_2 where **DARPO** directly falls into a local minimum. The initialization is a key step in non convex optimization problems and even if the current selection method is often satisfactory, further studies are required to improve it. In particular, the current initialization method does not enable to create poles in the case where there is not enough poles from the full-order model in the desired frequency band.

Figure 1 shows the decrease of the approximation error $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ for **DARPO** when Ω_4 is considered, highlighting the efficiency of the descent scheme. Indeed the algorithm starts with an error of 21% and ends with 10% only. Moreover it stops with a relative error better than the **ITIA**.

To illustrate what can be achieved when multiple frequency-bands are considered, one input to output transfer of the ISS model is reduced to an 16-th order over $\Omega = [6\ 12] \bigcup [35\ 70]$. The frequency responses of the full-order model and the reduced-order one are plotted in Figure 2. This feature can be useful from an engineering point of view (*e.g.* for control) since it enables to capture precisely some modes which are not necessarily in the same frequency band.

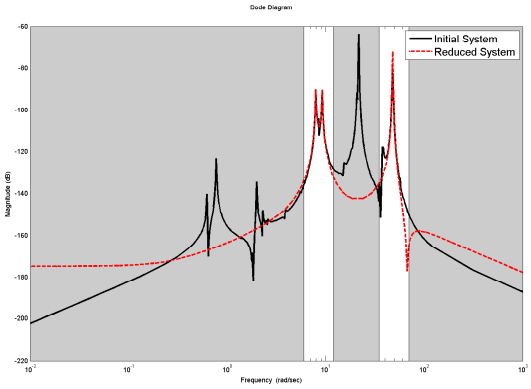

Fig. 2. Frequency responses of the full-order model and the reduced-order one (ISS model,$r = 16$, $\Omega = [6\ 12] \bigcup [35\ 70]$, **DARPO**).

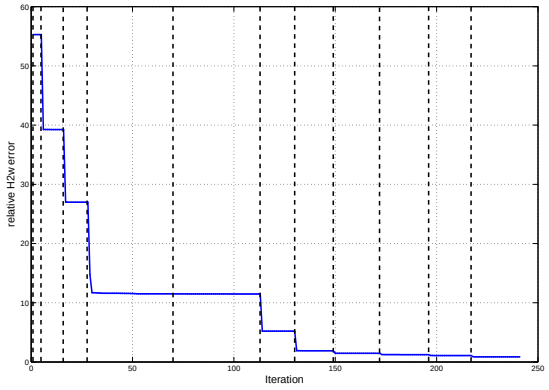

Fig. 3. Decrease of $\mathcal{J}_{\mathcal{H}_{2,\omega}}$ with respect to the iteration when the order is not fixed (ISS model, $\Omega = [0\ 12]$, **DARPO**).

Model-Ω	r	DARPO	FL-BT or BT	FL-ISTIA	ITIA
ISS-[0 12]	22	**0.8629**	0.8651	0.9947	-
ISS-[0 ∞]	36	0.8353	**0.7252**	0.9727	0.9835
LAH-[0 7]	16	**0.1849**	0.9599[2]	0.9269	-
LAH-[0 ∞]	30	0.5858	0.1355	0.1188	**0.1036**
CBM-[0 0.7]	4	**0.7366**	0.7543	0.7624	-
CBM-[0 ∞]	12	**0.5296**	1.0828	1.0551	**0.5296**

TABLE II

Relative $\mathcal{H}_{2,\omega}$ errors (in %) after reduction of the different models to the order r reached by **DARPO** when its relative error falls below 1%.

B. Fixed $\mathcal{H}_{2,\omega}$ error

In this section, three models are reduced by **DARPO** such that a relative error smaller than $E_{\mathcal{H}_{2,\omega}} = 1$ % (over a particular frequency band) is reached. The models are also reduced with other approaches for the resulting order r. The models used here are the ISS model, the Clamped Beam Model (CBM, 270 states, SISO) and the Los Angeles Hospital Model (LAH, 48 states, SISO) [9].

The results are reported in Table II. Moreover the decrease of the approximation error for the ISS model when $\Omega = [0\ 12]$ is plotted in Figure 3 where each black dashed vertical line corresponds to an increase of 2 in the order r.

Table II shows that all the methods lead to similar results for each case. Yet, only **DARPO** enables to really watch the approximation error directly and stop the increase of the approximation order when the relative error requirement is fulfilled while other techniques require a time consuming trial and error approach. As expected, Figure 3 shows that there is an important decrease of the $\mathcal{H}_{2,\omega}$ approximation error each time the order r increases.

C. Very large-scale settings

The main cost of the proposed approach lies in the computation of the full-order model eigenvalues and eigenvectors (step 1 in Algorithm 1). Indeed, to fully compute the $\mathcal{H}_{2,\omega}$-norm of the error and its gradient, all of the eigenvalues and eigenvectors are required (due to the cross terms). Despite the fact that very effective iterative algorithms exist for large-scale eigenvalues problems [16], it can becomes very time and memory consuming to compute the eigenvalues and eigenvectors. In this Section, some hints are given to bypass this difficulty.

[2]Here the result was unstable so the modified version of the frequency-limited balanced truncation, described in [6], has been used.

Ω	[0 3]	[0 12]	[0 30]	[0 50]
DARPO	0.6880(270)	11.5015(270)	18.9904(270)	23.0815(270)
sDARPO	0.6913(24)	11.5019(82)	18.9951(120)	23.0815(214)

TABLE III

Relative $\mathcal{H}_{2,\omega}$ errors (in %) obtained with **DARPO** and **sDARPO** after reduction of the ISS model to an order 10 for several Ω.

The $\mathcal{H}_2$-norm of a large-scale system can be quite well estimated by only considering the highest contribution among the poles/residues couples. These poles/residues couples can be determined efficiently by using the Dominant Poles Algorithm [15]. Yet it still requires a lot of computation. However, in the $\mathcal{H}_{2,\omega}$ case, when the frequency band is bounded, there is potentially less poles/residues couples that have a major impact on the $\mathcal{H}_{2,\omega}$-norm (this is especially true for very flexible structures where only poorly damped modes matter).

As an illustration, let reduce the ISS model with **DARPO** using *(i)* all the eigenvalues and their corresponding eigenvectors and *(ii)* only the eigenvalues which have a magnitude below ω. When only some eigenvalues are used, the algorithm is denoted **sDARPO** for sparse **DARPO**. As the error given by **sDARPO** is not exact, it is recomputed afterwards. The resulting relative errors obtained with **DARPO** and **sDARPO** are reported in Table III, where the number of used eigenvalues is indicated in brackets.

Since the system is very flexible, this procedure works well, indeed there is only few difference between the approximation given by **DARPO** and **sDARPO**. Yet, it might fail for models with a lot of real or highly damped modes (like heat flow model).

V. CONCLUSION

In this paper, model approximation over a bounded frequency range has been addressed as an optimisation problem considering the frequency-limited $\mathcal{H}_2$-norm of the approximation error between the full-order model and the reduced-order one. Based on the poles/residues formulation of the $\mathcal{H}_{2,\omega}$-norm, the approximation error has been expressed as

a function of the poles and residues of the reduced-order model and its gradient has been derived. This formulation enables a fast computation of the error and its gradient granted that the eigenvalues and eigenvectors of the full-order model have been computed. The error and its gradient have then been exploited in a complex descent algorithm which efficiency has been illustrated on different examples. This algorithm also enables to select a reduced-order model based on the approximation error instead of the classical reduced order. Since the eigenvalues and eigenvectors of the full-order model are required, the approach is dedicated to medium-scale models, but some hints have been given concerning its extension in very large-scale settings.

APPENDIX

Proof of Theorem 1

The proof comes from writing the $\mathcal{H}_{2,\omega}$-norm of the error in terms of the residues $\tilde{\phi}_i$ and poles $\tilde{\lambda}_i$, $i = 1, \ldots, n+r$ of the error system $\tilde{H}(s)$,

$$\mathcal{J}_{\mathcal{H}_{2,\omega}} = \sum_{i=1}^{n+r}\sum_{k=1}^{n+r} \frac{\mathbf{tr}\left(\tilde{\phi}_i\tilde{\phi}_k^T\right)}{\tilde{\lambda}_i+\tilde{\lambda}_k}\tilde{\mathbf{a}}_{\omega,i} - \sum_{i=1}^{n+r}\mathbf{tr}\left(\tilde{\phi}_i\tilde{D}^T\right)\tilde{\mathbf{a}}_{\omega,i}\ldots + \tfrac{1}{\pi}\omega\mathbf{tr}\left(\tilde{D}\tilde{D}^T\right). \tag{25}$$

The poles/residues of $\tilde{H}(s)$ are composed of the poles/residues of $H(s)$ and $-\hat{H}(s)$, indeed, by supposing they are ordered as

$$\tilde{\lambda}_i = \begin{cases} \lambda_i & \text{if } i = 1, \ldots, n \\ \hat{\lambda}_i & \text{if } i = n+1, \ldots, n+r, \end{cases} \tag{26}$$

$$\tilde{\phi}_i = \begin{cases} \phi_i & \text{if } i = 1, \ldots, n \\ -\hat{\phi}_i & \text{if } i = n+1, \ldots, n+r, \end{cases} \tag{27}$$

the sums in (25) can then be divided as follow,

$$\begin{aligned}\mathcal{J}_{\mathcal{H}_{2,\omega}} = &\sum_{i=1}^{n}\sum_{k=1}^{n} \frac{\mathbf{tr}\left(\phi_i\phi_k^T\right)}{\lambda_i+\lambda_k}\mathbf{a}_{\omega,i} - \sum_{i=1}^{n}\mathbf{tr}\left(\phi_i\tilde{D}^T\right)\mathbf{a}_{\omega,i}\ldots \\ &+\sum_{i=1}^{n}\sum_{k=1}^{r} \frac{\mathbf{tr}\left(-\phi_i\hat{\phi}_k^T\right)}{\lambda_i+\hat{\lambda}_k}\mathbf{a}_{\omega,i} + \sum_{i=1}^{r}\sum_{k=1}^{n} \frac{\mathbf{tr}\left(-\hat{\phi}_i\phi_k^T\right)}{\hat{\lambda}_i+\lambda_k}\hat{\mathbf{a}}_{\omega,i}\ldots \\ &+\sum_{i=1}^{r}\sum_{k=1}^{r} \frac{\mathbf{tr}\left(\hat{\phi}_i\hat{\phi}_k^T\right)}{\hat{\lambda}_i+\hat{\lambda}_k}\hat{\mathbf{a}}_{\omega,i} + \sum_{i=1}^{r}\mathbf{tr}\left(\hat{\phi}_i\tilde{D}^T\right)\hat{\mathbf{a}}_{\omega,i}\ldots \\ &+\tfrac{1}{\pi}\omega\mathbf{tr}\left(\tilde{D}\tilde{D}^T\right).\end{aligned} \tag{28}$$

Finally, regrouping the cross sums leads to the result,

$$\begin{aligned}\mathcal{J}_{\mathcal{H}_{2,\omega}} = &\sum_{i=1}^{n}\sum_{k=1}^{n} \frac{\mathbf{tr}\left(\phi_i\phi_k^T\right)}{\lambda_i+\lambda_k}\mathbf{a}_{\omega,i} - \sum_{i=1}^{n}\mathbf{tr}\left(\phi_i\tilde{D}^T\right)\mathbf{a}_{\omega,i}\ldots \\ &-\sum_{i=1}^{n}\sum_{k=1}^{r} \frac{\mathbf{tr}\left(\phi_i\hat{\phi}_k^T\right)}{\lambda_i+\hat{\lambda}_k}\left(\mathbf{a}_{\omega,i}+\hat{\mathbf{a}}_{\omega,k}\right) + \frac{1}{\pi}\omega\mathbf{tr}\left(\tilde{D}\tilde{D}^T\right)\ldots \\ &+\sum_{i=1}^{r}\sum_{k=1}^{r} \frac{\mathbf{tr}\left(\hat{\phi}_i\hat{\phi}_k^T\right)}{\hat{\lambda}_i+\hat{\lambda}_k}\hat{\mathbf{a}}_{\omega,i} + \sum_{i=1}^{r}\mathbf{tr}\left(\hat{\phi}_i\tilde{D}^T\right)\hat{\mathbf{a}}_{\omega,i}.\end{aligned} \tag{29}$$

REFERENCES

[1] M. R. Anderson, A. Emami-Naeni, and J.H. Vincent. Measures of merit for multivariable flight control. Technical report, Systems Control Technology Inc, Palo Alto, California, USA, 1991.

[2] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. Society for Industrial and Applied Mathematics, 2005.

[3] C.A. Beattie and S. Gugercin. A trust region method for optimal $\mathcal{H}_2$ model reduction. In *proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, pages 5370–5375, 2009.

[4] W. Gawronski and J. Juang. Model reduction in limited time and frequency intervals. *International Journal of Systems Science*, 21(2):349–376, 1990.

[5] S. Gugercin. An iterative SVD-Krylov based method for model reduction of large-scale dynamical systems. *Linear Algebra and its Applications*, 428(89):1964–1986, April 2007.

[6] S. Gugercin and A. C. Antoulas. A survey of model reduction by balanced truncation and some new results. *International Journal of Control*, 77(8):748–766, 2004.

[7] S. Gugercin, A. C. Antoulas, and C. Beattie. $\mathcal{H}_2$ model reduction for Large-Scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications*, 30(2):609–638, 2008.

[8] Howard E. Haber. The complex inverse trigonometric and hyperbolic functions, 2011. University of California.

[9] F. Leibfritz and W. Lipinski. Description of the benchmark examples in *COMPl$_e$ib* 1.0. Technical report, University of Trier, 2003.

[10] A. Masi, R. Wallin, A. Garulli, and A. Hansson. Robust finite-frequency $\mathcal{H}_2$ analysis. In *Proceedings of the 49th IEEE Conferance on Decision and Control*, pages 6876 –6881, December 2010.

[11] D. Petersson and J. Löfberg. Model reduction using a frequency-limited $\mathcal{H}_2$-cost. available as http://arxiv.org/abs/1212.1603, 2012.

[12] C. Poussot-Vassal. An Iterative SVD-Tangential Interpolation Method for Medium-Scale MIMO Systems Approximation with Application on Flexible Aircraft. In *Proceedings 50^{th} IEEE Conference on Decision and Control - European Control Conference (CDC-ECC'11)*, pages 7117–7122, Florida, 2011.

[13] C. Poussot-Vassal, T. Loquen, P. Vuillemin, O. Cantinaud, and J.P. Lacoste. Business jet large-scale model approximation and vibration control. In *Proceedings of the 11th IFAC International Workshop on Adaptation Learning in Control and Signal Processing (ALCOSP'13)*, pages 199–204, 2013.

[14] C. Poussot-Vassal and P. Vuillemin. Introduction to MORE: a MOdel REduction Toolbox. pages 776–781, Dubrovnik, Croatia, October 2012. Proceedings of the IEEE Multi Systems Conference (MSC CCA'12).

[15] J. Rommes and N. Martins. Efficient computation of multivariable transfer function dominant poles using subspace acceleration. 21:1471–1483, 2006.

[16] Yousef Saad. *Iterative Methods for Sparse Linear Systems, Second Edition*. Society for Industrial and Applied Mathematics, 2 edition, April 2003.

[17] L. Sorber, M. Van Barel, and L. De Lathauwer. Unconstrained optimization of real functions in complex variables. *SIAM J. Optim*, 22:879–898, 2012.

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

[19] P. Vuillemin, C. Poussot-Vassal, and D. Alazard. A spectral expression for the frequency-limited $\mathcal{H}_2$-norm. Available as http://arxiv.org/abs/1211.1858, 2012.

[20] P. Vuillemin, C. Poussot-Vassal, and D. Alazard. A frequency-limited $\mathcal{H}_2$ model approximation method with application to a medium-scale flexible aircraft. In *Proceedings of the AIAA CEAS Conference on Guidance Navigation and Control in aerospace (EuroGNC'13)*, 2013.