Resampling Plans for Sample Point Selection in Multipoint Model-Order Reduction

Luis Miguel Silveira, *Senior Member, IEEE*, and Joel R. Phillips, *Senior Member, IEEE*

***Abstract*—Multipoint projection methods have gained much notoriety in model-order reduction of linear, nonlinear, and parameter-varying systems. A well-known difficulty with such methods lies in the need for clever point selection to attain model compactness and accuracy. In this paper, the authors present a method for sample point selection in multipoint projection-based model-order reduction. The proposed technique, which is borrowed from the statistical modeling area, is based on resampling schemes to estimate error and can be coupled with recently proposed order reduction schemes to efficiently produce accurate models. Two alternative implementations are presented: 1) a rigorous linear-matrix-inequality-based technique and 2) a simpler, more efficient, heuristic search. The goal of this paper is to answer two questions. First, can this alternative metric be effective in selecting sample points in the sense of placing points in regions of high error without recourse to evaluation of the larger system? Second, if the metric is effective in this sense, under what conditions are substantial improvements in the model reduction efficiency achieved? Results are shown that indicate that the metric is indeed effective in a variety of settings, therefore opening the possibility for performing adaptive error control.**

***Index Terms*—Boosting, model order reduction, multipoint projection, parameter-varying systems, resampling.**

I. Introduction

MODEL reduction algorithms are now standard techniques in the integrated circuits community for analysis, approximation, and simulation of models arising from interconnect and electromagnetic structure analysis. Krylov subspace projection methods such as Padé-via-Lanczos (PVL) [1] and Passive Reduced-Order Interconnect Modeling Algorithm (PRIMA) [2] have been the most widely studied over the past decade. Model-order reduction methods have also been applied with some degree of success to modeling weakly nonlinear [3]–[5] and parameter-varying systems [6], [7]. A more sophisticated approach to projection-based model-order reduction is to construct the projection space from a rational, or multipoint, Krylov subspace [8]–[10]. For a given model order, the multipoint approximants tend to be more accurate but are usually more expensive to construct. Still, multipoint approximations are thought of as very effective because they are known to produce better quality models for a given effort. Multipoint methods also provide a simple framework for nonlinear systems or parameter-varying descriptions by allowing expansions to be performed under different frequencies, operating points, or parameter settings [7].

Although very appealing, multipoint projection methods raise many practical questions to ponder in an actual implementation. In this paper, we will focus solely on one of these questions, namely the issue of point selection: how many points to choose and how to place them. These issues are particularly troublesome in the context of Krylov projection methods due to what is known as a major drawback of such methods. There is no general agreement on how to control error in these procedures. Error estimators do exist for some methods [11]–[13], but they are seldom used in practice and require additional computations, which can be expensive and awkward to implement. Furthermore, most of the estimators we are aware of are developed for single expansion point techniques rather than multipoint approximations. However, error control with multipoint projection is more complicated since the quality and compactness of the models produced are dependent on a good choice of points for computing the samples that form the projection basis. Considerable effort has previously been devoted to multipoint projection from a variety of perspectives to improve accuracy in Krylov-based order reduction [8], [9], [13], [14]. These included placing points along the frequency axis in a uniform or random way or choosing a particular point for projection given some *a priori* knowledge of the system. The goal of such work was, however, not specifically directed to point placement, and as such, point placement was hardly analyzed at all. In fact, we are not aware of any systematic study of point selection strategies beyond variants of uniform refinement (e.g., interval bisection) nor of any work on structured multiple-input–multiple-output models (see [15] for examples of how point selection might be complicated in this case).

For nonlinear or parameter-varying systems [5], [7], [16], the increased dimensionality of the problem makes it even harder to appropriately choose the projection points. In fact, no clear strategy is known for appropriately choosing the points in such a setting: Uniform sampling and exhaustive search, the likely candidates, are not viable options due to exponential growth of computational cost with dimensionality. While in this paper we will only discuss linear system approximation, a primary goal of this paper is to lay the groundwork for point selection in these more difficult contexts.

An alternate class of model reduction schemes is the truncated balanced realization (TBR) family [17]. These are purported to produce "nearly optimal" models and have easy to compute *a posteriori* error estimators. However, their high

Manuscript received June 28, 2005; revised November 22, 2005. This paper was recommended by Associate Editor C.-J. R. Shi.

L. M. Silveira is with the Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa/Instituto Superior Técnico (INESC ID/IST), Technical University of Lisbon, 1169-047 Lisbon, Portugal, and also with Cadence Laboratories, Berkeley, CA 94704 USA (e-mail: lms@inesc-id.pt).

J. R. Phillips is with Cadence Laboratories, Berkeley, CA 94704 USA.

Digital Object Identifier 10.1109/TCAD.2006.882477

computational cost makes them unusable in practice for large-scale systems. In recent years, some effort has been devoted to efficiently solving the Lyapunov equations for approximate Gramians [10], [18]. Interestingly enough, such methods also rely on point selection algorithms and can also be cast as multipoint projection schemes, even if the goal of the projection is different there. As such, point selection is also of great concern in such a setting. Recently, a new algorithm, i.e., Poor Man's TBR (PMTBR), was introduced [19], which provides a direct connection between multipoint rational approximation techniques and TBR. The PMTBR algorithm can be described in terms of a probabilistic interpretation of the system Gramians that leads to the construction of a projection matrix based on samples of the system state. The number and the choice of samples are directly related to the model order, as well as to its accuracy.

This paper presents an algorithm for sample point selection that is based on three key ideas. First, we adopt the statistical PMTBR viewpoint that the projection vectors are constructed by principal components analysis over a space of samples of a random process. Thus, any set of samples with proper distribution is a valid choice of projectors. Second, a possible measure of the quality of a model (or family of models) is the degree of divergence, or the variance, between the different models in the family. Third, given a set of candidate sample vectors, the variance can be computed by "resampling" [20]–[22] on the candidate vectors. We point out that although our description is performed in the context of linear system order reduction, these operations are quite general and not restricted to linear modeling because our goal is to develop the foundations for a general point selection strategy. In particular, we are interested in parameter-varying problems where no viable strategies beyond random search have to date appeared.

This paper is organized as follows: In Section II, we review the standard techniques for order reduction, recall the recently introduced PMTBR algorithm, and discuss point selection issues, including resampling and variance-based point selection. In Section III, we discuss the computational implementation of the basic ideas. We first outline a rigorous procedure for point selection based on the proposed metric, i.e., the variance between models. Such a procedure demonstrates that the metric proposed can indeed be used for point selection. Unfortunately, the computational cost of the procedure makes it unfeasible for large-scale systems. Therefore, we propose a more efficient heuristic computation for the same task. Then, in Section IV, we show computational experiments to illustrate when the proposed approach is advantageous (and also when it is not) and present some conclusions in Section V.

II. Resampling-Based Error Estimation

A. Multipoint Projection Framework

For simplicity of exposition, consider for the moment the restricted case linear system models in the standard state-space form, i.e.,

$$\frac{dx}{dt} = Ax + Bu, \quad y = Cx + Du \tag{1}$$

with input u and output y, which are described by the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{p\times n}$, and $D \in \mathbb{R}^{p\times p}$. These algorithms take as input a linear system of the form (1) and produce a reduced model, i.e.,

$$\frac{dz}{dt} = \hat{A}z + \hat{B}u, \quad y = \hat{C}z + Du \tag{2}$$

where $\hat{A} \in \mathbb{R}^{q\times q}$, $B \in \mathbb{R}^{q\times p}$, and $C \in \mathbb{R}^{p\times q}$. This is achieved by constructing matrices W and V whose columns span a "useful" subspace and projecting the original equations in the column spaces of W and V as

$$\hat{A} \equiv W^T AV, \quad \hat{B} \equiv W^T B, \quad \hat{C} \equiv CV. \tag{3}$$

The most common choices are based on picking the columns of W, V to span a Krylov subspace [1], [2]. A more sophisticated approach is to construct the projection matrix V from a multipoint Krylov subspace [8]–[10]. Given m complex frequency points s_k, a projection matrix may be constructed whose kth block column is $k = 1, \ldots, m$ for every s_k, i.e.,

$$z_k = (s_k I - A)^{-1} B. \tag{4}$$

Multipoint projection is known to be an efficient reduction algorithm in that the number of columns, which determines the final model size, is usually smaller for a given allowable approximation error when compared to pure moment matching approaches. In this paper, we are addressing a practical question of implementation: How many points s_k should be used, and how should the s_k be chosen?

B. Review of the PMTBR Algorithm

An alternative method of model reduction, i.e., balanced truncation, is based on analysis of the controllability and observability Gramians X and Y, respectively. The Gramians are usually computed from the following Lyapunov equations:

$$AX + XA^T = -BB^T \tag{5}$$

$$A^T Y + YA = -C^T C. \tag{6}$$

Reduction is performed by projection onto the invariant subspaces associated with the dominant eigenvalues of the product of Gramians XY [17], [23].

The PMTBR analysis is based on an alternative frequency-domain expression for the Gramian, e.g.,

$$X = \int_{-\infty}^{\infty} (j\omega I - A)^{-1} BB^T (j\omega I - A)^{-H}\, d\omega \tag{7}$$

for the controllability Gramian. Given a quadrature scheme with nodes ω_k and weights w_k and defining

$$z_k = (j\omega_k I - A)^{-1} B \tag{8}$$

an approximation $\hat{X}$ to X could be computed as

$$\hat{X} = \sum_k w_k z_k z_k^H. \tag{9}$$

Alternatively, let Z be a matrix whose columns are z_k, and W a diagonal matrix with diagonal entries $W_{kk} = \sqrt{w_k}$, and perform the singular value decomposition (SVD) of ZW as

$$ZW = V_Z S_Z U_Z \tag{10}$$

where S_Z is a real diagonal, and V_Z and U_Z are unitary matrices. Clearly, then, $\hat{X} = V_Z S_Z^2 V_Z^T$; thus, in fact, the dominant singular vectors in V_Z, as can be identified from the singular values in S_Z, give the eigenvectors of $\hat{X}$. Therefore, V_Z converges to the eigenspace of X, and the Hankel singular values are obtained directly from the entries of S_Z. V_Z can then be used as the projection matrix in a model-order reduction scheme. The PMTBR method in fact advocates using the projection matrix $V = V_Z$ for model reduction in a congruence fashion; much like is done in [2]. As such, stability as well as passivity of the final model is guaranteed, at least for problems resulting from interconnect or electromagnetic structure analysis formulated using the standard techniques. It has been shown that constructing projection matrices by multipoint frequency sampling, as in (4), followed by an SVD, converges to the TBR algorithm [19]. One of the important features of TBR is an absolute bound on the error of approximation that is related to the singular values of the products of the Gramians. Likewise, the singular values of the matrix ZW can be used in the PMTBR procedure for error control. However, we will not use these quantities here because they are purely *a posteriori* estimates. Instead, we are motivated by another interesting interpretation of the PMTBR algorithm, which can readily be shown by writing

$$X = \int_{-\infty}^{\infty} z(\omega) z^H(\omega) p(\omega) d\omega = E\left\{ z(\omega) z^H(\omega) \right\} \tag{11}$$

where $z(\omega) = (j\omega I - A)^{-1} B$, and $p(\omega)$ can be interpreted as an assumed or known distribution [thus generalizing the integration weights from (9)]. In this interpretation, (7) is the counterpart of (11) assuming uniform distribution over all possible frequency values. In a sense, this corresponds to an uninformed modeling procedure since no distinction is made between frequencies. In other words, direct computation of the Gramians, as performed by TBR, corresponds to a prior assumption of distribution on the inputs that is uninformative (see [24] for a discussion of this interpretation in Bayesian terms and a direct relation to system entropy). The interesting thing about this interpretation is that any general knowledge one might have about the system structure or behavior can be incorporated into $p(\omega)$ to guide selection of the projection matrix. Under this interpretation, any knowledge that is available *a priori* or can be gained from measuring or otherwise directly analyzing the system can be used directly to guide the choice of the projection matrix. Such an approach was, for instance, exploited in [7], [15], and [24], where, for instance, frequency selection or knowledge of input characteristics was used with immediate consequences in model quality and efficiency.

Here, we will reverse the implication. Suppose we take M samples of ω drawn independent identically distributed with density given by $p(\omega)$ (the argument is independent of the precise density). This, in turn, generates M vectors that can be used, via projection and possibly SVD, to generate a reduced-order model of order M or less. As $M \to \infty$, the singular values and vectors of Z will converge to those of X calculated from (11). Now, suppose a "different" set of M samples is drawn with the same density. Under the statistical interpretation, these are all "equally reasonable" order-M reduced models for the state-space system excited by inputs with frequency power density related to $p(\omega)$. Each of the two sets of projection vectors—or for that matter, any size-M subset of the projection vectors—can generate a viable candidate model for consideration. It turns out that such an interpretation leads to an adaptive scheme for sample point selection.

C. Variance Metric

Clearly, multipoint projection relies on effective placement of sample points. Obvious strategies include random selection of points or uniform griding of the domain. Both of these can be shown as uninformed algorithms about either the structure or the behavior of the system, and as such, they may not be optimal. More importantly, for problems in multiple dimensions (e.g., nonlinear or parameter-varying systems), uniform sampling may quickly become too expensive, while random sampling may fail to recognize relevant areas of behavior. Furthermore, neither uniform nor random sampling, akin to Monte Carlo simulation, really addresses the question of where to place new points.

Intuitively, an efficient point selection algorithm should place new points where the error or model uncertainty is large. However, such practical reasoning introduces a "chicken-or-egg" problem: We do not know the model error without resort to evaluating the detailed model, which is expensive. Once a detailed model evaluation has been made, we might as well include the associated vector in the projection subspace. Immediately, we are faced again with the question of where to evaluate the next candidate. Instead, we desire a point selection approach that does not require any reference to the larger original system when evaluating candidate points.[1] In other words, we would like to perform most computation in reduced spaces and not resort to the original space at all.

The first insight into achieving our goal comes from the statistical interpretation of the PMTBR procedure. In this scheme, each matrix Z represents a sample drawn from a distribution that generates the Gramians. Suppose we construct "multiple" matrices $Z_1, \ldots, Z_B$, each of the same finite dimension q. Each of these in turn defines a model with transfer function $H_k(s),\ k = 1, \ldots, B$. As argued above, there is no reason to prefer one over the others, and so, if they do not give similar results, something has clearly gone wrong—some model aspect needs to be improved. We propose that the places where the transfer functions disagree the most are likely to be good[2] places to put a new sample point.

[1] Effectively, we are assuming that the cost of evaluating the detailed model is quite large, i.e., n is considerably larger than q.

[2] Note that error control, in terms of stopping the procedure, can be managed separately, e.g., by the normal singular value tests.

Specifically, if we also define a "nominal" or "mean" projector $Z^{(\mathrm{nom})}$ that is obtained from the column union of Z_k, $Z^{(\mathrm{nom})} = \bigcup_{k=1}^{B} Z_k$, then as a metric of model uncertainty, we propose the variance-like quantity, i.e.,

$$\mathrm{var}(s) = \frac{1}{B}\sum_{1}^{B} \left\| H_k(s) - H_{\mathrm{nom}}(s) \right\|^2 . \tag{12}$$

Note that for large and/or complex starting systems, evaluation of the variance in (12) can be considerably less expensive than evaluation of the full model (assuming that the reduced models themselves can be constructed inexpensively; see Section III-A for this critical step). This is true because all the models involved are reduced-order models of size $\mathcal{O}(q)$ (assuming fixed B). Therefore, evaluation of (12) at multiple points is acceptable if it will lead to a good choice of the next expansion point. However, we have introduced a new difficulty. We must now construct multiple reduced systems, i.e., construct and manipulate a very large number of reduced models that must be related in a statistically meaningful way. *De facto*, the main step boils down to efficient construction and manipulation of multiple projection matrices. A naive implementation would construct the B models independently, but more efficient alternatives exist as we now discuss.

D. Resampling for Variance Estimation

To address this new challenge, we borrow some ideas from what are known as "resampling plans" in the statistics literature. Resampling plans, such as the jackknife and bootstrap [22], construct "replica" data sets by appropriately selecting samples from an original data set. Once the replica data sets are available, robust estimates of statistical estimators can be obtained by considering each of the replica data sets as samples from a parent population. Such techniques are most relevant whenever large numbers of samples are not available or are too expensive to obtain.

The bootstrap is a nonparametric method for estimating the sampling distribution of a statistic. In the bootstrap setting, one generates a replica set $D_{(b)} = \{z_1^*, z_2^*, \ldots, z_M^*\}$ consisting of M data values drawn randomly with replacement from an original data set D_m of size m. $D_{(b)}$ is called a bootstrap sample. Consider performing B repetitions of the $D_{(b)}$ sampling. Suppose we are interested in a generic parameter θ (which could be mean, variance, or some other function of some parameter) and want to compute an estimate of it, i.e., $\hat{\theta}$. We can obtain such an estimate $\hat{\theta}_{(b)}$ from the statistic computed by $D_{(b)},\ b = 1, \ldots, B$. For instance, the bootstrap estimator of the variance of the estimator $\hat{\theta}$ is the variance of the set $\hat{\theta}_{(b)},\ b = 1, \ldots, B$, i.e.,

$$\mathrm{var}_{\mathrm{bs}}(\hat{\theta}) = \mathrm{var}_{\mathrm{bs}}\left(\hat{\theta}_{(b)}\right) = \frac{\sum_{b=1}^{B}\left(\hat{\theta}_{(b)} - \hat{\theta}_{(\bullet)}\right)^2}{B-1}$$

where $\hat{\theta}_{(\bullet)} = 1/B\sum_{b=1}^{B}\hat{\theta}_{(b)}$.

Note that the bootstrap employs sampling with replacement. In statistical terms, a random sample is a set of items that have been drawn from a population in such a way that each time an item was selected, every item in the population had an equal opportunity to appear in the sample. To meet the equal opportunity requirement, it is important that the sampling be done with replacement. That is, each time an item is selected, the relevant measure is taken and recorded. Then, the item must be replaced back into the population before the next item is drawn. If the items are not replaced in the population, each time an item is withdrawn, the probability of being selected for each of the remaining items will have been increased. It is also important to recognize that when sampling with replacement, it is possible for the same item to appear more than once in a sample, and it is possible to draw a random sample that is larger than the population from which it came. Notice also that it is possible to draw as many random samples as we like from a given population. The key idea here is that we either sample with replacement or draw our samples from a population that is so large that the withdrawal of successive items changes probability by an amount that is too small to be of concern.

Now, let us discuss resampling in the context of PMTBR and point selection. Assume we have a set of candidate projection vectors $z_1, \ldots, z_m$ as the m columns of the matrix Z, i.e.,

$$Z = [z_1, \ldots, z_m]$$

and consider, for instance, bootstrapping this data set by constructing B replica sets, each of size M, with replacement, from Z (M could equal m, be larger or smaller). For each of the B replica sets, we draw vectors randomly from Z, which form the columns of our bootstrap sample $Z_{\mathrm{bs}}^{(k)},\ k = 1, \ldots, B$. Each of these bootstrap samples should have approximately the same distribution from which the columns of Z were drawn. We are abusing notation here by referring to draws of the columns of Z since, in reality, the draws are of the frequency variable ω, i.e., draws determined by the density $p(\omega)$, but there is a one-to-one correspondence via (8) between ω_k and z_k. For the kth bootstrap sample, consider the vector $s^{(k)} \in \mathbb{R}^m$ such that the ith entry $s_i^{(k)}$ represents the number of occurrences of the vector z_i in the replica set $Z_{\mathrm{bs}}^{(k)}$. Let $S^{(k)} \in \mathbb{R}^{m\times m}$ be a diagonal matrix whose diagonal is formed by $s^{(k)}$. To form the projector, we construct the matrix for each replica set, i.e.,

$$Z^{(k)} = ZS^{(k)} \tag{13}$$

and proceed to perform an SVD on it, from which the projector of size q $(q \leq M)$ is obtained as described in Section II-B. This step and (13) are easily shown to be akin to the procedure followed in (10). This step is then repeated for each replica step, in each case leading to a different projector $V^{(k)}$. Once the B bootstrap projectors are obtained, one can form reduced models from each of them. The "nominal" model in this case can be constructed, for instance, by summing the individual s vectors and proceeding in the same fashion to obtain its projector. Given the projectors, it is a trivial matter to obtain the corresponding reduced-order models, which we denote with the stencil $[A_k, B_k, C_k]$.

III. Implementation

In this section, we consider the computational implementation of the previously proposed scheme.

A. Efficient Construction of Projectors

The efficient reuse of previously computed data (i.e., caching) is critical to a fast implementation of this algorithm. Explicit construction of the replica Z matrices, followed by SVD, followed by projection for each, would defeat our goal of having an estimator that does not require reference to the original system. The simplest procedure is to drop the SVD step and use the Z matrices directly for projection when constructing the estimators. The replica models can be constructed by computing and saving $z_i^T A z_j$. To illustrate, consider a concrete example. Suppose that three candidate vectors z_1, z_2, and z_3 have been computed. The reduced $\hat{A}$ is

$$\hat{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} z_1^T A z_1 & z_1^T A z_2 & z_1^T A z_3 \\ z_2^T A z_1 & z_2^T A z_2 & z_2^T A z_3 \\ z_3^T A z_1 & z_3^T A z_2 & z_3^T A z_3 \end{bmatrix}.$$

Consider a two-sample bootstrap replica (z_2, z_1). The needed reduced model is

$$\hat{A}_2 = \begin{bmatrix} A_{22} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}$$

which is trivially constructed from $\hat{A}$. Now, consider a four-sample replica (z_1, z_3, z_1, z_2). This is equivalent to $(2z_1, z_2, z_3)$. The required model is

$$\hat{A}_4 = \begin{bmatrix} 4A_{11} & 2A_{12} & 2A_{13} \\ 2A_{21} & A_{22} & A_{23} \\ 2A_{31} & A_{32} & A_{33} \end{bmatrix}.$$

A more sophisticated strategy involves reconstruction of the SVDs in the reduced space. To see this, note that if we have computed the SVD of Z as $Z = U\Sigma V^T = UR$, where R is an $m \times m$ matrix, then the SVD of ZS can be computed if we know the SVD of RS, which is another $m \times m$ matrix. When combined with the previously mentioned caching scheme for the projected matrices, all operations can be done in the reduced space. In the second example above, R will be a 3×3 matrix, and S will be a diagonal matrix with entries $(2, 1, 1)$. If we perform the SVD of RS, $RS = U_4 \Sigma_4 V_4^T$, the reduced system matrix is given by

$$\hat{A}_{4,\mathrm{SVD}} = U_4^T \hat{A} U_4.$$

It should be clear that the construction of the reduced $\hat{B}, \hat{C}$ is precisely similar. Note that the matrix multiplying dz/dt is not necessarily the identity if the congruence transform is not unitary.

B. Linear Matrix Inequality (LMI)-Based Approach

Next, we need to find the place where the variance is maximized over the replica models. It turns out that we can pose the problem of optimal point selection as optimization problem with LMI constraints by constructing a specific composite linear system whose norm will give the "variance." In the following, we outline a procedure to perform this computation. Existence of such a procedure shows that the proposed metric can be effectively computed without resorting to any reference to the original large system.

Let A, B, and C denote the system matrices for the current best model (assume $D = 0$ for now). Let A_k, B_k, and C_k denote the system matrices for the kth bootstrap replica. With the bootstrap replica transfer function $H_k(s) = C_k(sI - A_k)^{-1}B_k$, we interpret $H(s) = C(sI - A)^{-1}B$ as the "mean" model and then compute the variance as in (12). Now, construct

$$\bar{A}_k = \begin{bmatrix} A & \\ & A_k \end{bmatrix}, \quad \bar{B}_k = \begin{bmatrix} B \\ B_k \end{bmatrix}, \quad \bar{C}_k^T = \begin{bmatrix} C^T \\ -C_k^T \end{bmatrix}$$

and using all B bootstrap replicas

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & & & \\ & \bar{A}_2 & & \\ & & \ddots & \\ & & & \bar{A}_B \end{bmatrix}, \ \bar{B} = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \\ \vdots \\ \bar{B}_B \end{bmatrix}, \ \bar{C}^T = \begin{bmatrix} \bar{C}_1^T \\ \bar{C}_2^T \\ \vdots \\ \bar{C}_B^T \end{bmatrix}.$$

Then, $\|\bar{C}(sI - \bar{A})^{-1}\bar{B}\|^2$ is the required variance $\mathrm{var}(s)$. The objective of point selection is to find the frequency point $s_{\max}$ at which $\mathrm{var}(s)$ is maximized. This is a standard norm computation problem, which is solved by posing the norm constraint as an LMI [25], with the norm (which in our case we interpret as variance) and the point $s_{\max}$ found by solving a sequence of LMIs in a bisection procedure. Thus, in principle, we can guarantee that the quantities for the variance-based search can be computed in polynomial time. Unfortunately, the procedure outline leads to very large LMIs, which are expensive to solve in practice. This obviously limits the practical application of this technique. Nevertheless, the general framework is still valid. Furthermore, again in principle, it might be possible to exploit knowledge about the particular structure of our system, which is quite peculiar, to obtain significant complexity improvements. A technique for exploiting such knowledge was previously presented in [26] for passivity enforcing in LMI-based system identification. Instead of pursuing that approach, however, we note that the point search is only an adjunct to the reduced-model computation. As such, it is not necessary that the search be exact. An approximate point found at lower cost might be more desirable and certainly more efficient to determine. With this goal in mind, in the following section, we propose a heuristic procedure for determining a sampling point with high variance value.

C. Heuristic Search

In light of the comments in the preceding section, we have experimented with a simple heuristic stochastic search procedure. To begin the procedure, we generate at random an initial set of candidate search points, with enough points being chosen to cover the search space at very coarse resolution (e.g., in the experiments later, we have used around 20 points). At each

iteration, the variance is evaluated on the search points, and the point with maximal value of the variance is chosen as the next sample. To repopulate the search point population, we drop some fraction of the "worst" points at each iteration—for example, we may drop one third of points with the smallest variance. These points represent places that are presumably the best approximated. Since the system is already "reliably understood" at those points, it is unlikely that placing a sample there in the future is going to be a good expenditure of effort. We will discard such points and look somewhere else. Conversely, points with large variance are more likely to represent regions that are harder to approximate. We may wish to remember these points and consider them as possible future samples. The cost of this procedure is acceptable since determining the value of the metric at those points can be done using only systems of reduced size, as described by (12). Once the point with the maximal variance is computed, then we refer to the original system and include the new point in the projection scheme. The whole procedure is then repeated to determine the next sample point for projection. If enough points are already available, or the variance in all search points is considered acceptable, the procedure is terminated.

IV. Computational Experiments

In this section, we present computational experiments designed to illustrate various properties of the resampling procedures for point selection. As test cases, we took three RLC circuits that originate from partial element equivalent circuit (PEEC)-based analysis of actual three-dimensional (3-D) IC packages. The first example refers to a printed circuit board structure from Teradyne, Inc., whereas the second example refers to the model of a subset of a backplane connector. Finally, the third example is the lumped-element equivalent circuit for a 3-D problem modeled via PEEC. All three examples have previously appeared in the literature [1], [27], [28]. The details of the modeling procedure are not very relevant for our purposes here; thus, for the most part, we will omit them and refer to the examples as PEEC-A, PEEC-B, and PEEC-C.

A. Fidelity of Metric

First, we illustrate the behavior of the model construction process as the sample points are selected via the bootstrap procedure. In this experiment, the variance-versus-frequency curves were searched exhaustively[3] for the maximum variance point at which to take a new sample. For each sample, a circuit solved at a single (complex) frequency is performed, generating two new projection vectors. Fig. 1 shows snapshots for four (increasing) model orders (specifically, after 10, 14, 21, and 24 samples, with the first ten starting samples selected randomly), e.g., PEEC-A. As model construction proceeds, the model achieves closer and closer match to the original system. Likewise, the degree of variation of the bootstrap samples decreases as construction proceeds. Particularly, in the third plot, we see the signature characteristic of this method—areas where the error is large are also areas of uncertainty; thus, there is "instability" under bootstrapping, which results to large variance.

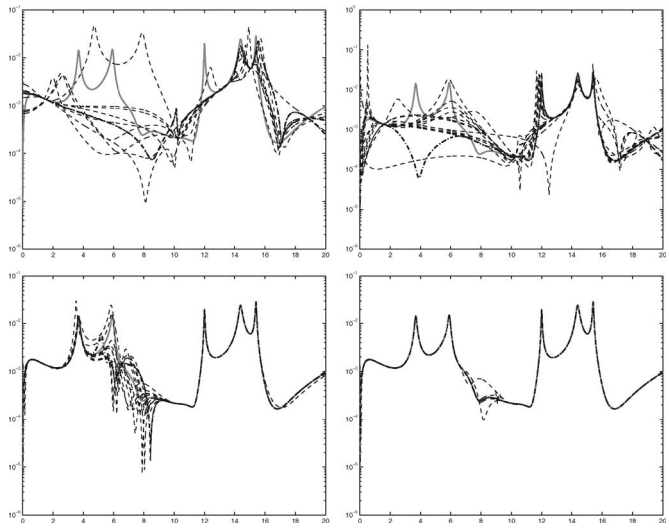

Fig. 1. Example PEEC-A. Snapshots of transfer function (solid), reduced-model approximations (dash), and bootstrap samples (light dash) for four progressive refinements of model. Note the large variance in bootstrap replicas near regions of large model error.

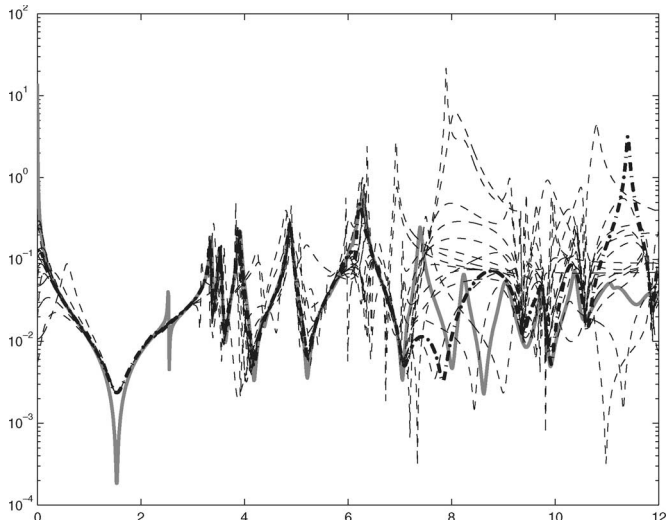

Fig. 2. Example PEEC-B. Snapshot of transfer function (solid), current reduced model (dash-dot), and bootstrap replicas (light dashed) for one step in adaptive approximation procedure.

Fig. 2 shows a single-model snapshot (e.g., PEEC-B) in the very early stage of the model construction process. Note that this example has much sharper resonances than example PEEC-A. A similar result is obtained with example PEEC-C, which has even sharper resonances (see Fig. 8).

Next, we investigate explicitly the point selection behavior of the bootstrapping variance estimation approach. One way to do this is to compare the point picked by the variance metric with the error in the current model. Fig. 3 shows four consecutive sample point selections from a run on model PEEC-A. (Note well that these are not the same scenarios as previously

[3]These would be the curves produced by an LMI-based search procedure.

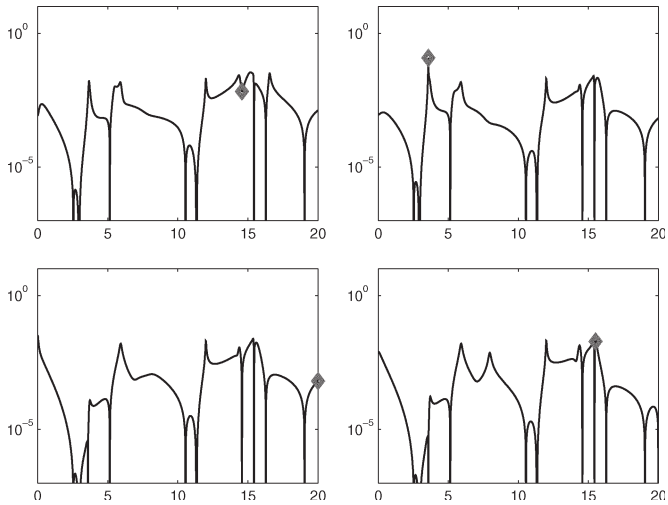

Fig. 3. Example PEEC-A. Transfer function error (line) and four consecutive sampling point selections (diamond).

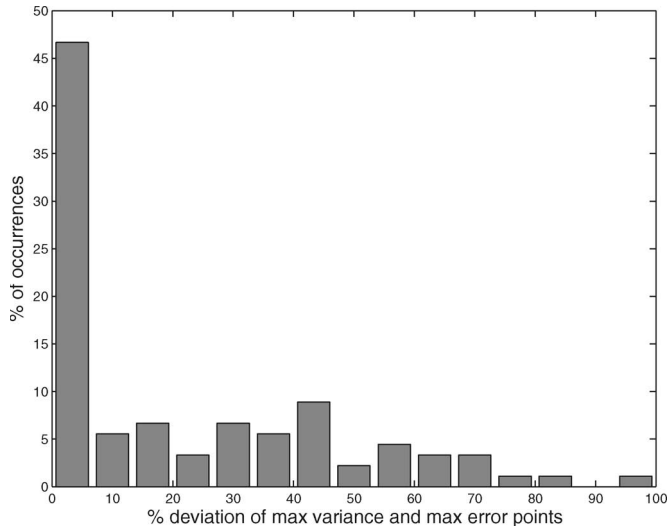

Fig. 4. Relation of variance-based to error-based sampling.

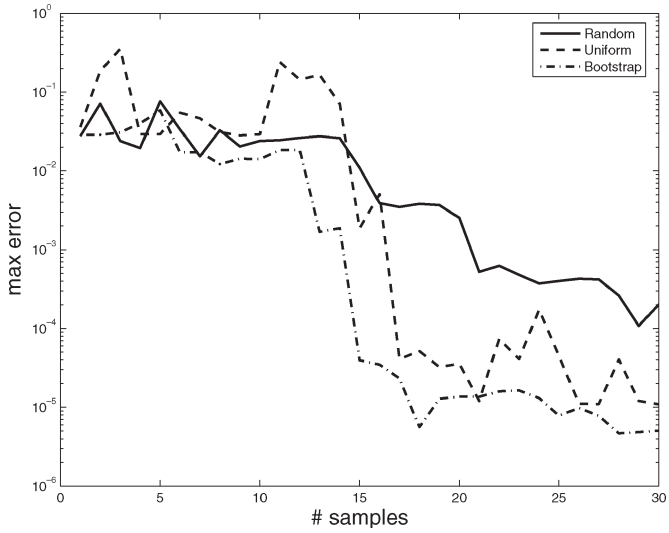

Fig. 5. Example PEEC-A. Comparison of sample selection methods.

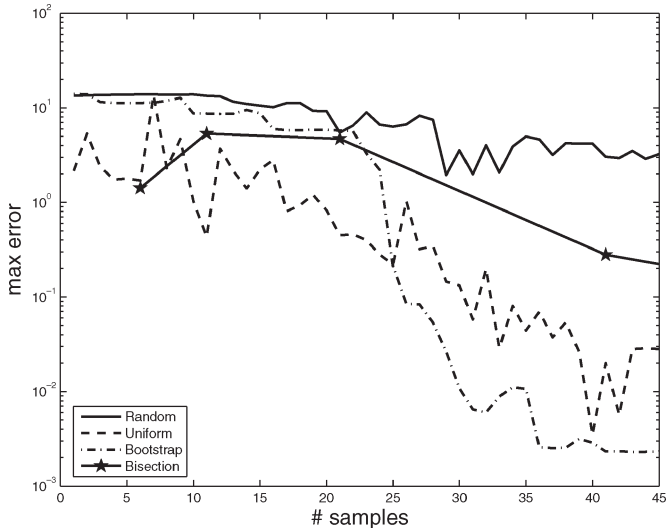

Fig. 6. Example PEEC-B. Comparison of sample selection methods.

discussed.) In our experience, these scenarios are very representative; in two cases, the maximal variance metric picked the point that also had maximal error, in one case, it picked a nearby point, and in one case, it picked a point that was in fact fairly far away from the maximal error point but whose error was not hugely smaller and that was in fact a local maximum in the error. Fig. 4 shows a histogram of the percentage deviation of points picked by the maximal variance when compared with points with maximal error. Note that the correspondence is not perfect—maximal variance is not the same as maximal error after all—but it is fairly good. This answers in the affirmative one of the main questions of this paper. The variance metric can be a useful proxy for error control in point placement. While we have not conducted systematic studies, we have also observed that the variance metric seems to pick points near poles or resonances in the transfer function, which is known to be a good point selection heuristic. While not shown here, we have been able to adapt the algorithm to select based on either absolute or relative error metrics.

B. *Efficiency of Sampling*

Next, we investigate the efficiency of the sampling procedures themselves. Figs. 5 and 6 show comparisons of the bootstrap procedure to random sampling, uniform sampling, and interval bisection. It turns out that the different properties of the two examples have an interesting interaction with the point selection. On example PEEC-A, the bootstrap procedure is disappointing—it performs somewhat, but not hugely, better than random sampling. It is about as good as the (much simpler) uniform sampling procedure. Again, of course, we wish to emphasize that uniform sampling cannot be performed adaptively and is not workable in multiple dimensions[4]; thus, this is not really a fair comparison. On example PEEC-B, as shown in Fig. 6, however, the bootstrap performs significantly better than random sampling and in fact slightly better than uniform sampling. We have also shown interval bisection for

[4]Multiple-dimensional sampling occurs, for example, in parametric model-order reduction problems, [7], [16].

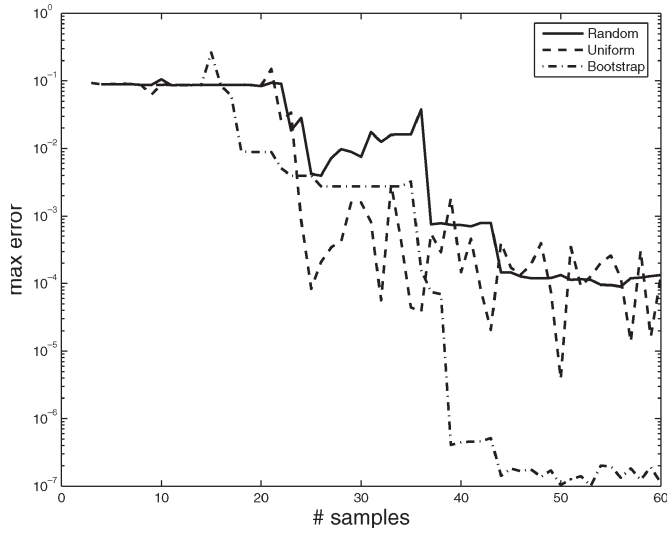

Fig. 7. Example PEEC-C. Comparison of sample selection methods.

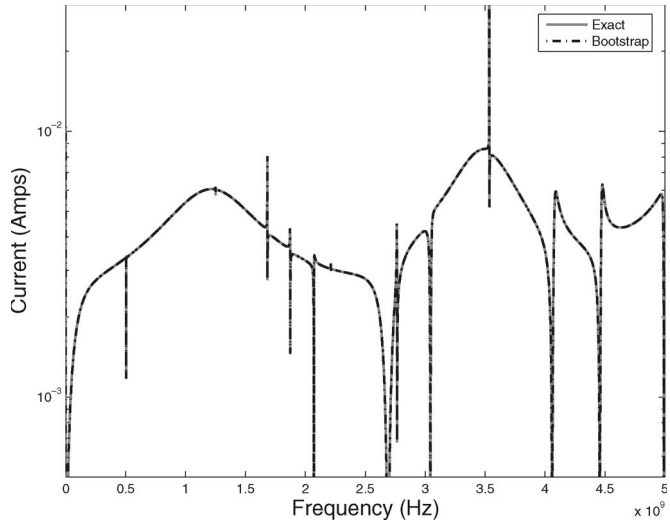

Fig. 8. Example PEEC-C. Exact transfer function and bootstrap generated model after 50 samples. The plots are virtually indistinguishable.

this example to illustrate, while unlike uniform sampling, it can be done adaptively, the coarseness in the number of possible sample points due to doubling the number at each stage limits its usefulness. The different behavior has to do with the properties of the examples. As can be shown in Figs. 1 and 2, example PEEC-B has sharp localized resonant features. An adaptive procedure is much better at selecting for these features. On the other hand, example PEEC-A has fairly smooth features. On this example, because of the global nature of the features, almost any selection scheme works very well. Fig. 7, which is the equivalent plot on example PEEC-C, further stresses this point. As previously mentioned, this example has extremely sharp resonances that are very hard to capture accurately with simplified point selection schemes. This can be shown in Fig. 8, where the exact transfer function and a model with 50 samples picked by the bootstrap procedure are plotted. In this case, the simplified selection schemes are unable to pick all the resonances and still show larger errors even

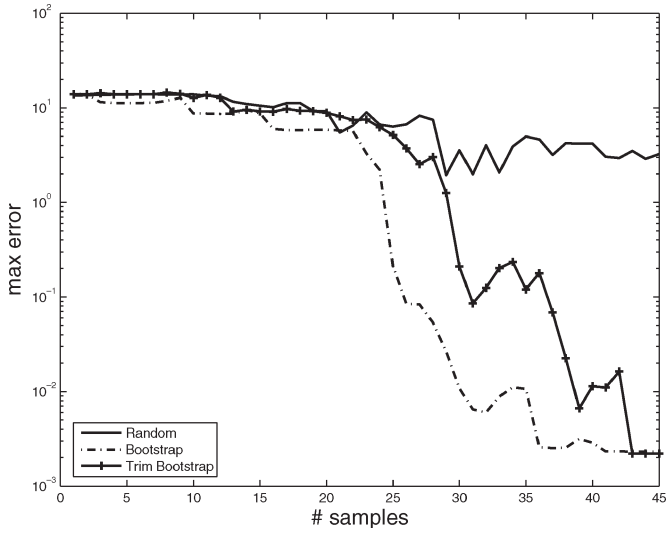

Fig. 9. Example PEEC-B. Comparison of exhaustive versus heuristic bootstrap sample searches.

after a large number of samples was picked. The bootstrap procedure, however, does a much better job of selecting appropriate points.

C. Comparisons of Schemes

Finally, we show some results from the simple heuristic search algorithm. Fig. 9 shows the results on example PEEC-B. We see some degradation in efficiency compared to exhaustive search, as is expected. The important result is that "trim" bootstrap procedure is conclusively superior to the competing random search strategy.

V. Conclusion

In this paper, we presented an approach to sample point selection in model-order reduction that is based on a combination of the statistical interpretation in [19], a variance-based metric, and a bootstrap-motivated procedure for computing the variance. The goal of this paper was twofold: on one hand, determining if the proposed metric can be effective in selecting sample points and, on the other hand, when the metric is effective in this sense, determining the conditions under which substantial improvements in the model reduction efficiency can be achieved.

First, we have shown that the variance metric, whose approximate computation involves only reduced systems, indeed turns out to be an effective strategy for point selection, in particular that it is reasonably correlated with the model error. This is a result of potentially practical significance in reducing very large systems because it means that adaptive error control can be done, at least in some circumstances, without having to resort to evaluating the larger system.

Second, it appears that it may be the case that a supposedly better point selection strategy often does not give significant overall benefits in the reduction procedure. While this is practically disappointing, it is theoretically interesting, further indicating that projection-based reduction is a powerful and robust

modeling technique. It may help to explain why the technique in [7] obtained good results using extremely dumb point selection strategies. On the other hand, we have shown that on at least some more difficult problems, the bootstrap-based procedure can be more robust than competing strategies; thus, when a new problem is being investigated, it may present itself as an interesting alternative.

Future work will involve extending this search strategy to function in multiple dimensions as is needed for construction of nonlinear and parametric models [7], [16]. In such situations, simplified selection schemes are virtually useless, and alternative procedures must be sought.

REFERENCES

[1] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Padé approximation via the Lanczos process," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, 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. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 17, no. 8, pp. 645–654, Aug. 1998.

[3] J. R. Phillips, "Projection-based approaches for model reduction of weakly nonlinear, time-varying systems," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 22, no. 2, pp. 171–187, Feb. 2003.

[4] N. Dong and J. Roychowdhury, "Piecewise polynomial nonlinear model reduction," in *Proc. 40th ACM/IEEE Des. Autom. Conf.*, Anaheim, CA, Jun. 2003, pp. 484–489.

[5] P. Li and L. T. Pileggi, "NORM: Compact model order reduction of weakly nonlinear systems," in *Proc. 40th ACM/IEEE Des. Autom. Conf.*, Anaheim, CA, Jun. 2003, pp. 472–477.

[6] Y. Liu, L. T. Pileggi, and A. J. Strojwas, "Model order reduction of RC(L) interconnect including variational analysis," in *Proc. 36th ACM/IEEE Des. Autom. Conf.*, Jun. 1999, pp. 201–206.

[7] J. Phillips, "Variational interconnect analysis via PMTBR," in *Proc. Int. Conf. Comput.-Aided Des.*, San Jose, CA, Nov. 2004, pp. 872–879.

[8] I. M. Elfadel and D. L. Ling, "A block rational Arnoldi algorithm for multipoint passive model-order reduction of multiport RLC networks," in *Proc. Int. Conf. Comput.-Aided Des.*, San Jose, CA, Nov. 1997, pp. 66–71.

[9] E. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, Coordinated-Sci. Lab., Univ. Illinois, Urbana-Champaign, IL, 1997.

[10] J.-R. Li, F. Wang, and J. White, "An efficient Lyapunov equation-based approach for generating reduced-order models of interconnect," in *Proc. 36th ACM/IEEE Des. Autom. Conf.*, New Orleans, LA, Jun. 1999, pp. 1–6.

[11] Z. Bai, R. Slone, W. Smith, and Q. Ye, "Error bound for reduced system model by Padé approximation via the Lanczos process," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 18, no. 2, pp. 133–141, Feb. 1999.

[12] A. Odabasioglu, M. Celik, and L. T. Pileggi, "Practical considerations for passive reduction of RLC circuits," in *Proc. Int. Conf. Comput.-Aided Des.*, San Jose, CA, Nov. 1999, pp. 214–219.

[13] G. Shi and C.-J. R. Shi, "Model-order reduction by dominant subspace projection: Error bound, subspace computation, and circuit applications," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 52, no. 5, pp. 975–993, May 2005.

[14] E. Chiprout and M. S. Nakhla, "Analysis of interconnect networks using complex frequency hopping (CFH)," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 14, no. 2, pp. 186–200, Feb. 1995.

[15] L. M. Silveira and J. Phillips, "Exploiting input information in a model reduction algorithm for massively coupled parasitic networks," in *Proc. 41st ACM/IEEE Des. Autom. Conf.*, San Diego, CA, Jun. 2004, pp. 385–388.

[16] L. Daniel, O. C. Siong, L. S. Chay, K. H. Lee, and J. K. White, "A multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 23, no. 5, pp. 678–693, May 2004.

[17] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. AC-26, no. 1, pp. 17–32, Feb. 1981.

[18] I. M. Jaimoukha and E. M. Kasenally, "Krylov subspace methods for solving large Lyapunov equations," *SIAM J. Numer. Anal.*, vol. 31, no. 1, pp. 227–251, Feb. 1994.

[19] J. R. Phillips and L. M. Silveira, "Poor Man's TBR: A simple model reduction scheme," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 24, no. 1, pp. 43–55, Jan. 2005.

[20] M. H. Quenouille, "Approximate tests of correlation in time series," *J. R. Stat. Soc., Ser. B*, vol. 11, no. 1, pp. 68–84, 1949.

[21] R. A. Fisher, *Statistical Methods and Scientific Inference*. Edinburgh, U.K.: Oliver and Boyd, 1956.

[22] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.

[23] M. G. Safonov and R. Y. Chiang, "A Schur method for balanced truncation model reduction," *IEEE Trans. Autom. Control*, vol. 34, no. 7, pp. 729–733, Jul. 1989.

[24] J. R. Phillips and L. M. Silveira, "Poor man's TBR: A simple model reduction scheme," in *Proc. DATE Exhib. Conf.*, Paris, France, Feb. 2004, vol. 2, pp. 938–943.

[25] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, ser. Studies in Applied Mathematics, vol. 15. Philadelphia, PA: SIAM, Jun. 1994.

[26] C. P. Coelho, J. R. Phillips, and L. M. Silveira, "A convex programming approach for generating guaranteed passive approximations for tabulated-frequency data," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 23, no. 2, pp. 293–301, Feb. 2004.

[27] M. Kamon, N. Marques, and J. White, "FastPep: A fast parasitic extraction program for complex three-dimensional geometries," in *Proc. Int. Conf. Comput.-Aided Des.*, San Jose, CA, Nov. 1997, pp. 456–460.

[28] N. Marques, M. Kamon, J. White, and L. M. Silveira, "An efficient algorithm for fast parasitic extraction and passive order reduction of 3-D interconnect models," in *Proc. DATE Exhib. Conf.*, Paris, France, Feb. 1998, pp. 538–543.

Luis Miguel Silveira (S'85–M'95–SM'00) was born in Lisbon, Portugal. He received the Engineer's (*summa cum laude*) and Master's degrees in electrical and computer engineering from the Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal, in 1986 and 1989, respectively, and the M.S., E.E., and Ph.D. degrees from Massachusetts Institute of Technology, Cambridge, in 1990, 1991, and 1994, respectively.

He is currently a Full Professor of Electrical and Computer Engineering with the Instituto Superior Técnico, Technical University of Lisbon, a Senior Researcher with the Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC ID-Systems and Computer Engineering Institute R&D), and a Founding Member of the Lisbon Center of the Cadence Laboratories. His research interests include various aspects of computer-aided design of integrated circuits, with emphasis on interconnect modeling and simulation, parallel computer algorithms, and the theoretical and practical issues concerning numerical simulation methods for circuit design problems.

Dr. Silveira has received notable paper citations at various conferences, including ECTC, ICCAD, DAC, DATE, and SBCCI. He is a member of Sigma Xi.

Joel R. Phillips (S'91–M'97–SM'05) received the S.B, M.S., and Ph.D. degrees from Massachusetts Institute of Technology, Cambridge, MA, USA, in 1991, 1993, and 1997, respectively.

He joined Cadence Design Systems in 1997 and is currently a Research Scientist with Cadence Berkeley Laboratories, Berkeley, CA. His research interests include the application of numerical simulation techniques to problems in electronic design automation.

Dr. Phillips received notable paper citations at various conferences, including DAC 2000, DAC 2002, SBCCI 2002, and the 1995 Copper Mountain Conference on Multigrid Methods.