Outstanding Issues in Model Order Reduction

Luís Miguel Silveira
INESC ID
Department of Electrical and Computer Engineering
IST - Technical University of Lisbon
Lisbon, Portugal
lms@inesc-id.pt
http://algos.inesc-id.pt/ lms

***Abstract*— With roots dating back many years and applications in a wide variety of areas, model order reduction has emerged in the last few decades as a crucial step in the simulation, control, and optimization of complex physical processes. Reducing the order or dimension of models of such systems, is paramount to enabling their simulation and verification. While much progress has been achieved in the last few years regarding the robustness, efficiency and applicability of these techniques, certain problems of relevance still pose difficulties or renewed challenges that are not satisfactorily solved with the existing approaches. Furthermore, new applications for which dimension reduction is crucial, are becoming increasingly important, raising new issues in our quest for increased performance. This talk is aimed at presenting an overview of some of those issues and at sparking renewed discussion about how to tackle them.**

Keywords— Model Order Reduction, Krylov Projection, Massively coupled systems, Parametric System Description, Circuit Simulation.

I. Basics of MOR

Model reduction algorithms are now standard techniques in many areas, including the integrated circuits community. Such techniques are commonly used for approximation and analysis of models arising from interconnect and electromagnetic structure analysis. The need for higher accuracy while accounting for all relevant physical effects, implies that the mathematical formulations used to describe and model such systems often result in very large scale systems. Reducing the order or dimension of such models is paramount to enabling the simulation and verification of these systems. In recent years the need for reduction techniques for large-scale systems has triggered a revival of research activities in model order reduction. Initial interest in model reduction techniques stemmed from efforts to accelerate analysis of circuit interconnect. More recently, however, model reduction has come to be viewed as a method for generating compact models from all sorts of physical system modeling tools. Such techniques are now routinely used to generate lumped-circuit approximations of distributed electronic circuit elements, such as the interconnect or package of VLSI chips, or in simulations of microelectromechanical systems (MEMS), which have both electrical and mechanical components. While enormous progress has been achieved both from a theoretical as well as practical standpoint, still greater challenges lie ahead as new and exciting applications are being researched for which order reduction is again a crucial step.

Model order reductions seeks to replace a large-scale models of a physical system by a system of lower dimension which exhibits similar behavior, typically measured in terms of its frequency or time response characteristics. Existing methods for model reduction can be broadly characterized into two types: those that are based on projection methods, and those based on balancing techniques. Among the first, Krylov subspace projection methods such as PVL [1] and PRIMA [2] have been the most widely studied over the past decade. They are very appealing due mostly to their simplicity and their strong performance in terms of efficiency and accuracy. However, Krylov projection methods are known to have a few drawbacks in practical application. First, there is no general agreement on how to control error in these methods. Error estimators do exist for some methods but they are seldom used in practice as they require additional computations, which can be expensive and awkward to implement. Second, moment based methods such as PRIMA, are known in some cases to produce models that are "too high" in order with the obvious consequences in terms of analysis or simulation cost [3], [4]. Third, any guarantees on retaining relevant properties of the underlying system, such as passivity of the models, are dependent on the structure of the system representation, a non-issue for RLC-type models, but a troublesome problem for other types of models. An alternate class of model reduction schemes are the truncated balanced realization (TBR) family [5]. These are purported to produce "nearly optimal" models and have easy to compute a-posteriori error bounds. However, they are expensive to apply, as they require the computation of the system Gramians by solving a large Lyapunov or Sylvester-type equation. This is an expensive procedure, which limits their applicability to small to medium sized systems. Balanced realization-based techniques are also awkward to implement correctly and Both types of methods have carved their niche in specific segments or applications. Furthermore hybrid techniques that combine some of the features of each type of method have also been presented. Examples of these include the solution of the large Lyapunov equations via a Krylov subspace method [6], [7]. Other approaches include a two-step method where an initial reduced model is first obtained via projection and then further compressed using a TBR method [3]. Recently a new technique has been proposed that attempts to establish a connection between the two techniques. The Poor Man's TBR [8] explores a connection between multipoint projection and a different interpretation of the system Gramians. It leads to a projection scheme where the projection matrix approximately spans the dominant eigenspace of the Gramian matrix and provides an interesting platform for bridging between the two types of techniques. Still the technique is not without drawbacks, as it relies on proper choice of sampling points, a non-trivial task in general.

II. Issues with MOR

In spite of the issues previously mentioned, both projection-based and balanced-realization type methods are in widespread use nowadays. Clearly such issues, irrespective of their seriousness, have not affected the popularity of those methods. Still, there are situations where neither method presents itself as a satisfactory solution. Additionally, new challenges are being posed that required further research into these methods.

As an example of the first type of problems consider the problem of reducing systems with a large number of input/output ports, also known as "massively coupled" systems. Such systems typically occur in substrate and package parasitic networks. Algorithms such as PRIMA [2] and PVL [1] are considered impractical for such networks. They rely on block iterations, where the size of the block equals the number of input/output ports. Therefore each block iteration considerably increases the size of the model. For example, if a moment-matching (Krylov-subspace) algorithm is used to reduce a network with 1000 ports, and if only two (block) moments are to be matched at each port, the resulting model will have 2000 states, and the reduced system matrices will be dense. This makes simulation in the presence of nonlinear elements impractical. TBR is intrinsically somewhat less sensitive to the number of inputs ports. Unfortunately such systems are typically very large which makes reduction based on balancing techniques impractical. Hybrid methods are also not helpful as the initial projection-based reduction is ineffective, making the TBR-based 2^{nd} step too expensive. In the PMTBR framework, however, it is possible to exploit circuit functional information that results in correlations between the waveforms incident on the parasitic network ports. By exploiting this information, an *input-correlated* variant of the PMTBR procedure can be derived that enables significant further model order reduction. Unfortunately such reduction is highly dependent upon knowledge of the input time behavior and does not generalize well. Clearly further research in needed into this problem.

As an example of new challenges that arise for model order reduction techniques is the problem of order reduction of parameterized systems. Parameter-based descriptions are now starting to be used for variability-aware design and verification. Operating conditions, such as temperature, as well as relevant process and geometric features, will parameterize such models. Parameter description may also reflect geometric concerns, free parameters to be optimized by design optimization procedures, layout issues as well as coupling-related information. For high frequency, at nanoscale feature sizes, process variability effects as well as dependence on operating conditions becomes extremely relevant and should be accounted for in the models. Existing techniques for handling such systems are straightforward extensions to the basic order reduction algorithms [9], [10]. Projection-based techniques match Taylor-series coefficients, which in parameter-based descriptions are multidimensional moments. Unfortunately this technique has exponential cost increase with the number of parameters and is thus expensive except for small size and small number of parameters. Building a projection space assuming small perturbations around the nominal operating point is also problematic: it is hard to do anything beyond first-order and thus it is not clear how to dial in accuracy. Sampling the parameter space also presents a challenge, as it is not clear where to place sample point in such a multidimensional space. Still if some information regarding the statistical distribution of the parameter values is available, this can be used to guide the sampling and to build the model accordingly.

In this talk, these and other issues will be discussed in order to stimulate discussion leading to further advances in this field.

References

[1] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Padé approximation via the Lanczos process," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, no. 5, pp. 639–649, May 1995.

[2] A. Odabasioglu, M. Celik, and L. T. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. Computer-Aided Design*, vol. 17, no. 8, pp. 645–654, August 1998.

[3] Mattan Kamon, Frank Wang, and Jacob White, "Generating nearly optimally compact models from Krylov-subspace based reduced-order models," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 4, pp. 239–248, April 2000.

[4] Nuno Marques, Mattan Kamon, Jacob White, and L. Miguel Silveira, "A mixed nodal-mesh formulation for efficient extraction and passive reduced-order modeling of 3d interconnects," in 35^{th} *ACM/IEEE Design Automation Conference*, San Francisco, CA, USA, June 1998, pp. 297–302.

[5] Bruce Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction," *IEEE Transactions on Automatic Control*, vol. AC-26, no. 1, pp. 17–32, February 1981.

[6] Jing-Rebecca Li, F. Wang, and J. White, "Efficient model reduction of interconnect via approximate system grammians," in *International Conference on Computer Aided-Design*, SAN JOSE, CA, November 1999, pp. 380–383.

[7] I. M. Jaimoukha and E. M. Kasenally, "Krylov subspace methods for solving large Lyapunov equations," *SIAM Journal on Numerical Analysis*, vol. 31, pp. 227–251, 1994.

[8] Joel R. Phillips and L. Miguel Silveira, "Poor Man's TBR: A simple model reduction scheme," *IEEE Trans. Computer-Aided Design*, vol. 24, no. 1, pp. 43–55, Jan. 2005.

[9] L. Daniel, O. C. Siong, S. C. Low, K. H. Lee, and J. K. White, "A multiparameter moment-matching model-reduction approach for generating geometrically parametrized interconnect performance models," *IEEE Trans. Computer-Aided Design*, vol. 23, pp. 678–693, May 2004.

[10] X. Li, P. Li, and L. Pileggi, "Parameterized interconnect order reduction with Explicit-and-Implicit multi-Parameter moment matching for Inter/Intra-Die variations," in *International Conference on Computer Aided-Design*, San Jose, CA, November 2005, pp. 806–812.