Raport Badawczy

Research Report

RB/28/2015

On the asymptotic accuracy of reduced-order models

D. Casagrande, W. Krajewski,
U. Viaro

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Zbigniew Nahorski

Warszawa 2015

On the asymptotic accuracy of reduced–order models

Daniele Casagrande, Wiesław Krajewski, and Umberto Viaro

Abstract: Popular model reduction methods can easily be adapted to retain the asymptotic response to inputs with rational transform. To this purpose, the forced response of the high–order system is decomposed into a transient and a steady–state component. Then, the reduced–order model is obtained by combining the unaltered steady–state component with an approximation of the transient component. Examples show that forcing the reduced–order model to retain the steady–state component does not compromise the transient accuracy.

Keywords: Model reduction, Steady–state response, Transient response, Balanced truncation, Hankel norm

1. INTRODUCTION

The model reduction problem has received considerable attention since the dawn of system and control theory (see [1] [2] [3] [4] for recent surveys) and, in many respects, the subject has reached a high level of maturity, even if important new impulses to research are still coming from emerging areas [5] [6] [7] and clever ideas [8] [9] [10]. In fact, despite the increase of computing capabilities witnessed in recent years, reduced–order models are being widely used for both analysis and design purposes; suffice it here to recall the control of flexible mechanical structures [11], the simulation of integrated circuits [12], the study of power grid networks [13] and climate modelling [14].

Usually, reduced–order models are obtained by referring, either explicitly or implicitly, to the minimisation of a norm or a semi-norm of the approximation error in the response to suitable inputs. For instance, the classic Padé technique and the so–called moment–matching methods set to zero the error transform at specific frequencies ($s = 0$ in the classic Padé method), which, for stable systems, entails zeroing asymptotically the error in the response to harmonic input signals at these frequencies (step and, possibly, other canonical inputs in the classic Padé method). These techniques, however, do not ensure stability retention, at least in their original versions. On the other hand, popular stability–preserving methods, such as balanced truncation [15], Hankel–norm approximation [16] and L_2 model reduction [1] [22], do not ensure the retention of the asymptotic response to any canonical input (except possibly for the step, at the expense of exact optimality, by resorting to singular perturbation approximation [17] [18]), even if their *transient* accuracy is usually very good compared to alternative techniques. Moreover, fairly robust (and easily accessible) algorithms exist for their implementation [19] [20] [21] [22] [23], which makes them applicable to the reduction of very large–scale systems despite their non–negligible numerical complexity (related to the solution of high–dimensional Lyapunov equations in the case balanced truncation or the iterative solution of interpolation problems in the case of L_2 approximation).

This note suggests a simple way to ensure the retention of the asymptotic response to predetermined persistent inputs with rational transform. To this purpose, the transform of the related original forced response is decomposed into the sum of a system component and an input component, according to the terminology in [24] [25]. For stable systems, the first component accounts for the system's transient behaviour, whereas the second accounts for its asymptotic behaviour. Then, the approximation procedure is applied only to the system component whereas the input component is reproduced exactly. In a sense, a similar approach has been followed in [26] [27] [28] [29] for approximating unstable systems via stable/antistable decomposition.

2. DECOMPOSITION OF THE FORCED RESPONSE

The suggested reduction technique is based on the additive decomposition of the forced response outlined next. Denote the strictly–proper rational transfer function of the original system by

$$W(s) = \frac{B(s)}{A(s)}, \tag{1}$$

Manuscript received
Daniele Casagrande and Umberto Viaro are with the Department of Electrical, Management and Mechanical Engineering, University of Udine, via delle Scienze 206, 33100 Udine, Italy (e-mail: {daniele.casagrande, umberto.viaro}@uniud.it).
Wiesław Krajewski is with the Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland (e-mail: krajewsk@ibspan.waw. pl).

where $A(s)$ and $B(s)$ are coprime polynomials, and the proper rational Laplace transform of the input $u(t)$ by

$$U(s) = \frac{N(s)}{D(s)}, \tag{2}$$

where $N(s)$ and $D(s)$ are also coprime polynomials.

Throughout, it is further assumed that $B(s)$ and $D(s)$ are coprime as well as $N(s)$ and $A(s)$. In this way, no cancellation occurs in the expression of the strictly–proper Laplace transform of the forced response $y_f(t)$

$$Y_f(s) = W(s)U(s) = \frac{B(s)N(s)}{A(s)D(s)}. \tag{3}$$

If polynomials $A(s)$ and $D(s)$ at the denominator of (3) have common roots, they can be expressed as

$$A(s) = \bar{A}(s)C_A(s), \tag{4}$$
$$D(s) = \bar{D}(s)C_D(s), \tag{5}$$

where $C_A(s)$ is a factor containing all the roots of $A(s)$ (with their multiplicities) that are also roots of $D(s)$ and $C_D(s)$ is a factor containing all the roots of $D(s)$ (with their multiplicities) that are also roots of $A(s)$. Denoting the product of these two factors by

$$C(s) := C_A(s)C_D(s), \tag{6}$$

the three pairs $(\bar{A}(s),\bar{D}(s))$, $(\bar{A}(s),C(s))$ and $(C(s),\bar{D}(s))$ are all formed by coprime polynomials, so that $Y_f(s)$ can be expressed *uniquely* as the sum of three strictly–proper rational functions as:

$$Y_f(s) = Y_w(s) + Y_u(s) + Y_c(s), \tag{7}$$

where

$$Y_w(s) = \frac{N_w(s)}{\bar{A}(s)},\ Y_u(s) = \frac{N_u(s)}{\bar{D}(s)},\ Y_c(s) = \frac{N_c(s)}{C(s)}, \tag{8}$$

as may easily be proved by resorting to the Bezout identity (see, e.g., [30, p. 204]).

Since the inverse Laplace transform $y_w(t)$ of the first addendum on the right–hand side of (7), whose poles are those of $W(s)$ not in common with $U(s)$, is a combination of system modes, it is called *system component* in [24]. Similarly, $y_u(t) = \mathrm{LT}^{-1}[Y_u(s)]$ is a combination of the modes of $u(t)$ that are not in common with $w(t)$ and is therefore called *input component*. The poles of the third addendum $Y_c(s)$ in (7) are those common to $W(s)$ and $U(s)$, but their multiplicity is the sum of the multiplicities of the same poles in $W(s)$ and $U(s)$; therefore, $y_c(t) = \mathrm{LT}^{-1}[Y_c(s)]$ can reasonably be called *resonant component*. A characterisation of these components in terms of the solution of either a homogeneous or a non–homogeneous differential equation is provided in [24] together with some suggestions on how to extend the definitions to non–rational input and system transforms.

In most cases, $A(s)$ and $D(s)$ have no common factors so that the forced–response is given by the sum of the system and input components only, i.e.:

$$y_f(t) = y_w(t) + y_u(t). \tag{9}$$

This simplifying assumption will be made in the sequel since the consideration of the more general case would entail a substantial increase in notation without a corresponding gain in insight.

Although the decomposition (9) holds for stable and unstable systems, as well as for both persistent and vanishing inputs, it is particularly meaningful for the purpose of model reduction when $u(t)$, and thus $y_u(t)$, is persistent and $w(t)$, and thus $y_w(t)$, tends asymptotically to zero, i.e., the system is BIBO stable. In this case, $y_u(t)$ corresponds to the *steady–state* (or *asymptotic*) *response* and $y_w(t)$ to the *transient response*. In the sequel, reference is made to such a situation.

3. REDUCTION PROCEDURE

According to the previous considerations, it is assumed next that:
(i) the original system is BIBO stable,
(ii) all of the poles of $U(s)$ are in the closed right half–plane so that $A(s)$ and $D(s)$ are necessarily coprime and all of the modes of $u(t)$ are persistent.

The suggested reduction procedure consists of the following steps.

1. Find an approximation

$$Y_{w_r}(s) = \frac{N_{w_r}(s)}{A_r(s)} \tag{10}$$

of the original transient response transform $Y_w(s)$ according to a suitable reduction criterion (e.g., balance truncation, Hankel–norm approximation or L_2 reduction).

2. Form the reduced–order forced response transform $Y_{f_r}(s)$ as

$$Y_{f_r}(s) = Y_{w_r}(s) + Y_u(s) + Y_a(s), \tag{11}$$

where

$$Y_a(s) = \frac{N_a(s)}{A_a(s)} \tag{12}$$

is an auxiliary strictly–proper stable rational function whose task is to guarantee that the successive step admits a solution.

3. Determine the strictly–proper reduced–order transfer function $W_r(s)$ as

$$W_r(s) = \frac{Y_{f_r}(s)}{U(s)} = \frac{B_r(s)}{A_r(s)A_a(s)}, \tag{13}$$

where $\deg[B_r(s)] = \deg[A_r(s)] + \deg[A_a(s)] - 1$.

A few considerations are in order.

Remark 1: The auxiliary term (12) is needed because, otherwise, the reduction procedure would not admit a solution, except for the case of impulsive inputs (see Remark 3 below).

Remark 2: The poles of $Y_a(s)$ must lie in the open left half–plane not to influence the asymptotic response. Also, these poles should lie far away from the poles of $Y_{w_r}(s)$ not to affect appreciably the transient behaviour at the dominant frequencies of the reduced model.

Remark 3: The order of $Y_a(s)$, i.e., the degree of its denominator, depends on the order of the component $Y_u(s)$ that must be retained and, thus, on the order of the input. To clarify this point, consider equation (13) which, in view of (10), (11) and (12), leads to the polynomial identity:

$$B_r(s)N(s) =$$

$$N_{w_r}(s)D(s)A_a(s) + N_u(s)A_r(s)A_a(s) + N_a(s)A_r(s)D(s) \quad (14)$$

whose degree, given the properness of $U(s)$ and the strict properness of $Y_{w_r}(s)$, $Y_u(s)$ and $Y_a(s)$, is

$$N = \deg[D(s)] + \deg[A_r(s)] + \deg[A_a(s)] - 1.$$

It follows that, by equating the coefficients of the equal powers of s on both sides of (14), a system of $N+1$ equations is obtained. In order for this system to admit a unique solution, an equal number of unknowns must be present. Now, if the denominator $A_a(s)$ of $Y_a(s)$ is fixed beforehand to ensure that its roots are far away to the left, the only unknowns are: (i) the coefficients of $B_r(s)$, whose number is $\deg[A_r(s)] + \deg[A_a(s)]$, and (ii) the coefficients of $N_a(s)$, whose number is $\deg[A_a(s)]$. In conclusion, the system of equations admits a unique solution only if $N+1 = \deg[A_r(s)] + 2\deg[A_a(s)]$, that is,

$$\deg[D(s)] = \deg[A_a(s)]. \quad (15)$$

Remark 4: A consequence of Remark 3 is that the order of the reduced model $W_r(s)$ is greater than the order of the simplified transient term $Y_{w_r}(s)$ by an amount equal to $\deg[D(s)]$. For steps, ramps and sinusoidal inputs, $\deg[D(s)] \leq 2$ which is usually quite negligible compared to the order of the approximating model of a very high–dimensional system [31].

Remark 5: The aforementioned approximation procedure ensures that the reduced–order model is strictly proper like the original system, whereas the the balanced truncation procedure and its DC–gain–preserving variant, illustrated, e.g., in [18], lead, in general, to exactly–proper reduced–order models.

4. EXAMPLES

To show the performance of the suggested reduction technique, two benchmark examples are worked out next by assuming that the asymptotic response to be retained is the one in the **ramp response.** In both cases, the reduced–order transient response to such an input is determined by minimising the Hankel norm of the approximation error using the Matlab® function [28] `hankelmr`. Of course, other simplification methods could be employed as well. The results of the suggested approximation procedure are then compared with those afforded by the balanced truncation and Hankel–norm approximation methods applied directly to the original system without consideration of the asymptotic response to a ramp input. Note, however, that the Matlab® function `balred` adjusts the reduced model derived by truncating the original balanced realisation in such a way that the steady–state value in the response to a *step input* is retained exactly [28]. This adjustment obviously entails that the reduced model is no longer a perfect truncation of the original balanced realisation.

4.1. Hospital building

This example, whose original equations can be found in [2] and [31], describes the dynamics of a hospital building with 8 floors, each having 3 degrees of freedom (horizontal and vertical displacements and rotation). The order of the original model (state–space dimension) is 48 (twice the number of spacial coordinates). The input $u(t)$ is the force acting in the horizontal direction of the first floor (state $x_1(t)$). The output of interest coincides with the derivative of the first coordinate (i.e., the 25th state $x_{25}(t)$). Fig. 1 depicts the original model (OM) response to a unit step input as well as the step responses of the 6th–order models obtained using Hankel–norm approximation (HN), balanced truncation (with and without DC gain adjustment) (BY and BN, respectively) and the method (AR) outlined in Section 3. The last model has been obtained from the 4th–order approximation of the transient component in the ramp response; of course, the approximation would be better if a 6th–order approximation were used for this purpose, but the resulting reduced–order transfer function would then be of order 8. Figs. 2 and 3 compare, respectively, ramp responses and Bode diagrams. Observe that, although the ramp response of the balanced–truncation model adjusted to retain the DC gain seems to coincide asymptotically with the original ramp response, it slightly differs from it for $t \to \infty$. Clearly, the frequency responses of the reduced models obtained by minimising the Hankel norm of the error (HN) and by retaining the asymptotic ramp response (AR) are much closer to the original one at high frequencies since they are strictly proper like the original system.

4.2. 1006th–order system

A 1006th–order model has been described in [32] and detailed in [31]. Fig. 4 depicts the original model response to a *ramp input* (OM) as well as the ramp responses of the 9th–order models obtained according to the Hankel–

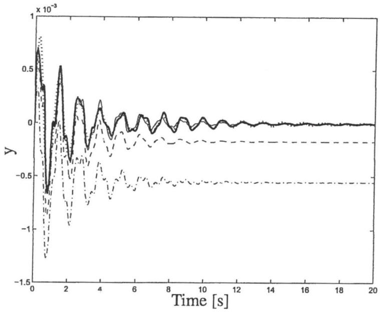

Fig 1: Step responses of the models of the hospital building: OM (bold solid line), BN (dashed line), BY (thin solid line), HN (dash–dotted line), and AR (bold dotted line). The responses of models OM, BY and AR overlap asymptotically.

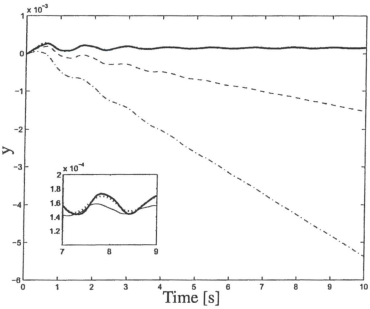

Fig 2: Ramp responses of the models of the hospital building: OM (bold solid line), BN (dashed line), BY (thin solid line), HN (dash–dotted line), and AR (bold dotted line). Only the last response tends to reproduce exactly the original response.

norm approximation method (HN), the balanced truncation method (BY) (modified to match the original DC gain) and the method retaining the asymptotic ramp response (AR) outlined in Section 3. The last model has been formed from the 7th–order approximation of the transient component of the ramp response (of course, better results are obtained if a 9th–order approximation of the transient term is used instead). This time, the model resulting from balanced truncation without DC gain adjustment has been excluded because it leads to much poorer results. Observe that, although the ramp response of model BY seems to coincide with the original ramp response, it actually exhibits a small asymptotic error equal to 0.074. Fig. 5 shows the original Bode diagrams together with the Bode diagrams of all of the aforementioned 9th–order models. Clearly, forcing the retention of the asymptotic response ($t \to \infty$) to canonical inputs (steps and ramps) leads to a better approximation in the low frequency range ($\omega \to 0$) as is the case with the classic Padé approximation. Note, finally, that, the magnitude plot of the gain–adjusted model BY diverges from the original magnitude plot for $\omega \to \infty$ because its transfer function is not strictly proper.

5. CONCLUSIONS

It has been shown how model reduction techniques can be adapted to reproduce the asymptotic response to inputs with rational transform. The suggested procedure is not computationally demanding and entails only a small increment of the reduced model order with respect to the order of the approximating transient term. Benchmark examples have shown that the responses of the reduced–order models obtained in this way compare favourably with those afforded by conventional reduction techniques that do not take care *ab initio* of the asymptotic response.

REFERENCES

[1] A.C. Antoulas, *Approximation of Large–Scale Dynamical Systems*. Society for Industrial and Applied Mathematics (SIAM), Series Advances in Design and Control, Philadelphia, USA, 2005.

[2] A.C. Antoulas, D.C. Sorensen, and S. Gugercin, "A survey of model reduction methods for large-scale systems", American Mathematical Society, Contemporary Mathematics Series, vol. 280, pp. 193–220, 2001.

[3] U. Baur, C. Beattie, P. Benner, and S. Gugercin, "Interpolatory projection methods for parameterized model reduction", SIAM J. Scientific Computing, vol. 33, no. 5, pp. 2489–2518, 2011.

[4] A. Bultheel and B. De Moor, "Rational approximation in linear systems and control", J. Comput. Appl. Math., vol. 121, nos. 1–2, pp. 355–378, 2000.

[5] A. Astolfi, "Model reduction by moment matching for linear and nonlinear systems", IEEE Trans. Automat. Contr., vol. 55, no. 10, pp. 2321–2336, 2010.

[6] P. Benner, S. Gugercin, and K. Willcox, "A survey of model reduction methods for parametric systems", Max Planck Institute, Magdeburg, preprint no. MPIMD/13-14, 2013.

[7] S. Gugercin, R.V. Polyuga, C. Beattie, and A. van der Schaft, "Structure–preserving tangential interpolation for model reduction of port–Hamiltonian systems", Automatica, vol. 48, no. 9, pp. 1963–1974, 2012.

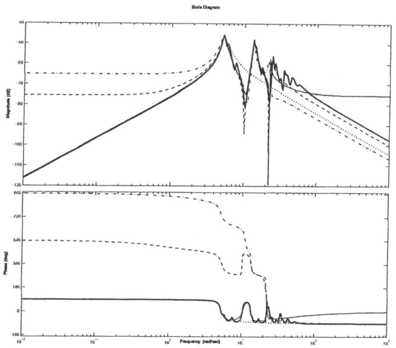

Fig 3: Bode diagrams of the models of the hospital building: OM (solid line), BN (dashed line), BY (thin solid line), HN (dash–dotted line), and AR (bold dotted line). The magnitude plot of model BY tends to a constant value as $\omega \to \infty$.

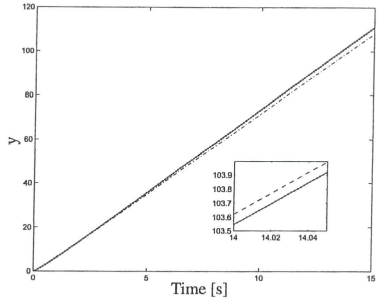

Fig 4: Ramp responses of the models for the 1006th–order example: OM (solid line), BY (dashed line), HN (dash–dotted line), and AR (bold dotted line).

[8] S.R. Desai and R. Prasad, "A new approach to order reduction using stability equation and big bang big crunch optimization", Systems Science & Control Engineering: An Open Access Journal, vol. 1, no. 1, pp. 20–27, 2013.

[9] D. Petersson and J. Löfberg, "Model reduction using a frequency–limited H_2 cost", Systems and Control Lett., vol. 67, no. 12, pp. 32–39, 2012.

[10] A. Sikander and R. Prasad, "A novel order reduction method using cuckoo search algorithm", IETE J. Research, vol. 61, no. 2, pp. 83–90, 2015.

[11] W.K. Gawronski, *Dynamics and Control of Structures: A Modal Approach*. Springer, New York, USA, 1998.

[12] P. Benner, M. Hinze, and E. Jan W. ter Maten (Eds.), *Model Reduction for Circuit Simulation*. Springer, Dordrecht, The Netherlands, 2011.

[13] Z. Zhang, X. Hu, C.-K. Cheng, and N. Wong, "A block-diagonal structured model reduction scheme for power grid networks", Proc. Design, Automation & Test in Europe Conference & Exhibition (DATE), Grenoble, France, 14–18 March 2011, pp. 44–49.

[14] D. Kondrashov, M.D. Chekroun, A.W. Robertson, and M. Ghil "Low-order stochastic model and 'past-noise forecasting' of the Madden-Julian Oscillation", Geophysical Research Letters, vol. 40, no. 19, pp. 5305–5310, 2013.

[15] B.C. Moore, "Principal component analysis in linear systems: controllability, observability, and model reduction", IEEE Trans. Automat. Contr., vol. 26, no. 1, pp. 17–32, 1981.

[16] K. Glover, "All optimal Hankel–norm approximations of linear multivariable systems and their L_∞ error bounds", Int. J. Control, vol. 39, no. 6 , pp. 1115–1193,1984.

[17] P.V. Kokotovic, H.K. Khail, and J.O. Reilly, *Singular Perturbation Methods in Control: Analysis and Design*. Academic Press, London, UK, 1986.

[18] P. Benner and E.S. Quintana–Ortí, "Model reduction based on spectral projection methods", in P. Benner, V. Mehrmann, and D.C. Sorensen, *Dimension Reduction of Large–Scale Systems*. Lecture Notes in Computational Science and Engineering, vol. 45. Springer, Berlin, Germany, 2005, pp. 5–48.

[19] A.C. Antoulas, C.A. Beattie, and S. Gugercin, "Interpolatory model reduction of large–scale dynamical systems", in J. Mohammadpour and K.M. Grigoriadis (Eds.), *Efficient Modelling and Control of Large–Scale Systems*. Springer, New York, USA, 2010.

[20] D.K. Chaturvedi, *Modelling and Simulation of Systems Using Matlab*. Ch. 4. CRC Press, Boca Raton, Fl, USA. 2009.

[21] I.M. Jaimoukha and E.M. Kasenally, "Krylov subspace methods for solving large Lyapunov equations", SIAM J. Numerical Analysis, vol. 31, no. 1, pp. 227–251, 1994.

[22] W. Krajewski and U. Viaro, "Iterative–interpolation algorithms for L_2 model reduction", Control and Cybernetics, vol. 38, no. 2, pp. 543–554, 2009.

[23] P. Rabiei and M. Pedram, "Model order reduction of large circuits using balanced truncation", Proc. Asia and South Pacific Design Automation Conf., Whan Chai, Hong Kong Island, Southern China, 18–21 Jan.1999, vol. 1, pp. 237-240.

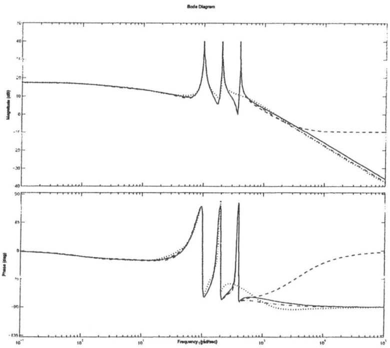

Fig 5: Bode diagrams of the models for the 1006th–order example: OM (solid line), BY (dashed line), HN (dash–dotted line), and AR (bold dotted line). The magnitude plot of model BY tends to a constant value as $\omega \to \infty$.

[24] P. Dorato, A. Lepschy, and U. Viaro, "Some comments on steady-state and asymptotic responses", IEEE Trans. Education, vol. 37, no. 3, pp. 264–268, 1994.

[25] A. Ferrante, W. Krajewski, A. Lepschy, and U. Viaro, "Remarks on the steady–state accuracy of a feedback control system", Control and Cybernetics, vol. 29, no. 1, pp. 51–67, 2000.

[26] S. Barrachina, P. Benner, E.S. Quintana–Ortí, and G. Quintana–Ortí, "Parallel algorithms for balanced truncation of large–scale unstable systems", Proc. 44th Conf. Decision and Control and European Control Conf., Seville, Spain, 12–15 Dec. 2005, pp. 2248–2253, 2005.

[27] P. Benner , M. Castillo , E.S. Quintana–Ortí , and G. Quintana–Ortí, "Parallel model reduction of large-scale unstable systems", in G.R. Joubert, W.E. Nagel, F.J. Peters, and W.V. Walter (Eds.), *Parallel Computing: Software Technology, Algorithms, Architectures and Applications*. Series Advances in Parallel Computing, vol. 13. Elsevier (North–Holland), Amsterdam, The Netherlands, 2004, pp.251–258.

[28] Control System Toolbox™ of Matlab® (Model Simplification), "Approximate model with unstable or near–unstable pole", MathWorks®, http://it.mathworks.com/help/control/ug/low-order-approximation-for-model-with-unstable-pole.html

[29] K. Zhou, G. Salomon, and E. Wu, "Balanced realization and model reduction for unstable systems", Int. J. Robust Nonlinear Control, vol. 9, no. 3, pp. 183–198, 1999.

[30] A. Ferrante, A. Lepschy, and U. Viaro, *Introduzione ai Controlli Automatici* [Introduction to Automatic Control]. UTET, Torino, Italy, 2000.

[31] Y. Chahlaoui and P. Van Dooren, "A collection of benchmark examples for model reduction of linear time invariant dynamical systems", SLICOT Working Note 2002-2: February 2002 (http://slicot.org/20-site/126-benchmark-examples-for-model-reduction).

[32] T. Penzl, "Algorithms for model reduction of large dynamical systems", Linear Algebra and its Applications, vol. 415, nos. 2–3, pp. 322–343, 2006.