Optimal model reduction for non-rational functions

Mark R. Opmeer[1]

***Abstract*— We survey several algorithm for $H2$ optimal model reduction with a particular emphasis on the case of approximating irrational functions. Irrational transfer functions arise for systems modeled by partial differential equations or delay differential equations. We then compare the performance of these algorithms on two examples of irrational transfer functions: one arising from a heat equation and one arising from a beam equation.**

I. Introduction

Model reduction has been a topic of considerable interest in control theory for several decades now and is a topic in several control theory textbooks (e.g. [5], [16], [17]). Especially desirable is optimal approximation of the transfer function in the H^∞ norm. However, this is generally considered to be computationally unfeasible. Optimal approximation in other norms is computationally more feasible. Optimal Hankel norm approximation [4] and optimal H^2 norm approximation are two such cases. In this paper we consider optimal H^2 approximation of the transfer function (or equivalently: optimal L^2 approximation of the impulse response). It is usually assumed that the to-be-approximated transfer function is rational. It is interesting that several algorithms proposed in the literature for H^2 optimal approximation carry through unchanged for the case of irrational transfer functions. Irrational transfer functions arise for systems modeled by partial differential equations or delay differential equations. In this paper we first survey several algorithms for H^2 optimal approximation which have been considered in the literature. We then apply these algorithms to two examples of irrational transfer functions: one that arises from a heat equation and one that arises from a beam equation.

II. Problem statement

The problem considered in this paper is the following. Given $g \in L^2(0,\infty)$, find $g_r \in L^2(0,\infty)$ such that $\hat{g}_r$ is a rational function of degree at most r and such that

$$\|g - g_r\|^2_{L^2(0,\infty)} = \min \|g - h\|^2_{L^2(0,\infty)},$$

where the minimization is over $h \in L^2(0,\infty)$ with $\hat{h}$ a rational function of degree at most r.

III. A survey of L^2 optimal approximation algorithms

In this section we briefly survey some algorithms which have been proposed in the literature for the solution of the L^2 optimal approximation problem mentioned in Section II. The purpose is on the one hand to highlight similarities and differences and on the other hand to point out that these algorithms apply equally well to irrational functions (though they were generally introduced for rational functions only).

A. Nonlinear optimization algorithms

For simplicity we assume that the optimal approximation g_r is such that $\hat{g}_r$ has simple poles. Then we can search over the set of $h \in L^2(0,\infty)$ with $\hat{h}$ a rational function of degree at most r and with simple poles. This set can be parametrized as follows (the pole-residue parameterization):

$$h(t) = \sum_{j=1}^{r} d_j \mathrm{e}^{-\alpha_j t},$$

or equivalently in terms of the Laplace transform:

$$\hat{h}(s) = \sum_{j=1}^{r} \frac{d_j}{s + \alpha_j},$$

where $d_j \in \mathbb{C}$ and $\alpha_j \in \mathbb{C}$ with $\mathrm{Re}(\alpha_j) > 0$.

Substituting this parametrization in the error and expanding gives (this calculation is also performed in e.g. [8]):

$$\begin{aligned}
&\|g - h\|^2_{L^2(0,\infty)} \\
&\quad = \|g\|^2_{L^2(0,\infty)} - 2\mathrm{Re}\langle g, h\rangle_{L^2(0,\infty)} + \|h\|^2_{L^2(0,\infty)} \\
&\quad = \|g\|^2_{L^2(0,\infty)} - 2\mathrm{Re}\sum_{j=1}^{r} \overline{d_j}\hat{g}(\overline{\alpha}_j) + \sum_{j,\ell=1}^{r} \frac{d_j\overline{d_\ell}}{\overline{\alpha}_j + \alpha_\ell}.
\end{aligned}$$

Minimizing the error with respect to h is therefore equivalent to

$$\min -2\mathrm{Re}\sum_{j=1}^{r} \overline{d_j}\hat{g}(\overline{\alpha}_j) + \sum_{j,\ell=1}^{r} \frac{d_j\overline{d_\ell}}{\overline{\alpha}_j + \alpha_\ell},$$

over $d_j \in \mathbb{C}$ and $\alpha_j \in \mathbb{C}$ with $\mathrm{Re}(\alpha_j) > 0$. This is a nonlinear optimization problem in $2r$ complex variables (or equivalently: in $4r$ real variables). Specific algorithms can be devised which utilize the special strucure, or general purpose nonlinear optimization algorithms (such as those available through the matlab function `fmincon`) can be used.

B. Two-step iterative algorithms

Another class of algorithms for solving the L^2 optimal approximation problem mentioned in Section II can be described as follows.

a) Make an initial guess for the denominator of $\hat{g}_r$.
b) Solve the optimization problem from Section II over all $h \in L^2(0,\infty)$ with $\hat{h}$ a rational function of degree at most r *and with as denominator the initial guess*. Since

[1]Mark Opmeer is with the Department of Mathematical Sciences, University of Bath, BA2 7AY Bath, United Kingdom `m.opmeer@bath.ac.uk`

the set over which is minimized is now a linear space, the minimizer is easily found: it is the orthogonal projection of the given function g onto the subspace determined by the initial guess.

c) Solve the optimization problem from Section II over all $h \in L^2(0,\infty)$ with $\hat{h}$ a rational function of degree at most r *and with as numerator the numerator obtained in the previous step*. This is a nonlinear optimization problem and is typically solved by a gradient based iterative method.

d) Repeat step b) but with the initial guess for the denominator replaced by the denominator obtain in step c).

Since step c) is part of an iterative process, often in the gradient based iterative method for the nonlinear optimization algorithm, only one step is performed. The details in which the algorithms proposed in the literature differ is in the parametrization used for h. Spanos et al. [14] use the numerator-denominator parametrization of h. Hwang and Hwang [7] use numerator coefficients, but use Routh parameters of the denominator instead of denominator coefficients.

A reason for doing the above two-step optimization process rather than solving the original nonlinear optimization problem directly (as in Section III-A) is that in the original problem there are $2r$ complex variables whereas in the subproblem in step c) there are only r complex variables.

Note that we may alternatively use the pole-residue parametrization from Section III-A and in step c) replace "and with as numerator the numerator obtained in the previous step" by "and with as residues the residues obtained in the previous step". In more detail:

a) Make an initial guess for the poles (i.e for α_j from Section III-A).

b) Solve the optimization problem from Section III-A for d_j *with these fixed* α_j. Since the set over which is minimized is now a linear space, the minimizer is easily found: it is the orthogonal projection of the given function g onto the subspace determined by the initial guess. We comment further on this in Section III-D.

c) Solve the optimization problem from Section III-A for α_j for these fixed d_j.

d) Repeat step b) but with the initial guess for the poles replaced by the poles obtain in step c).

C. Algorithms based on optimality conditions

First order optimality conditions for the L^2 optimal approximation problem mentioned in Section II can be obtained by "differentiating the cost function with respect to h". More precisely, the set of $h \in L^2(0,\infty)$ with $\hat{h}$ a rational function of degree at most r is parametrized by $2r$ complex parameters and the derivative of the cost function with respect to the corresponding $4r$ real variables is set to zero. As in Section III-B, different parametrizations are possible. The most commonly used one is the pole-residue parameterization. The obtained conditions are:

$$\begin{aligned} \hat{g}(-\overline{p}_j) &= \hat{g}_r(-\overline{p}_j), \\ \hat{g}'(-\overline{p}_j) &= \hat{g}_r'(-\overline{p}_j), \end{aligned} \tag{1}$$

where p_j $(j = 1, \ldots, r)$ are the poles of $\hat{g}_r$. The above idea goes back at least to Meier and Luenberger [12]. A review of alternative formulations of these optimality conditions can be found in [6].

The nonlinear equations (1) are typically not algebraically solvable. Already in [12] use of Newton's method was suggested to solve these equations numerically. An iterative method somewhat similar in spirit to the method discussed in Section III-B for solving (1) is as follows.

i) Make an initial guess for the poles p_j of $\hat{g}_r$.

ii) Solve the equations (1) for the numerator and denominator coefficients of $\hat{g}_r$.

iii) Calculate the poles of the obtained $\hat{g}_r$ and repeat step ii) with these values for the p_j.

This iterative method based on first order optimality conditions goes back at least to Lepschy et al. [10] and Lucas [11]. Local convergence is considered in Krajewski et al. [8] and Flagg et al. [3]. An efficient algorithm (for the case where g is rational) based on state space formulas and Krylov subspaces is given in Gugercin et al. [6]. An implementation using only evaluations of the transfer function and its derivatives is described in [1] (as in this paper, non-rational functions are considered in [1]).

D. The method from Mi et al.

In this section we give an exposition of the results from Mi et al. [13]. Those results are strongly related to the two-step iterative algorithms reviewed in Section III-B. In fact, what is done in [13] is that the optimal numerator is calculated for an arbitrary denominator (in terms of the zeros of the denominator). The nonlinear optimization problem then only has to be solved once (but for a more complicated cost function).

1) Optimal model reduction with a fixed denominator: In this section we consider the case where the minimization is over all $h \in L^2(0,\infty)$ with $\hat{h}$ a rational function of degree at most r *with a fixed denominator*. The importance of this assumption of a fixed denominator is that the set over which is minimized is now a linear space. The minimizer is then easily found: it is the orthogonal projection of the given function g onto the given subspace. To compute this orthogonal projection it is convenient to utilize an orthonormal basis for the given subspace. We now recall the construction of this basis (which essentially goes back to Takenaka and Malmquist; see Walsh [15]).

Let the fixed denominator be given by

$$\prod_{k=1}^{r} s + \alpha_k,$$

with $\mathrm{Re}(\alpha_k) > 0$. Define the functions ϕ_j and ψ_j by

$$\begin{aligned} \phi_1 &= t \mapsto \mathrm{e}^{-\alpha_1 t}, \\ \psi_1 &= \sqrt{2\mathrm{Re}(\alpha_1)} \cdot \phi_1, \\ \phi_j &= \phi_{j-1} - (\alpha_j + \overline{\alpha}_{j-1}) \cdot (\mathrm{e}^{-\alpha_j \cdot} * \phi_{j-1}), \\ \psi_j &= \sqrt{2\mathrm{Re}(\alpha_j)} \cdot \phi_j, \end{aligned}$$

where $*$ denotes the convolution product, i.e. $(g*h)(t) = \int_0^t g(t-\tau)h(\tau)\,d\tau$. Then $(\psi_k)_{k=1}^r$ is an orthonormal set in $L^2(0,\infty)$. Taking Laplace transforms we obtain that $(\widehat{\psi}_k)_{k=1}^r$ is an orthonormal set in the Hardy space $H^2(\mathbb{C}_0^+)$ and that

$$\begin{aligned}
\widehat{\phi}_1(s) &= \tfrac{1}{s+\alpha_1},\\
\widehat{\psi}_1(s) &= \sqrt{2\mathrm{Re}(\alpha_1)}\cdot\widehat{\phi}_1(s),\\
\widehat{\phi}_j(s) &= \widehat{\phi}_{j-1}(s) - (\alpha_j+\overline{\alpha}_{j-1})\cdot\tfrac{1}{s+\alpha_j}\cdot\widehat{\phi}_{j-1}(s),\\
\widehat{\psi}_j(s) &= \sqrt{2\mathrm{Re}(\alpha_j)}\cdot\widehat{\phi}_j(s).
\end{aligned}$$

We obtain by induction that

$$\widehat{\psi}_j(s) = \frac{\sqrt{2\mathrm{Re}(\alpha_j)}}{s+\alpha_j}\cdot\prod_{\ell=1}^{j-1}\frac{s-\overline{\alpha}_\ell}{s+\alpha_\ell}.$$

Note that $(\psi_k)_{k=1}^r$ is an orthonormal basis for our given subspace. Hence the optimal approximation with fixed denominator is

$$g_r = \sum_{k=1}^r \langle g,\psi_k\rangle\psi_k.$$

The involved inner-products can be explicitly calculated. For simplicity of exposition we will assume that $\alpha_i \neq \alpha_j$ for $i\neq j$. We have

$$\langle g,\psi_k\rangle = \frac{1}{i2\pi}\int_{i\mathbb{R}}\widehat{g}(s)\overline{\widehat{\psi}_k(s)}\,ds.$$

Defining

$$\widehat{\Psi}_k(s) := \frac{\sqrt{2\mathrm{Re}(\alpha_k)}}{-s+\overline{\alpha}_k}\cdot\prod_{\ell=1}^{k-1}\frac{-s-\alpha_\ell}{-s+\overline{\alpha}_\ell},$$

we have

$$\langle g,\psi_k\rangle = \frac{1}{i2\pi}\int_{i\mathbb{R}}\widehat{g}(s)\widehat{\Psi}_k(s)\,ds.$$

By the Residue Theorem we then have

$$\langle g,\psi_k\rangle = -\sum_{j=1}^k \widehat{g}(\overline{\alpha}_j)\ \mathrm{Res}[\widehat{\Psi}_k,\overline{\alpha}_j].$$

The residue of $\widehat{\Psi}_k$ at $\overline{\alpha}_j$ can be calculated as follows (at this point it is used that $\overline{\alpha}_j$ is a simple pole of $\widehat{\Psi}_k$): for $j<k$ we have

$$\begin{aligned}
\mathrm{Res}[\widehat{\Psi}_k,\overline{\alpha}_j] &= \lim_{s\to\overline{\alpha}_j}(s-\overline{\alpha}_j)\widehat{\Psi}_k(s)\\
&= \frac{\sqrt{2\mathrm{Re}(\alpha_k)}}{\overline{\alpha}_k-\overline{\alpha}_j}\cdot(\overline{\alpha}_j+\alpha_j)\cdot\prod_{\ell=1,\ell\neq j}^{k-1}\frac{\overline{\alpha}_j+\alpha_\ell}{\overline{\alpha}_j-\overline{\alpha}_\ell},
\end{aligned}$$

and for $j=k$ we have

$$\begin{aligned}
\mathrm{Res}[\widehat{\Psi}_k,\overline{\alpha}_k] &= \lim_{s\to\overline{\alpha}_k}(s-\overline{\alpha}_k)\widehat{\Psi}_k(s)\\
&= -\sqrt{2\mathrm{Re}(\alpha_k)}\cdot\prod_{\ell=1}^{k-1}\frac{\overline{\alpha}_k+\alpha_\ell}{\overline{\alpha}_k-\overline{\alpha}_\ell}.
\end{aligned}$$

Combining the above, we obtain a slightly cumbersome but very explicit formula for the optimal approximation in the case of a fixed denominator:

$$g_r = \sum_{k=1}^r c_k\psi_k,$$

where

$$\begin{aligned}
c_k := &-\sum_{j=1}^{k-1}\widehat{g}(\overline{\alpha}_j)\ \frac{\sqrt{2\mathrm{Re}(\alpha_k)}}{\overline{\alpha}_k-\overline{\alpha}_j}\cdot(\overline{\alpha}_j+\alpha_j)\cdot\prod_{\ell=1,\ell\neq j}^{k-1}\frac{\overline{\alpha}_j+\alpha_\ell}{\overline{\alpha}_j-\overline{\alpha}_\ell},\\
&+\widehat{g}(\overline{\alpha}_k)\ \sqrt{2\mathrm{Re}(\alpha_k)}\cdot\prod_{\ell=1}^{k-1}\frac{\overline{\alpha}_k+\alpha_\ell}{\overline{\alpha}_k-\overline{\alpha}_\ell}. \qquad (2)
\end{aligned}$$

It is possible to obtain an alternative expression for g_k as a sum of exponentials. We have

$$\psi_k(t) = \sum_{j=1}^k \mathrm{Res}[\widehat{\psi}_k,-\alpha_j]\mathrm{e}^{-\alpha_j t},$$

which gives

$$g_r(t) = \sum_{k=1}^r c_k\sum_{j=1}^k \mathrm{Res}[\widehat{\psi}_k,-\alpha_j]\mathrm{e}^{-\alpha_j t},$$

so that (changing the order of summation)

$$g_r(t) = \sum_{j=1}^r\sum_{k=j}^r c_k\mathrm{Res}[\widehat{\psi}_k,-\alpha_j]\mathrm{e}^{-\alpha_j t},$$

i.e.

$$g_r(t) = \sum_{j=1}^r d_j\mathrm{e}^{-\alpha_j t},$$

with

$$d_j := \sum_{k=j}^r c_k\mathrm{Res}[\widehat{\psi}_k,-\alpha_j].$$

The expression $\mathrm{Res}[\widehat{\psi}_k,-\alpha_j]$ can be calculated similarly as $\mathrm{Res}[\widehat{\Psi}_k,\overline{\alpha}_j]$ was calculated above: for $j<k$

$$\mathrm{Res}[\hat{\psi}_k,-\alpha_j] = \frac{\sqrt{2\mathrm{Re}(\alpha_k)}}{\alpha_j-\alpha_k}\cdot(\alpha_j+\overline{\alpha}_j)\cdot\prod_{\ell=1,\ell\neq j}^{k-1}\frac{\alpha_j+\overline{\alpha}_\ell}{\alpha_j-\alpha_\ell}.$$

and for $j=k$

$$\mathrm{Res}[\hat{\psi}_k,-\alpha_k] = \sqrt{2\mathrm{Re}(\alpha_k)}\cdot\prod_{\ell=1}^{k-1}\frac{\alpha_k+\overline{\alpha}_\ell}{\alpha_k-\alpha_\ell}.$$

Note that

$$\mathrm{Res}[\hat{\psi}_k,-\alpha_j] = -\overline{\mathrm{Res}[\widehat{\Psi}_k,\overline{\alpha}_j]}.$$

The error $\|g-g_r\|^2$ can be calculated as follows. We first recall a general Hilbert space result. Let $\mathscr{H}$ be a Hilbert space, let $g\in\mathscr{H}$ and let $(\psi_k)_{k=1}^r$ be an orthonormal set in $\mathscr{H}$. Then

$$\begin{aligned}
\left\|g-\sum_{k=1}^r\langle g,\psi_k\rangle\psi_k\right\|^2 &= \|g\|^2 - \left\|\sum_{k=1}^r\langle g,\psi_k\rangle\psi_k\right\|^2\\
&= \|g\|^2 - \sum_{k=1}^r|\langle g,\psi_k\rangle|^2.
\end{aligned}$$

Applied to the problem at hand this gives

$$\|g - g_r\|^2 = \|g\|^2 - \sum_{k=1}^{r} |c_k|^2,$$

where c_k is given by (2).

2) Optimal approximation with a variable denominator: The solution of the general optimal approximation problem formulated in Section II can be obtained by minimizing the solution obtained in Section III-D.1 over all possible denominators. Therefore, the general optimal approximation problem is reduced to

$$\min \|g\|^2 - \sum_{k=1}^{r} |c_k|^2,$$

where c_k is given by (2) and the minimum is taken over all $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$ with positive real parts. This is equivalent to

$$\min - \sum_{k=1}^{r} |c_k|^2,$$

over all $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$ with positive real parts. This is a nonlinear programming problem which can be solved by various standard techniques from that area.

In the numerical examples presented in Section IV, we use the function `fmincon` from matlab (as is done in [13]). This function `fmincon` optimizes over real variables and therefore the α_i need to be split into their real and imaginary parts and the optimization takes place over $2r$ real variables. In principle `fmincon` only needs to be able to evaluate the objective function; by (2) this means that evaluation of $\hat{g}$ needs to be possible. Optionally, the gradient and Hessian can be provided to `fmincon`. Using (2) it is possible to explicitly calculate these; they will involve evaluations of $\hat{g}'$ and $\hat{g}''$. In the numerical experiments reported in Section IV, we have not used these gradient and Hessian options.

We note that it is more natural to use optimization algorithms based directly on the complex variables $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$. It is indeed possible to use gradient based iterative methods directly for complex variables (see e.g. the exposition in [9]). However, we have not implemented these and instead simply rely on the matlab function `fmincon`.

IV. Numerical examples

We consider several numerical examples. The transfer functions are non-rational and arise from PDEs. We have taken these examples from Curtain and Morris [2].

We have implemented the method described in Section III-A, the method described in Section III-D and the method described in Section III-C. In the first two cases we use the matlab function `fmincon` (with the three optional algorithms `interior-point`, `sqp`, `active-set`). In the third case we use the implementation TF-IRKA as described in [1]. As stopping criterion in TF-IRKA we use that the relative change in the cost function from one iteration to the next is smaller than some prescribed tolerance.

A. A heat equation

We consider the following heat equation (taken from Curtain and Morris [2, Section 1.3]):

$$\begin{aligned} &\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial \xi^2}, \qquad t > 0, \quad \xi \in (0,1), \\ &w(0,t) = 0, \quad \frac{\partial w}{\partial \xi}(1,t) = u(t), \\ &y(t) = w(\xi_0, t). \end{aligned}$$

The transfer function (computed in [2, Section 1.3]) is:

$$\hat{g}(s) = \frac{\sinh(\sqrt{s}\xi_0)}{\sqrt{s}\cosh(\sqrt{s})}.$$

In the computations below we take $\xi_0 = 1/3$.

Utilizing any of the mentioned methods, it quickly becomes clear that $r = 2$ is the "correct" order of the approximant. The approximation error for $r = 1$ is significantly larger than for $r = 2$, but for $r \geq 3$ the approximation error isn't significantly smaller than for $r = 2$.

When using `fmincon` in this example, we force the parameters to be real. Allowing them to be complex results in the output of `fmincon` having very small imaginary part.

As initial guess for the poles of the approximant we use the poles of $\hat{g}$ which are closest to zero. The poles of $\hat{g}$ are given by $\lambda_k = (k\pi + \pi/2)^2$ for $k = 0, 1, \ldots$ (see [2, Section 1.3]), so that the initial guess for the poles for $r = 2$ is $(\pi/2)^2 \approx 2.47$ and $(3\pi/2)^2 \approx 22.2$. The method from Section III-A also needs an initial guess for the residues. For this we use the optimal residues for the initial guess of poles (as calculated in Section III-D). We see from the results presented in Table I that all considered methods perform equally well (but using the `fmincon` option `active-set` is slightly worse than the other options).

Method	$\|g\|^2 - \|g - g_r\|^2$
III-A `interior-point`	0.1217
III-A `sqp`	0.1217
III-A `active-set`	0.1207
III-D `interior-point`	0.1217
III-D `sqp`	0.1217
III-D `active-set`	0.1207
III-C `TF-IRKA`	0.1217

TABLE I

Approximations for the heat equation

B. A beam equation

We consider the following beam equation (taken from Curtain and Morris [2, Section 4.2]):

$$\frac{\partial^2 w}{\partial t^2} + \beta \frac{\partial^5 w}{\partial \xi^4 \partial t} + \alpha \frac{\partial^4 w}{\partial \xi^4} = 0, \qquad t > 0, \quad \xi \in (0, L),$$
$$w(0,t) = 0, \quad \frac{\partial w}{\partial \xi}(0,t) = 0,$$
$$\beta \frac{\partial^3 w}{\partial \xi^2 \partial t}(L,t) + \alpha \frac{\partial^2 w}{\partial \xi^2}(L,t) = 0,$$
$$-\beta \frac{\partial^4 w}{\partial \xi^2 \partial t}(L,t) - \alpha \frac{\partial^3 w}{\partial \xi^2}(L,t) = u(t),$$
$$y(t) = \frac{\partial w}{\partial t}(L,t).$$

The transfer function (computed in [2, Section 4.2]) is:

$$\hat{g}(s) = \frac{sN(s)}{m^3(s)(\alpha + \beta s)D(s)},$$

where

$$m(s) = \left(\frac{-s^2}{\alpha + \beta s}\right)^{1/4},$$
$$N(s) = \cosh(Lm(s))\sin(Lm(s)) - \sinh(Lm(s))\cos(Lm(s)),$$
$$D(s) = 1 + \cosh(Lm(s))\cos(Lm(s)).$$

In the computations we take $L = 2$, $\alpha = 1.129$, $\beta = 3.89 \times 10^{-4}$. A Bode plot is given in Figure 1. This transfer function is considerably harder to approximate than the one from the heat equation given in Section IV-A.

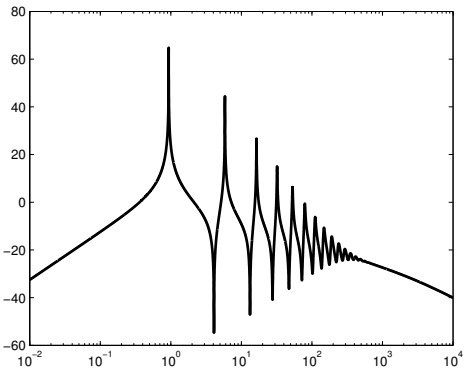

Fig. 1. Bode plot for the beam equation

The poles of the transfer function are given in terms of the solution of a transcendental equation. As shown in [2, Section 4.2], the solutions of this transcendental equation are asymptotically given by $\alpha_k \approx \frac{2k-1}{2L}\pi$ ($k = 1, 2, \ldots$). We use the matlab function `fsolve` with $\frac{2k-1}{2L}\pi$ as initial guess to solve this transcendental equation and obtain the poles. As shown in [2, Section 4.2], the residues can be calculated in terms of the poles. Using this information, we can form a modal truncation. This modal truncation is then used as initial guess for the algorithms described in Sections III-A, III-D (with the default method `interior point` used in `fmincon`) and III-C. The results are summarized in Table II; the same values are obtained for modal truncation and for the methods described in Sections III-A, III-C and III-D. We also consider the case where we use $\frac{2k-1}{2L}\pi$ instead of α_k in the formulas for the modal truncation. Using this "approximate modal truncation" as initial guess gives the results summarized in Table III. It can be seen that the method from Section III-A finds the minimum for $r = 2$, but fails to find it for $r = 4$ and $r = 8$. The method from Section III-C find the minimum in all three cases. The method from Section III-D finds the minimum for $r = 2$, fails to find it for $r = 4$ and comes close to the minimum for $r = 8$. We also consider the case $r = 2$ with initial guess for the poles corresponding to the approximation $\frac{3}{2L}$ of α_2 rather than to the approximation of α_1. As expected, the methods from Sections III-C and III-D converge to a local minimum corresponding to the "second peak" in the Bode plot; see Table IV and Figure 2. The method from Section III-A does even worse. These results clearly indicate that for a transfer function where the Bode plot has "peaks", the initial guess has to reasonably accurately capture the location of these peaks; otherwise the considered algorithms will converge to a local minimum.

r	$\|g\|^2 - \|g - g_r\|^2$
2	6654
4	6824
8	6851

TABLE II

APPROXIMATIONS FOR THE BEAM EQUATION: MODAL TRUNCATION AS INITIAL GUESS

r	method	$\|g\|^2 - \|g - g_r\|^2$
2	III-A	6654
	III-C	6654
	III-D	6654
4	III-A	6672
	III-C	6824
	III-D	6657
8	III-A	310
	III-C	6851
	III-D	6847

TABLE III

APPROXIMATIONS FOR THE BEAM EQUATION: APPROXIMATE MODAL TRUNCATION AS INITIAL GUESS

REFERENCES

[1] C. Beattie and S. Gugercin. Realization-independent H2-approximation. In *IEEE 51st Conference on Decision and Control (CDC)*, pages 4953–4958, Dec 2012.

method	$\|g\|^2 - \|g - g_r\|^2$
III-A	18
III-C	169
III-D	169

TABLE IV

APPROXIMATIONS FOR THE BEAM EQUATION: $r = 2$ AND INITIAL GUESS CLOSE TO THE "SECOND PEAK"

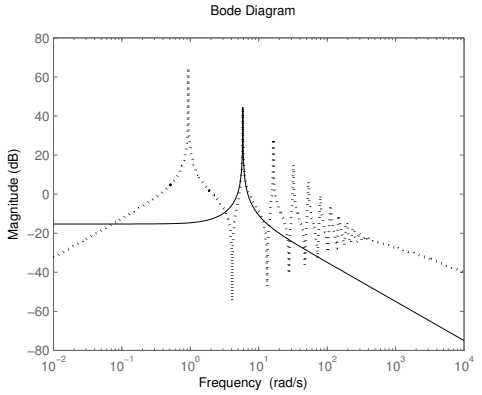

Fig. 2. Bode plot for the beam equation and its $r = 2$ approximations III-C and III-D starting from an initial guess close to the "second peak"

[2] Ruth Curtain and Kirsten Morris. Transfer functions of distributed parameter systems: a tutorial. *Automatica J. IFAC*, 45(5):1101–1116, 2009.

[3] Garret Flagg, Christopher Beattie, and Serkan Gugercin. Convergence of the iterative rational Krylov algorithm. *Systems Control Lett.*, 61(6):688–691, 2012.

[4] Keith Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds. *Internat. J. Control*, 39(6):1115–1193, 1984.

[5] Michael Green and David J.N. Limebeer. *Linear robust control*. Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[6] S. Gugercin, A. C. Antoulas, and C. Beattie. H_2 model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30(2):609–638, 2008.

[7] Chyi Hwang and Jyh-Haur Hwang. A new two-step iterative method for optimal reduction of linear SISO systems. *J. Franklin Inst. B*, 333(5):631–645, 1996.

[8] W. Krajewski, A. Lepschy, M. Redivo Zaglia, and U. Viaro. A program for solving the L_2 reduced-order model problem with fixed denominator degree. *Numer. Algorithms*, 9(3-4):355–377, 1995.

[9] Ken Kreutz-Delgado. The complex gradient operator and the $\mathbb{CR}$-calculus.

[10] A. Lepschy, G.A. Mian, G. Pinato, and U. Viaro. Rational l2 approximation: a non-gradient algorithm. In *IEEE 30th Conference on Decision and Control (CDC)*, pages 2321–2323 vol.3, Dec 1991.

[11] T. N. Lucas. Optimal model reduction by multipoint Padé approximation. *J. Franklin Inst.*, 330(1):79–93, 1993.

[12] L. Meier and D.G. Luenberger. Approximation of linear constant systems. *IEEE Trans. Automat. Control*, 12(5):585–588, 1967.

[13] Wen Mi, Tao Qian, and Feng Wan. A fast adaptive model reduction method based on Takenaka-Malmquist systems. *Systems Control Lett.*, 61(1):223–230, 2012.

[14] J. T. Spanos, M. H. Milman, and D. L. Mingori. A new algorithm for L_2 optimal model reduction. *Automatica J. IFAC*, 28(5):897–909, 1992.

[15] J. L. Walsh. *Interpolation and approximation by rational functions in the complex domain*. Fourth edition. American Mathematical Society Colloquium Publications, Vol. XX. American Mathematical Society, Providence, R.I., 1965.

[16] Kemin Zhou and John C. Doyle. *Essentials of robust control*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1999.

[17] Kemin Zhou, John C. Doyle, and Keith Glover. *Robust and optimal control*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.