

A Rational Krylov Iteration for Optimal \mathcal{H}_2 Model Reduction

Serkan Gugercin, Athanasios C. Antoulas and Christopher A. Beattie

In this paper, we address the optimal \mathcal{H}_2 approximation of a stable, single-input single-output large-scale dynamical system. The problem we consider is as follows: Given an n^{th} order linear dynamical system $G(s) = C(sI - A)^{-1}B$ where $A \in \mathbb{R}^{n \times n}$, and $B, C^T \in \mathbb{R}^n$, find a stable r^{th} order reduced system $G_r(s) = C_r(sI_r - A_r)^{-1}B_r$ with $r < n$, such that $G_r(s)$ minimizes the \mathcal{H}_2 error, i.e.

$$G_r(s) = \arg \min_{\deg(\hat{G})=r} \|G(s) - \hat{G}(s)\|_{\mathcal{H}_2}. \quad (1)$$

where $\|G\|_{\mathcal{H}_2} := \left(\int_{-\infty}^{\infty} |G(jw)|^2 dw \right)^{1/2}$.

In the sequel, we will construct the reduced order models $G_r(s)$ through *Krylov* projection methods. Toward this end, we construct matrices $V \in \mathbb{R}^{n \times r}$ and $Z \in \mathbb{R}^{n \times r}$ that span certain Krylov subspaces with the property that $Z^T V = I_r$. The reduced order model $G_r(s)$ will then be obtained as

$$A_r = Z^T A V, \quad B_r = Z^T B, \quad \text{and} \quad C_r = C V. \quad (2)$$

The corresponding oblique projection is given by VZ^T .

Many researchers have worked on the problem (1); see [18], [16], [7], [5], [3], [17], [4] and references therein. Since obtaining a global minimum is a hard task, the goal is to generate a reduced-order model that satisfies the first-order conditions for (1). However, these methods require solving a (sequence) of large-scale Lyapunov equations and hence dense matrix operations including inversion, which rapidly become intractable as the dimension increases in large-scale settings. Indeed some of these methods are unsuitable even for medium scale problems. Here, we propose an iterative rational Krylov algorithm which efficiently seeks a minimizer to the problem (1). The method is based on the computationally proven approaches utilizing Krylov subspaces. The proposed method is suitable for large-scale settings where the order of the system, n , can grow to the order of many thousands of state variables.

Our starting point is the interpolation based first-order conditions for the optimal \mathcal{H}_2 approximation obtained by Meier and Luenberger [5]:

Theorem 1: Let $G_r(s)$ solve the optimal \mathcal{H}_2 problem and let $\hat{\lambda}_i$ denote the eigenvalues of A_r , i.e. $\hat{\lambda}_i$ are the Ritz values. For simplicity, assume that $\hat{\lambda}_i$ has multiplicity

one. Then, the first-order necessary conditions for \mathcal{H}_2 optimality are

$$\left. \frac{d^k}{ds^k} G(s) \right|_{s=-\hat{\lambda}_i} = \left. \frac{d^k}{ds^k} G_r(s) \right|_{s=-\hat{\lambda}_i}, \quad k = 0, 1. \quad (3)$$

Theorem 1 states that the reduced model has to *interpolate* $G(s)$ and its first derivative at the mirror images of the Ritz values. For equivalence between the interpolation-based framework of the optimal \mathcal{H}_2 problem [5] and Lyapunov-based framework [17], see Gugercin *et al.* [12].

The following result by Grimme [10] for interpolation-based (Krylov-based) model reduction will be the main tool for the proposed method.

Theorem 2: [10] Given $G(s) = C(sI - A)^{-1}B$ and r interpolation points $\{\sigma_i\}_{i=1}^r$, let $V \in \mathbb{R}^{n \times r}$ and $Z \in \mathbb{R}^{n \times r}$ be obtained as follows:

$$\begin{aligned} \text{Im}(V) &= \text{Span}\{(\sigma_1 I - A)^{-1}B, \dots, (\sigma_r I - A)^{-1}B\} \\ \text{Im}(Z) &= \text{Span}\{(\sigma_1 I - A)^{-T}C^T, \dots, (\sigma_r I - A)^{-T}C^T\} \end{aligned}$$

with $Z^T V = I_r$. Then, the reduced model $G_r(s) = C_r(sI_r - A_r)^{-1}B_r$ obtained as in (2) interpolates $G(s)$ and its first derivative at $\{\sigma_i\}_{i=1}^r$.

Proposed Algorithm: We propose a numerical algorithm which efficiently produces a reduced order model $G_r(s)$ satisfying the interpolation-based first-order necessary conditions (3). Our approach exploits the connection between the Krylov-based reduction and interpolation. Since the interpolation points in (3) depend on the final reduced model and are not known *a priori*, we use rational Krylov steps to iteratively correct the reduced-order model $G_r(s)$ so that the next (corrected) reduced-order model interpolates the full-order model at mirrored Ritz values $-\hat{\lambda}_i(A_r)$ from the previous reduced-order model. This continues until Ritz values from consecutive reduced-order models stagnate. Below, we give a sketch of the algorithm:

Algorithm 1: An Iterative Rational Krylov Algorithm (IRKA):

- 1) Make an initial shift selection σ_i for $i = 1, \dots, r$
- 2) $Z = [(\sigma_1 I - A^T)^{-1}C^T, \dots, (\sigma_r I - A^T)^{-1}C^T]$
- 3) $V = [(\sigma_1 I - A)^{-1}B, \dots, (\sigma_r I - A)^{-1}B]$
- 4) $Z = Z(Z^T V)^{-T}$ (to make $Z^T V = I_r$)
- 5) while (not converged)
 - a) $A_r = Z^T A V$,
 - b) $\sigma_i \leftarrow -\hat{\lambda}_i(A_r)$ for $i = 1, \dots, r$
 - c) $Z = [(\sigma_1 I - A)^{-T}C^T, \dots, (\sigma_r I - A)^{-T}C^T]$
 - d) $V = [(\sigma_1 I - A)^{-1}B, \dots, (\sigma_r I - A)^{-1}B]$
 - e) $Z = Z(Z^T V)^{-T}$ (to make $Z^T V = I_r$)
- 6) $A_r = Z^T A V$, $B_r = Z^T B$, $C_r = C V$

S. Gugercin and C.A. Beattie are with the Department of Mathematics, Virginia Tech., Blacksburg, VA, USA, {gugercin,beattie}@math.vt.edu. The work of these author was supported in part by the NSF through Grants DMS-050597 and DMS-0513542.

A.C. Antoulas is with the Department of Electrical and Computer Eng., Rice University, Houston, TX, USA, aca@ece.rice.edu. The work of this author was supported in part by the NSF through Grants CCR-0306503 and ACI-0325081.

Upon convergence, desired interpolation conditions (3), i.e. the first-order conditions, will be satisfied. It should be noted that solution of the optimal \mathcal{H}_2 reduction problem is obtained via Krylov projection methods only and its computation is suitable in large-scale setting.

We have implemented the above algorithm for many different large-scale systems. In most of our numerical examples, the algorithm worked very efficiently and converged after a small number of steps, resulted in stable reduced systems. For some case problem where the global optimal is known, Algorithm 1 has converged to this global optimal. We have also developed a Newton formulation of this algorithm, see [12]. The Newton formulation seemed to result in faster convergence and prevent rare convergence failures of Algorithm 1.

Corollary 1: Let $G_r(s)$ be the reduced model resulting from Algorithm 1. Then, $G_r(s)$ is the optimal approximation of $G(s)$ with respect to the \mathcal{H}_2 norm among all reduced order systems having the same reduced system poles as $G_r(s)$. Therefore, Algorithm 1 generates a reduced model $G_r(s)$ which is the optimal solution for a restricted \mathcal{H}_2 problem.

CD Player Example: The original model describes the dynamics between a lens actuator and the radial arm position in a portable CD player. The model has 120 states, i.e., $n=120$, with a single input and a single output. As illustrated in [6], even though the Krylov-based methods resulted in good local behavior, they are observed to yield large \mathcal{H}_∞ and \mathcal{H}_2 error compared to balanced truncation [13], [14] which is well known to yield small \mathcal{H}_∞ and \mathcal{H}_2 error norms, see [6], [11].

In this example, we compare the performance of the proposed method **IRKA**, Algorithm 1, with that of balanced truncation. We reduce the order to r as r varies from 2 to 40; and for each r value, we compare the \mathcal{H}_2 error norms due to balanced truncation and due to Algorithm 1. For the proposed algorithm, the initial shifts are *randomly* selected with real parts in the interval $[10^{-1}, 10^3]$ and the imaginary parts in the interval $[1, 10^5]$. The results showing the relative \mathcal{H}_2 error for each r are depicted in Figure 1. The figure reveals that the proposed method with initial random shift selection outperforms balanced truncation for almost all the r values except $r = 2, 24, 36$. However, even for these three r values, the resulting \mathcal{H}_2 error is not far away from the one due to balanced truncation. For the range $r = [12, 22]$, **IRKA** clearly outperforms the balanced truncation. We would like to emphasize that these results were obtained by a *random* shift selection and staying in the numerically efficient Krylov projection framework *without* requiring any solutions to large-scale Lyapunov

equations. This is the main difference of our algorithm from the existing methods and this makes the proposed algorithm numerically efficient in large-scale settings.

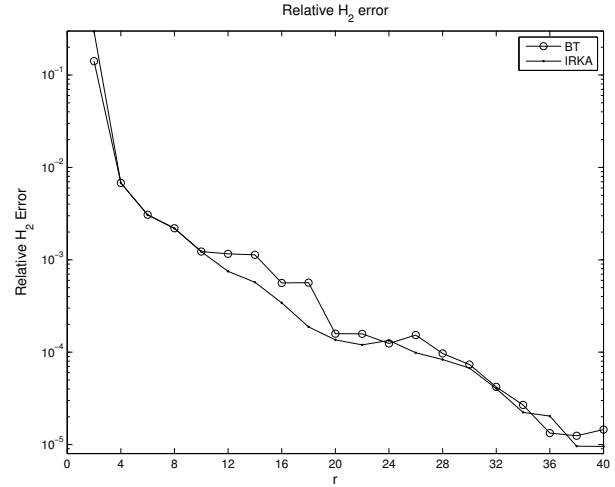


Fig. 1. Relative \mathcal{H}_2 norm of the error system vs r

To examine the convergence behavior, we reduce the order to $r = 8$ and $r = 10$ using Algorithm 1 and at each step of the iteration, we compute the \mathcal{H}_2 error due to the current estimate and plot this error vs iteration index. The results are shown in Figure 2. The figure illustrates at each step of the iteration, the \mathcal{H}_2 norm of the error is reduced and the algorithm converges after 4 steps.

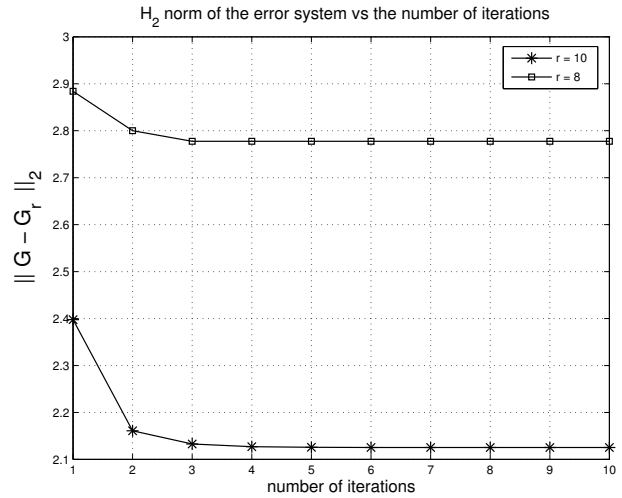


Fig. 2. \mathcal{H}_2 norm of the error system vs the number of iterations

A Semi-discretized Heat Transfer Problem: This problem arises during a cooling process in a rolling mill when

different steps in the production process require different temperatures of the raw material. The problem is modeled as boundary control of a two dimensional heat equation. A finite element discretization results in a system of the form

$$E\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t).$$

with state-dimension $n = 20209$, i.e., $A, E \in \mathbb{R}^{20209 \times 20209}$, $B \in \mathbb{R}^{20209 \times 7}$, $C \in \mathbb{R}^{6 \times 20209}$. Note that in this case $E \neq I_n$, but the algorithm works fine with the obvious modifications. For details regarding the modeling, discretization, optimal control design, and model reduction for this example, see [15], [1], [2]. We consider the full-order SISO system that associates the sixth input of this system with the second output. We apply our algorithm and reduced the order to $r = 6$. Amplitude Bode plots of $G(s)$ and $G_r(s)$ are shown in Figure 3.

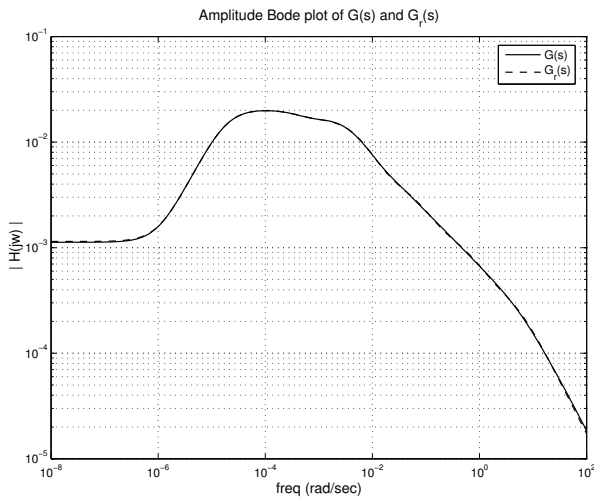


Fig. 3. Amplitude Bode plots of $G(s)$ and $G_r(s)$

The output response of $G_r(s)$ is virtually indistinguishable from $G(s)$ in the frequency range considered. **IRKA** converged in 7 iteration steps, although some interpolation points converged in the first 2-3 steps. The relative \mathcal{H}_∞ error obtained with our sixth order system was 7.85×10^{-3} . Note that in order to apply Lyapunov-based methods, e.g. [3], [16], for this example, one would need to solve 2 generalized Lyapunov equations (since $E \neq I_n$) of order 20209 at each step of the iteration.

REFERENCES

- [1] P. Benner, *Solving Large-Scale Control Problems*, IEEE Control Systems Magazine, Vol. 24, No. 1, pp. 44-59, 2004.

- [2] P. Benner and J. Saak, *Efficient numerical solution of the LQR-problem for the heat equation*, submitted to Proc. Appl. Math. Mech., 2004.
- [3] D.C. Hyland and D.S. Bernstein, *The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore*, IEE. Trans. Automat. Contr., Vol. 30, No. 12, pp. 1201-1211, 1985.
- [4] A. Lepschy, G.A. Mian, G. Pinato and U. Viaro, *Rational L_2 approximation: A non-gradient algorithm*, 30th CDC, 1991.
- [5] L. Meier and D.G. Luenberger, *Approximation of Linear Constant Systems*, IEE. Trans. Automat. Contr., Vol. 12, pp. 585-588, 1967.
- [6] A. C. Antoulas, D. C. Sorensen, and S. Gugercin, *A survey of model reduction methods for large scale systems*, Contemporary Mathematics, AMS Publications, **280**: 193-219, 2001.
- [7] A.E. Bryson and A. Carrier, *Second-order algorithm for optimal model order reduction*, J. Guidance Contr. Dynam., pp-887-892, 1990.
- [8] D. Gaier, *Lectures on Complex Approximation*, Birkhauser, 1987.
- [9] K. Gallivan, E. Grimme, and P. Van Dooren, *A rational Lanczos algorithm for model reduction*, Numerical Algorithms, 2(1-2):33-63, April 1996.
- [10] E.J. Grimme, *Krylov Projection Methods for Model Reduction*, Ph.D. Thesis, ECE Dept., U. of Illinois, Urbana-Champaign, 1997.
- [11] S. Gugercin and A. C. Antoulas, *A comparative study of 7 model reduction algorithms*, Proceedings of the 39th IEEE Conference on Decision and Control, Sydney, Australia, December 2000.
- [12] S. Gugercin, A.C. Antoulas and C.A. Beattie, *An iterative rational Krylov approach to optimal \mathcal{H}_2 model reduction*, Technical Report, Department of Mathematics, Virginia Tech., September 2004.
- [13] B. C. Moore, *Principal component analysis in linear system :controllability, observability and model reduction*, IEEE Transactions on Automatic Control, **AC-26**:17-32, 1981.
- [14] C. T. Mullis and R. A. Roberts, *Synthesis of minimum roundoff noise fixed point digital filters*, IEEE Trans. on Circuits and Systems, **CAS-23**, pp: 551-562, 1976.
- [15] Thilo Penzl, *Algorithms for model reduction of large dynamical systems*, Technical Report SFB393/99-40, Sonderforschungsbereich 393 Numerische Simulation auf massiv parallelen Rechnern, TU Chemnitz, 09107, FRG, 1999. Available from <http://www.tu-chemnitz.de/sfb393/sfb99pr.html>.
- [16] J.T. Spanos, M.H. Millman, and D.L. Mingori, *A new algorithm for \mathcal{L}_2 optimal model reduction*, Automatics, pp. 897-909, 1992.
- [17] D.A. Wilson, *Optimum solution of model reduction problem*, in Proc. Inst. Elec. Eng., pp. 1161-1165, 1970.
- [18] W-Y. Yan and J. Lam, *An approximate approach to \mathcal{H}_2 optimal model reduction*, IEEE Transactions on Automatic Control, **AC-44**, pp. 1341-1358, 1999.