Error Bound for Reduced System Model by Padé Approximation via the Lanczos Process

Zhaojun Bai, Rodney D. Slone, *Student Member, IEEE*, William T. Smith, *Member, IEEE*, and Qiang Ye

***Abstract*— Recently, there has been a great deal of interest in using the Padé Via Lanczos (PVL) technique to analyze the transfer functions and impulse responses of large-scale linear circuits. In this paper, a matrix-based derivation of the error between the original circuit transfer function and the reduced-order transfer function generated using the PVL technique is presented. This error measure may be used for the development of an automated termination of the Lanczos process in the PVL technique and achieve the desired accuracy of the approximate transfer function. PVL coupled with such an error bound will be referred to as the PVL-WEB algorithm.**

***Index Terms*—Asymptotic waveform evaluation (AWE), Lanczos, Padé, Padé Via Lanczos (PVL), Padé Via Lanczos with error bound (PVL-WEB).**

I. Introduction

THE extreme complexity and high density of printed circuit boards and multichip module layouts continue to drive the need for improved circuit-analysis techniques and efficient solution algorithms. Due to the inhomogeneity of the circuit layouts, the solution algorithms must be capable of handling the large systems of equations necessary to model these types of interconnect devices. In addition, the high clock speeds coupled with subnanosecond signal transients impose the need for wide-band solutions that can range in frequency from dc to several gigahertz.

The need for accurate yet efficient solution algorithms served as part of the motivation for the development of asymptotic waveform evaluation (AWE) [18]. The AWE formulation provides straightforward efficient circuit analysis in either the time or frequency domains. AWE uses moment matching to approximate the transfer function of a large system with a lower order approximate transfer function. The response of the transfer function for a circuit depends on the system poles and residues. AWE extracts the dominant poles and residues using Padé approximations [2] and provides an accurate estimation of the system response. AWE has been used to solve networks with resistance–capacitance trees, lumped elements, lossy coupled transmission lines with quasi-transverse electromagnetic mode propagation, partial element equivalent circuit (PEEC) networks, nonlinear terminations, and networks with frequency-dependent parameters [3], [6], [7], [13], [16], [18], [21].

Using the AWE formulation, the approximate transfer function will be most accurate in a neighborhood near the point of expansion. Maclaurin series moments are generated when expanding at $s = 0$. Therefore, the approximate system response obtained using the AWE computed poles and residues is most accurate near $s = 0$. Decreasing accuracy occurs for poles located at frequencies far removed from the expansion point. The problem is overcome by computing the poles and residues at multiple expansion points in the complex s plane. This technique is known as complex frequency hopping (CFH) [5], [6]. The CFH algorithm is somewhat difficult to automate and often requires user supervision to ensure accuracy.

Recently, the Padé Via Lanczos (PVL) algorithm was applied to circuit-analysis problems and was shown to be capable of producing accurate approximations of the circuit transfer functions, impulse responses, and pole predictions over a broad frequency range [8], [10]. In [8], the PVL algorithm was applied to lumped-type circuit modeling including PEEC models [19]. In [17], an adaptive block Lanczos algorithm was applied for solution of multiconductor transmission line (MTL) problems. In that work, a least square fitting procedure with frequency partitioning was used to obtain high-order approximations of the MTL parameters [17]. In a more recent work [4], the PVL algorithm was applied to MTL problems by using Chebyshev polynomials to represent the spatial variation of the transmission-line voltages and currents.

In all of the above works using the Lanczos process, the results demonstrated that the algorithms produce efficient and accurate approximations for the transfer functions and impulse responses of high-speed interconnect problems. The accuracy was demonstrated by comparing various orders of the PVL solutions with other accurate but much less efficient solutions. However, the existing approach cannot provide an unsupervised measure of the error for a given order of Padé approximation and, therefore, termination of the Lanczos procedure is largely heuristic. One commonly used convergence

Manuscript received July 30, 1997; revised June 11, 1998. This work was supported by the National Science Foundation (NSF) under Grant DMS-9 508 543. The work of Z. Bai was supported in part by the NSF under Grant ASC-9 313 958 and in part by the Department of Energy under Grant DE-FG03-94ER25219 via subcontracts from the University of California at Berkeley. The work of R. D. Sloane was supported by fellowships from the University of Kentucky. This paper was recommended by Associate Editor D. Ling.

Z. Bai is with the Department of Mathematics, University of Kentucky, Lexington, KY 40506 USA (e-mail: bai@ms.uky.edu).

R. D. Slone was with the Department of Electrical Engineering, University of Kentucky, Lexington, KY 40506 USA. He is now with the ElectroScience Laboratory, Department of Electrical Engineering, The Ohio State University, Columbus, OH 43212 USA (e-mail: rdslon01@ieee.org).

W. T. Smith is with the Department of Electrical Engineering, University of Kentucky, Lexington, KY 40506-0046 USA (e-mail: bsmith@engr.uky.edu).

Q. Ye is with the Department of Applied Mathematics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2 Canada (e-mail: ye@gauss.amath.umanitoba.ca).

Publisher Item Identifier S 0278-0070(99)01011-8.

criteria is to test the difference between successive orders of approximation until the difference becomes small. This does not, however, imply a small error between the original and approximate impulse responses.

In this paper, a matrix-based derivation of the error between the original circuit transfer function and the reduced-order transfer function generated using PVL is presented. Model error estimation was also discussed in [12]. However, the approaches used in [12] are computationally more expensive than the approach presented below. The error measure derived in this paper reveals the intrinsic properties of convergence for the PVL algorithm and allows for development of an automated termination of the Lanczos process in the PVL technique to ensure the desired accuracy of the approximate transfer function. The PVL coupled with this error bound will be called the Padé Via Lanczos with error bound (PVL-WEB) algorithm.

The rest of this paper is organized as follows. In Section II, the PVL-WEB algorithm is formulated. The section begins with a review of the quantities associated with the Lanczos process. A new matrix-based derivation of the error between the original transfer function and the reduced transfer function resulting from PVL is the focus of this paper and completes Section II. The matrix-based derivation provides for a computationally inexpensive error bound. Implementing an automated stopping criterion in PVL using the error bound is discussed in Section III. Numerical examples of PVL-WEB are presented in Section IV. A summary is given in Section V.

II. PVL WITH ERROR BOUND

In this section, a new matrix-based derivation of the PVL technique is discussed. This derivation is not only able to show the order of the approximation of the PVL algorithm but also gives a computable error bound for the approximation. This error bound can be used as an automated stopping criterion for the PVL iterations.

A. System Matrices and the PVL Model Reduction

Modified nodal analysis matrices [15] are commonly used with the PVL algorithm. These matrices can be derived from the state variable equations of the system [8], or from other methods such as shown in [4]. In either case, the transfer function is given as

$$H(s) = l^H(G + sC)^{-1}b$$

where $s = j2\pi f$ for the frequency f, l^H is an N-vector that selects the output of interest, b is the excitation N-vector of the network, and C and G are N by N matrices that represent the contribution of memory and memoryless elements. Next set

$$s = s_0 + \sigma, \quad A = -(G + s_0C)^{-1}C, \quad r = (G + s_0C)^{-1}b$$

where s_0 is the point of expansion in the s plane. Then the transfer function $H(s)$ can be rewritten as

$$H(s_0 + \sigma) = l^H(I - \sigma A)^{-1}r. \tag{1}$$

The PVL algorithm uses the iterative Lanczos process to reduce the matrix A to a tridiagonal matrix. The poles and residues of a Padé approximant of the transfer function $H(s)$ can be accurately and efficiently computed from an eigendecomposition of the tridiagonal matrix [8]. In [8], the governing matrix equations are given that define the predominant version of the Lanczos process used in the electrical engineering community. An equivalent formulation that explicitly shows the complex nature of the equations is now presented. These equations are valid for $n = 1, 2, \cdots, q$, where q is the final step in the iteration. The matrix equations are

$$AV_n = V_nT_n + v_{n+1}\rho_{n+1}e_n^T, \tag{2}$$

$$A^HW_n = W_n\tilde{T}_n + w_{n+1}\eta_{n+1}e_n^T \tag{3}$$

where T_n and $\tilde{T}_n$ are the tridiagonal matrices

$$T_n = \begin{bmatrix} \alpha_1 & \beta_2 & & \\ \rho_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_n \\ & & \rho_n & \alpha_n \end{bmatrix}, \quad \tilde{T}_n = \begin{bmatrix} \alpha_1^* & \gamma_2 & & \\ \eta_2 & \alpha_2^* & \ddots & \\ & \ddots & \ddots & \gamma_n \\ & & \eta_n & \alpha_n^* \end{bmatrix}.$$

$V_n = [v_1 \ v_2 \ \cdots \ v_n]$ and $W_n = [w_1 \ w_2 \ \cdots \ w_n]$ are matrices of the Lanczos vectors $\{v_i\}$ and $\{w_i\}$, which satisfy the biorthogonality properties

$$W_n^HV_n = D_n = \text{diag}(\delta_1, \delta_2, \cdots, \delta_n) \tag{4}$$

$W_n^Hv_{n+1} = V_n^Hw_{n+1} = 0$ and have unit vector length $\|v_n\|_2 = \|w_n\|_2 = 1$. Furthermore, T_n and $\tilde{T}_n^H$ are related by

$$\tilde{T}_n^H = D_nT_nD_n^{-1}. \tag{5}$$

Using these equations, one derives the following algorithm to compute all required quantities.

Algorithm 1: q steps of the Lanczos process.

$\rho_1 \leftarrow \|r\|_2, \eta_1 \leftarrow \|l\|_2, w_0 \leftarrow 0, v_0 \leftarrow 0, \delta_0 \leftarrow 1, n \leftarrow 0.$
$v_1 \leftarrow r(\rho_1)^{-1}, w_1 \leftarrow l(\eta_1)^{-1}.$
for $n = 1, 2, \cdots, q$ do
  $f \leftarrow Av_n.$
  $\delta_n \leftarrow w_n^Hv_n, \quad \alpha_n \leftarrow w_n^Hf(\delta_n)^{-1}.$
  $\beta_n \leftarrow \delta_n\eta_n^*/\delta_{n-1}, \quad \gamma_n \leftarrow (\delta_n\rho_n/\delta_{n-1})^*.$
  $f \leftarrow f - v_n\alpha_n - v_{n-1}\beta_n, \quad \rho_{n+1} \leftarrow \|f\|_2.$
  $v_{n+1} \leftarrow f(\rho_{n+1})^{-1}.$
  $f \leftarrow A^Hw_n - w_n\alpha_n^* - w_{n-1}\gamma_n, \quad \eta_{n+1} \leftarrow \|f\|_2.$
  $w_{n+1} \leftarrow f(\eta_{n+1})^{-1}.$
endfor

The transfer function $H(s_0 + \sigma)$ in (1) can be expressed in terms of Lanczos process quantities. Use (1) and (4) and the initial step of Algorithm 1 to obtain

$$r = v_1\rho_1, \quad l^H = w_1^H\eta_1^*, \quad l^Hr = \rho_1\eta_1^*\delta_1$$

and

$$\begin{aligned} H(s_0 + \sigma) &= l^H(I - \sigma A)^{-1}r \\ &= \eta_1^*\rho_1w_1^H(I - \sigma A)^{-1}v_1 \\ &= (l^Hr)(\delta_1)^{-1}w_1^H(I - \sigma A)^{-1}v_1. \end{aligned} \tag{6}$$

The nth-order reduced transfer function is defined to be (following [8])

$$H_n(s_0 + \sigma) = (l^Hr)e_1^T(I - \sigma T_n)^{-1}e_1. \tag{7}$$

In [8], the derivation of the approximation for PVL is linked to the moment-matching techniques of AWE, which are connected with Padé approximation theory. It is shown that the order of approximation of the reduced transfer function $H_n(s_0+\sigma)$ to the transfer function $H(s_0+\sigma)$ is $2n-1$. In the PVL algorithm, it is straightforward to see the approximation up to order $2n-2$. However, the proof for the approximation of order $2n-1$ is rather involved [11]. Furthermore, this order of approximation does not indicate the exact error between the reduced and the original system transfer functions.

B. Matrix-Based Derivation of the Order of Approximation of $H_n(s_0+\sigma)$

In this section, a new matrix-based derivation of the PVL method is presented. The most important outcome, however, is not the new derivation itself but rather an expression of the error of the PVL method. From this expression, within a certain frequency range, one can derive a computationally inexpensive error bound on the approximate transfer function, which then can be used as a stopping criterion for the Lanczos iteration. The initial idea of the matrix-based derivation of the PVL algorithm presented in this paper was developed at the same time as the work of [1]. In this paper, the Lanczos governing equations (2)–(5), which are more commonly found in the electronic circuit simulation community, are used. The focus here is on developing a practical error estimator and the associated implementation details. On the other hand, the work of [1] uses a different set of Lanczos governing equations and focuses on the mathematical aspects of error estimation and convergence analysis of the PVL algorithm and the extension to the multiinput, multioutput case.

Starting with (2), assume that n steps have been run, and rearrange to obtain

$$\begin{aligned} AV_n - v_{n+1}\rho_{n+1}e_n^T &= V_nT_n \\ V_n - \sigma AV_n + \sigma v_{n+1}\rho_{n+1}e_n^T &= V_n - \sigma V_nT_n \\ (I-\sigma A)\big(V_n + \sigma\rho_{n+1}(I-\sigma A)^{-1}v_{n+1}e_n^T\big) &= V_n(I-\sigma T_n) \end{aligned}$$

and

$$\begin{aligned} \big(V_n + \sigma\rho_{n+1}(I-\sigma A)^{-1}v_{n+1}e_n^T\big)(I-\sigma T_n)^{-1} \\ = (I-\sigma A)^{-1}V_n \end{aligned}$$

which results in the following:

$$\begin{aligned} V_n(I-\sigma T_n)^{-1} = (I-\sigma A)^{-1}V_n \\ - \sigma\rho_{n+1}(I-\sigma A)^{-1}v_{n+1}e_n^T(I-\sigma T_n)^{-1}. \end{aligned} \tag{8}$$

By a similar derivation, use (3) and again assume that n steps have been taken to obtain

$$\begin{aligned} \big(I-\sigma\tilde{T}_n^H\big)^{-1}W_n^H = W_n^H(I-\sigma A)^{-1} - \sigma\big(I-\sigma\tilde{T}_n^H\big)^{-1} \\ \times \eta_{n+1}^* e_n w_{n+1}^H(I-\sigma A)^{-1}. \end{aligned} \tag{9}$$

Premultiply (9) by e_1^T and then postmultiply by v_{n+1}. From (4), notice that $W_n^Hv_{n+1}=0$, so the left-hand side of (9) is equal to zero. Therefore, obtain

$$\begin{aligned} w_1^H(I-\sigma A)^{-1}v_{n+1} = \sigma e_1^T\big(I-\sigma\tilde{T}_n^H\big)^{-1}e_n\eta_{n+1}^*w_{n+1}^H \\ (I-\sigma A)^{-1}v_{n+1}. \end{aligned} \tag{10}$$

Now premultiply (8) by w_1^H and postmultiply by e_1, and note that $w_1^HV_n = \delta_1e_1^T$ to obtain

$$\begin{aligned} \delta_1e_1^T(I-\sigma T_n)^{-1}e_1 = w_1^H(I-\sigma A)^{-1}v_1 - \sigma\rho_{n+1}w_1^H \\ \times (I-\sigma A)^{-1}v_{n+1}e_n^T(I-\sigma T_n)^{-1}e_1. \end{aligned} \tag{11}$$

Last, multiply (11) by l^Hr, divide by δ_1, and use (6) and (7) to derive the following relationship between the transfer function and the reduced transfer function

$$\begin{aligned} H_n(s_0+\sigma) = H(s_0+\sigma) - (l^Hr)\sigma\rho_{n+1}w_1^H \\ \times (I-\sigma A)^{-1}v_{n+1}e_n^T(I-\sigma T_n)^{-1}e_1/\delta_1. \end{aligned}$$

Now substitute (10) into the above equation to give

$$\begin{aligned} H(s_0+\sigma) - H_n(s_0+\sigma) = (l^Hr)\sigma^2\rho_{n+1}\eta_{n+1}^*e_1^T \\ \times \big(I-\sigma\tilde{T}_n^H\big)^{-1}e_nw_{n+1}^H \\ \times (I-\sigma A)^{-1}v_{n+1} \\ \times e_n^T(I-\sigma T_n)^{-1}e_1/\delta_1. \end{aligned} \tag{12}$$

Using (5), the term $e_1^T(I-\sigma\tilde{T}_n^H)^{-1}e_n$ in the above equation yields

$$\begin{aligned} e_1^T\big(I-\sigma\tilde{T}_n^H\big)^{-1}e_n &= e_1^T\big(D_nD_n^{-1} - \sigma D_nT_nD_n^{-1}\big)^{-1}e_n \\ &= e_1^T\big(D_n(I-\sigma T_n)D_n^{-1}\big)^{-1}e_n \\ &= e_1^TD_n(I-\sigma T_n)^{-1}D_n^{-1}e_n \\ &= \delta_1e_1^T(I-\sigma T_n)^{-1}e_n/\delta_n. \end{aligned}$$

Substitute the above expression into (12) so

$$\begin{aligned} H(s_0+\sigma) - H_n(s_0+\sigma) = (l^Hr)\frac{\rho_{n+1}\eta_{n+1}^*}{\delta_n} \\ \times \sigma^2e_1^T(I-\sigma T_n)^{-1} \\ \times e_nw_{n+1}^H(I-\sigma A)^{-1}v_{n+1} \\ \times e_n^T(I-\sigma T_n)^{-1}e_1. \end{aligned}$$

Let

$$\tau_{ij}(\sigma) = e_i^T(I-\sigma T_n)^{-1}e_j$$

i.e., $\tau_{ij}(\sigma)$ denotes the (i,j) entry of the inverse of the tridiagonal matrix $I-\sigma T_n$. The above equation can now be written as

$$\begin{aligned} H(s_0+\sigma) - H_n(s_0+\sigma) = (l^Hr)\frac{\rho_{n+1}\eta_{n+1}^*}{\delta_n}\sigma^2\tau_{1n}(\sigma)\tau_{n1}(\sigma) \\ \times \big[w_{n+1}^H(I-\sigma A)^{-1}v_{n+1}\big] \end{aligned} \tag{13}$$

which is an exact expression for all σ in the complex plane except at the discrete points where $\tau_{1n}(\sigma)$, $\tau_{n1}(\sigma)$, or $(I-\sigma A)^{-1}$ is singular. When T_n is a good approximation for A, then these discrete points coincide. Note that if (13) is singular

because $(I-\sigma A)^{-1}$ is singular, then $H(s_0+\sigma)$ is also singular for the same reason from (6). Note that

$$(I-\sigma T_n)^{-1} = \frac{\mathrm{adj}(I-\sigma T_n)}{\det(I-\sigma T_n)}$$

where $\mathrm{adj}(I-\sigma T_n)$ is the classical adjoint matrix made up of $(n-1)\times(n-1)$ cofactors of $I-\sigma T_n$. For a tridiagonal matrix such as $I-\sigma T_n$, this results in

$$\tau_{1n}(\sigma) = e_1^T(I-\sigma T_n)^{-1}e_n = \frac{\sigma^{n-1}(\beta_2\beta_3\cdots\beta_n)}{\det(I-\sigma T_n)} \tag{14}$$

and

$$\tau_{n1}(\sigma) = e_n^T(I-\sigma T_n)^{-1}e_1 = \frac{\sigma^{n-1}(\rho_2\rho_3\cdots\rho_n)}{\det(I-\sigma T_n)}. \tag{15}$$

Substituting these back into (13) gives

$$\begin{aligned} H(s_0+\sigma) - H_n(s_0+\sigma) = {} & (l^H r) \\ & \times \frac{\sigma^{2n}(\beta_2\beta_3\cdots\beta_n\rho_2\rho_3\cdots\rho_n)}{\det(I-\sigma T_n)^2} \\ & \times \frac{\rho_{n+1}\eta_{n+1}^*}{\delta_n} \\ & \times \left[w_{n+1}^H(I-\sigma A)^{-1}v_{n+1}\right] \end{aligned} \tag{16}$$

which is an exact expression that shows an order σ^{2n} of approximation of the reduced transfer function $H_n(s_0+\sigma)$ to $H(s_0+\sigma)$. This also characterizes the Padé approximation.

Expression (13), which gives the error between $H_n(s_0+\sigma)$ and $H(s_0+\sigma)$, is essentially valid for all frequency values of σ, as indicated previously. Numerical experiments indicate that the terms $|\tau_{1n}(\sigma)|$ and $|\tau_{n1}(\sigma)|$ decrease steadily as the number of Lanczos iterations increase. In addition, the values of the term $|w_{n+1}^H(I-\sigma A)^{-1}v_{n+1}|$ essentially remain the same near convergence. $|\tau_{1n}(\sigma)|$ and $|\tau_{n1}(\sigma)|$ are the primary contributors to the convergence of the PVL algorithm. A theoretical justification of this observation is given in [1].

To use (13) as an error estimation for the approximation between $H_n(s_0+\sigma)$ and $H(s_0+\sigma)$ in the PVL algorithm, the major concern is on the cost of estimating the term $|w_{n+1}^H(I-\sigma A)^{-1}v_{n+1}|$ because the terms $|\tau_{1n}(\sigma)|$ and $|\tau_{n1}(\sigma)|$ can be computed cheaply (see the discussion in Section III). For all σ such that $|\sigma| < 1/\|A\|$, use the Cauchy–Schwartz inequality to obtain

$$\left|w_{n+1}^H(I-\sigma A)^{-1}v_{n+1}\right| \le \left\|w_{n+1}^H\right\|\|(I-\sigma A)^{-1}\|\|v_{n+1}\|$$

where recall $\|w_{n+1}^H\| = \|v_{n+1}\| = 1$. Furthermore, it is well known that

$$\|(I-\sigma A)^{-1}\| \le \frac{1}{1-|\sigma|\|A\|} \quad \text{when } |\sigma| < 1/\|A\|. \tag{17}$$

Then the following error bound for the approximation of the reduced transfer function is obtained:

$$\begin{aligned} |H(s_0+\sigma) - H_n(s_0+\sigma)| \le {} & |l^H r|\left|\frac{\rho_{n+1}\eta_{n+1}^*}{\delta_n}\right| \\ & \times \left|\frac{\sigma^2\tau_{1n}(\sigma)\tau_{n1}(\sigma)}{1-|\sigma|\|A\|}\right|. \end{aligned} \tag{18}$$

When $|\sigma| \ge 1/\|A\|$, the error bound (18) is no longer valid because of the inequality (17). However, as previously pointed out, since the primary contributors to the convergence of the PVL algorithm are the terms $|\tau_{1n}(\sigma)|$ and $|\tau_{n1}(\sigma)|$, one only needs to have an estimation for the term $|w_{n+1}^H(I-\sigma A)^{-1}v_{n+1}|$. In numerical experiments, it was observed that $|(1-|\sigma|\|A\|)^{-1}|$ might be used as an estimation of the term $\|(1-\sigma A)^{-1}\|$, and the bound (18) is still a plausible estimation for the approximation error even when $|\sigma| \ge 1/\|A\|$. Numerical examples in Section IV demonstrate this. An alternate approach for estimation of the term $|w_{n+1}^H(I-\sigma A)^{-1}v_{n+1}|$ is discussed in [1]. The optimal estimation of the term $|w_{n+1}^H(I-\sigma A)^{-1}v_{n+1}|$ when $|\sigma| \ge 1/\|A\|$ remains an open problem.

III. Implementation Issues of PVL-WEB

In this section, implementation issues for evaluation of this error bound are discussed.

A. Computing $\|A\|$

To compute $\|A\|$, one may use Hager and Higham's norm estimator [14], which only uses the matrix-vector multiplications Az and A^Hz. These operations are already available because they are required for the Lanczos process (see Algorithm 1).

Algorithm 2: Hager and Higham's norm estimator.

Choose any $x \in \mathbf{R}^N$ such that $\|x\|_1 = 1$
repeat
  $g \leftarrow Ax$.
  for $i = 1, 2, \cdots, N$ do
    $\zeta_i \leftarrow \begin{cases} g_i/|g_i| & \text{if } g_i \neq 0, \\ 1 & \text{if } g_i = 0. \end{cases}$
  endfor
  $z \leftarrow A^H\zeta$
  $z \leftarrow \mathrm{real}(z)$.
  if $\|z\|_\infty \le z^Tx$ then
    return $\|g\|_1$
  else
    find j such that $|z_j| = \|z\|_\infty$
    $x \leftarrow e_j$
  endif
endrepeat

The scalar $\|g\|_1$ returned from the above algorithm is an estimation of $\|A\|$. This is essentially the same algorithm used in the function `normest` in MATLAB.

B. Choosing σ for the Error Bound

The error bound in (18) is a function of σ. The term $|\sigma^2\tau_{1n}(\sigma)\tau_{n1}(\sigma)|$ can be evaluated with an order of n^3 flops for each frequency parameter σ. Under certain assumptions, the bound is an increasing function of the quantity $|\sigma|$. In general, although the bound does not strictly monotonically increase, the error bound is still somewhat of an increasing function of frequency. Therefore, each time the stopping criterion is tested, the error bound only needs to be computed at one bounding value of σ in the frequency range of interest. The following is an attempt to justify this observation in detail.

Note that in (18), the term $(1-|\sigma|||A||)^{-1}$ increases over $0 \le |\sigma| < 1/||A||$. Now consider the term $|\sigma^2\tau_{1n}(\sigma)\tau_{n1}(\sigma)|$ and rename it $\chi(\sigma)$ for convenience. By (14) and (15), obtain

$$\chi(\sigma) = \frac{|\sigma|^{2n}|\beta_2\beta_3\cdots\beta_n\rho_2\rho_3\cdots\rho_n|}{|\det(I-\sigma T_n)^2|}.$$

Let the eigenvalues of T_n be λ_i for $1 \le i \le n$ and let $\sigma = |\sigma|e^{j\theta}$ so

$$\chi(\sigma) = \frac{|\beta_2\beta_3\cdots\beta_n\rho_2\rho_3\cdots\rho_n|}{|(|\sigma|^{-1}-e^{j\theta}\lambda_1)|^2\cdots|(|\sigma|^{-1}-e^{j\theta}\lambda_n)|^2}.$$

From the last equation, when maximizing $\chi(\sigma)$, the variation of θ is of no importance when $|\sigma|$ is small and $0 \le |\sigma| < 1/||T_n||$. In addition, by considering plots of s, s_0, σ, and θ in the s plane, it is straightforward to see that θ varies slowly when $|\sigma|$ is large [20]. Therefore, if it is assumed that θ is a constant such that $e^{j\theta}\lambda_i = a_i + jb_i$, then

$$\chi(\sigma) = \frac{|\beta_2\beta_3\cdots\beta_n\rho_2\rho_3\cdots\rho_n|}{[(|\sigma|^{-1}-a_1)^2+b_1^2]\cdots[(|\sigma|^{-1}-a_n)^2+b_n^2]}.$$

It can be shown that this is an increasing function of $|\sigma|$ for $0 \le |\sigma| < |\sigma_0| < 1/||T_n||$ where σ_0 is the bounding point. However, since in practice and in the numerical simulations in Section IV θ is not a constant, the error bound does not strictly monotonically increase with $|\sigma|$. It is explained in [20] that although the error bound does not monotonically increase with $|\sigma|$ for all frequencies, the frequency regions where it may decrease are small. Outside these regions the bound will continue to increase. The maximum number of regions where it may decrease is equal to the dimension of A. Thus for all σ in the range, the bound only needs to be computed at one bounding point σ_0. It should be noted that this observation is in agreement with Padé approximations in general. Typically, the error in an approximation increases with the distance from the expansion point.

C. Entries in the Matrix $(I-\sigma T_n)^{-1}$

The quantities $\tau_{1n}(\sigma)$ and $\tau_{n1}(\sigma)$ are the $(1,n)$ and $(n,1)$ entries of the inverse of the tridiagonal matrix $I-\sigma T_n$. Since $n \ll N$ and $I-\sigma T_n$ is tridiagonal, these quantities are inexpensive to compute using an LU decomposition. In addition, as the order of approximation n increases, the leading rows and columns of T_n do not change. Therefore the LU decomposition does not have to be recomputed each time but rather can be inexpensively updated with each iteration. Using the LU decomposition of $I-\sigma T_n$, one may run forward and backward substitution on e_1 to find $\tau_{n1}(\sigma)$ and e_n to find $\tau_{1n}(\sigma)$. Therefore, it is not necessary to compute $(I-\sigma T_n)^{-1}$. Last, note that when running the forward and backward solver on e_1, the quantity $\tau_{11}(\sigma)$ will also be computed. This quantity will provide the frequency-domain response since $H_n(s_0+\sigma) = (l^Hr)e_1^T(I-\sigma T_n)^{-1}e_1$. Therefore, if the time-domain response is not required, the poles and residues need not be computed.

D. Criterion Used to Test if the Lanczos Iterations Can be Terminated

As pointed out earlier, unlike the existing PVL algorithm, the important advantage of the PVL-WEB algorithm is that the number of Lanczos iterations is not prescribed and no user supervision is required for termination of the process. An adaptive scheme can be incorporated into Algorithm 1 that increases the order of approximation n until an acceptably small error of the approximation of the reduced transfer function $H_n(s_0+\sigma)$ is achieved. Specifically, when $|\sigma| < 1/||A||$, the statement in Algorithm 1

$$\text{for } n = 1, 2, \cdots, q \text{ do}$$

can be replaced by

$$\begin{aligned}&\text{while (not converged) do}\\ &\quad n \leftarrow n+1\end{aligned}$$

where the criterion used to test if the Lanczos iteration can be terminated is

$$\text{while}\left(\left|\frac{\sigma^2(l^Hr)}{(1-|\sigma|||A||)}\right|\left|\frac{\eta_{n+1}^*\rho_{n+1}\tau_{1n}(\sigma)\tau_{n1}(\sigma)}{\delta_n}\right| > tol\right) \text{ do} \tag{19}$$

where *tol* is the error tolerance the user defines on the approximation.

It should be noted that the above criterion is only guaranteed for $|\sigma| < 1/||A||$. Additional work is required to obtain an equation similar to (19) from (13) for a reliable and efficient stopping criterion for $|\sigma| \ge 1/||A||$; see the discussion at the end of Section II.

E. Calculating the Poles and Residues

After the Lanczos process has been terminated, the dominant poles and residues of the system can be computed and used to visualize the characteristics of the system. The poles and residues can also be used to determine the time-domain impulse response. The procedure for determining the poles and residues is shown in the following algorithm, which is taken from [8].

Algorithm 3: Computing poles and residues.

Compute the eigendecomposition
$$T_q = S_q\,\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_q)S_q^{-1}.$$
set $\mu = S_q^Te_1$, $\nu = S_q^{-1}e_1$ and $k_\infty = 0$.
for $j = 1, 2, \cdots, q$ do
  if $\lambda_j \neq 0$ then
    $p_j = 1/\lambda_j$ and $k_j = -(l^Hr)\mu_j\nu_j/\lambda_j$
  else
    $k_\infty = k_\infty + (l^Hr)\mu_j\nu_j$.
  endif
endfor

Note that p_j given above is in the shifted coordinates s_0.

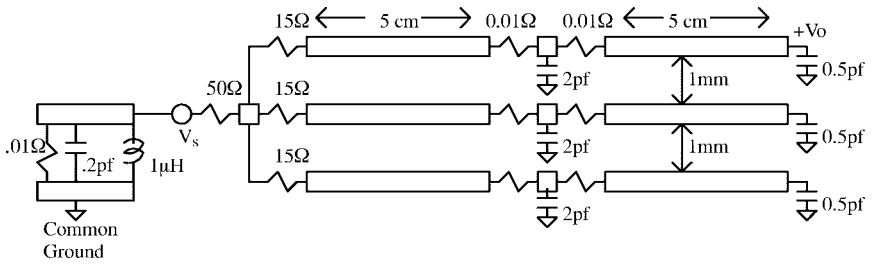

Fig. 1. MTL circuit modeled using PEEC.

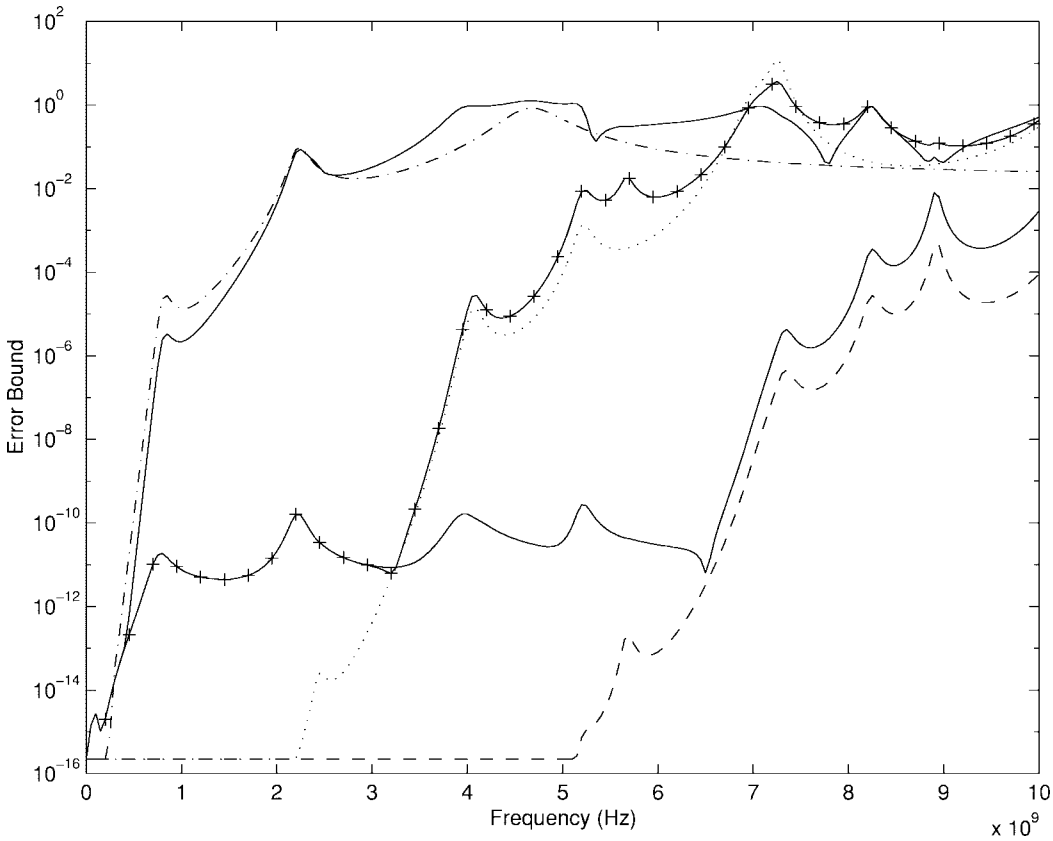

Fig. 2. Computed and PVL-WEB estimated errors. Top solid line: computed error for nine iterations; dash–dot line: estimated error for nine iterations; solid line with pluses: computed error for 31 iterations; dot–dot line: estimated error for 31 iterations; bottom solid line: computed error for 65 iterations; dash–dash line: estimated error for 65 iterations.

IV. Numerical Example

The numerical example presented in this section is from a PEEC model of an electromagnetics problem. The model consists of a set of six flat, lossy transmission lines (strips) with discrete capacitors, inductors, and resistors arranged to give rejection over certain narrow frequency bands (see Fig. 1). The example is a modified version of an MTL circuit evaluated in [4]. There are 306 capacitors, 294 inductors, and 294 resistors used in the PEEC model of the strips. In addition, there are 11 lumped resistors, one lumped inductor, and seven lumped capacitors used in this simulation. The size of the resulting matrices C and G is $N = 918$. The norm of the matrix A is estimated to be 5.73×10^{-10}. The expansion point is $s_0 = 0$ and $s = j2\pi f$ for the frequency f. Last, $\sigma = s - s_0$, and no assumption is made on the angle θ of σ.

Unlike previous PVL simulations, the number of Lanczos iterations is *not* prescribed for the PVL-WEB algorithm. Instead, the user sets up a tolerance value *tol* for the error between the original transfer function and reduced transfer function and an upper bounding value of $|\sigma|$ where it is desired to have the approximation within the tolerance value. The PVL-WEB will terminate automatically when the desired accuracy at the upper bounding value of $|\sigma|$ is satisfied. However, it should be noted that since the error bound is computationally efficient, it can be calculated at each frequency point of interest, as is shown in the following simulations. The termination criteria could be for the error bound across the band to be less than a prescribed value, but in this work only one upper bounding point was checked for convergence although the error bound was calculated for all the points.

In the first simulation, the tolerance value is set to $tol = 10^{-4}$ at a bounding frequency of 1 GHz. The PVL-WEB took nine Lanczos iterations to meet this error criterion. With the PVL-WEB technique, the error bound on the approximation is also obtained. This is plotted in Fig. 2, where the dash–dot line

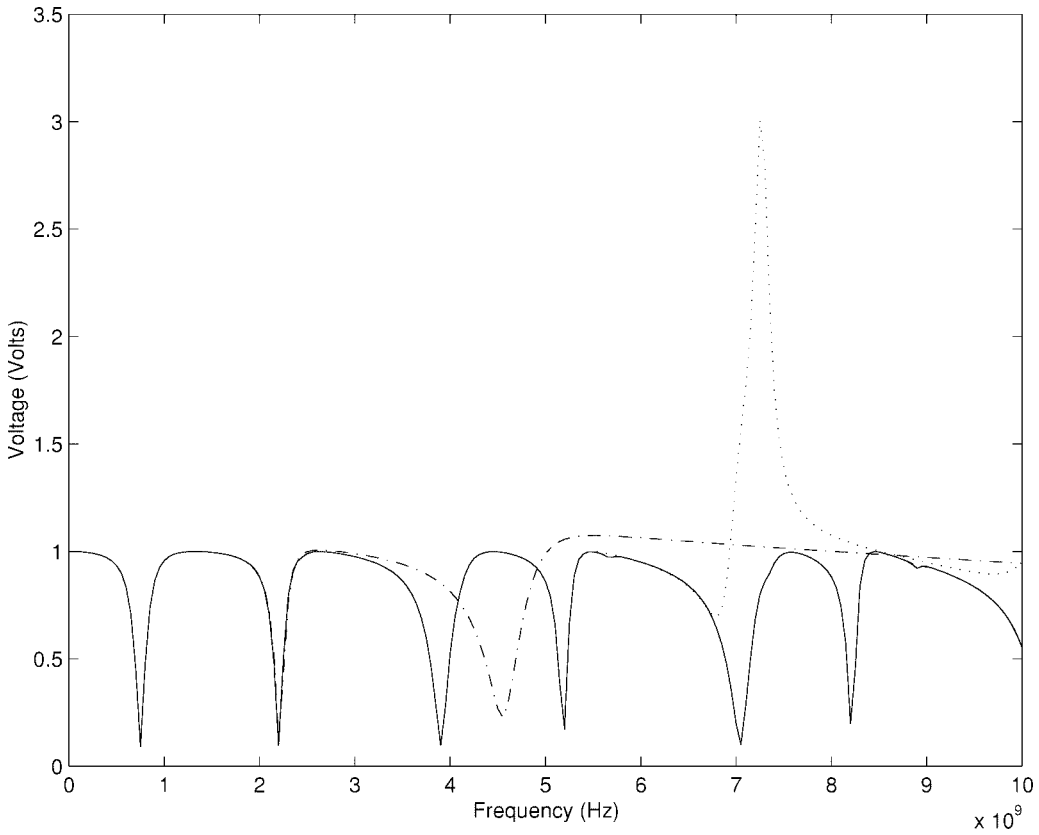

Fig. 3. Magnitude of the frequency responses. Solid line: $\bar{H}(s)$; dash–dot line: $H_n(s)$ for *tol* $= 10^{-4}$ at 1 GHz; dot–dot line: $H_n(s)$ for *tol* $= 10^{-4}$ at 5 GHz; dash–dash line: $H_n(s)$ for *tol* $= 10^{-4}$ at 10 GHz.

is the PVL-WEB estimated error between the original response $H(s)$ and the approximate response $H_n(s)$. Once the convergence criterion was met, $\tilde{H}(s)$ (which is $H(s)$ computed with an LU decomposition in finite precision arithmetic) was calculated to illustrate the validity of the PVL-WEB. The top solid curve in Fig. 2 is the computed error between $\tilde{H}(s)$ and the approximate response $H_n(s)$. The solid curves in Figs. 3 and 4 are the magnitude and phase of $\tilde{H}(s)$. The dash–dot lines in these two figures are the magnitude and phase of the reduced frequency response at convergence for the first simulation. The reduced order model shows good agreement out to the upper bounding $|\sigma|$ and slightly beyond.

In the second simulation, the tolerance value is still *tol* $= 10^{-4}$, but now the error measurement is taken at a bounding frequency of 5 GHz. With these conditions, the PVL-WEB took 31 Lanczos iterations to converge. The dotted line in Fig. 2 is the PVL-WEB estimated error between the original response and the approximate response. The computed error between $\tilde{H}(s)$ and $H_n(s)$ is plotted as a solid line with pluses. The computed results of the magnitude and the phase for this case are plotted in Figs. 3 and 4 using dot–dot lines. Again, good agreement between $\tilde{H}(s)$ and $H_n(s)$ is shown to 5 GHz and slightly beyond.

In the third simulation, the tolerance value is again *tol* $= 10^{-4}$, but the error measurement is taken at 10 GHz. The PVL-WEB took 65 Lanczos iterations to converge. The dash–dash line in Fig. 2 is the PVL-WEB estimated error, and the bottom solid line is the computed error between $\tilde{H}(s)$ and $H_n(s)$. The responses for this case are plotted in Figs. 3 and 4 using dash–dash lines. Note that because of the scaling used in Figs. 3 and 4, the approximate response $H_n(s)$ (dash–dash lines) is indistinguishable from the computed response $\tilde{H}(s)$ (solid line). Last, the exact response took 14 320.5 s to compute, and the third PVL-WEB simulation took 75.59 s to compute. This includes calculating the bound at not only the upper bounding value of σ but also all frequencies of interest as plotted.

Notice that in Fig. 2, the computed errors between $\tilde{H}(s)$ and $H_n(s)$ (solid lines) are larger than the PVL-WEB estimated errors in certain frequency regions. There are two factors contributing to this phenomenon. The first factor is numerical errors in computing $\tilde{H}(s)$. The condition numbers of the matrix $I - \sigma A$ are as large as order 10^6. From standard numerical error analysis, it is known that $\tilde{H}(s)$ cannot be computed with more accuracy than 10^{-10} using double machine precision. The second factor is the simple inequality (17) that is used to bound the approximation error (18). As discussed in Section II, in the high-frequency region, $|\sigma| \geq 1/\|A\|$, the inequality (17) is no longer valid. In general, in the lower $|\sigma|$ region, the first factor contributes to the discrepancy between the computed and PVL-WEB estimated errors. At the higher $|\sigma|$ region, however, the second factor dominates. Nevertheless, the PVL-WEB estimated errors essentially still track the computed errors. A more refined and computationally costly scheme could be developed to bound (18) without using the inequality (17) and overcome the limitation of the

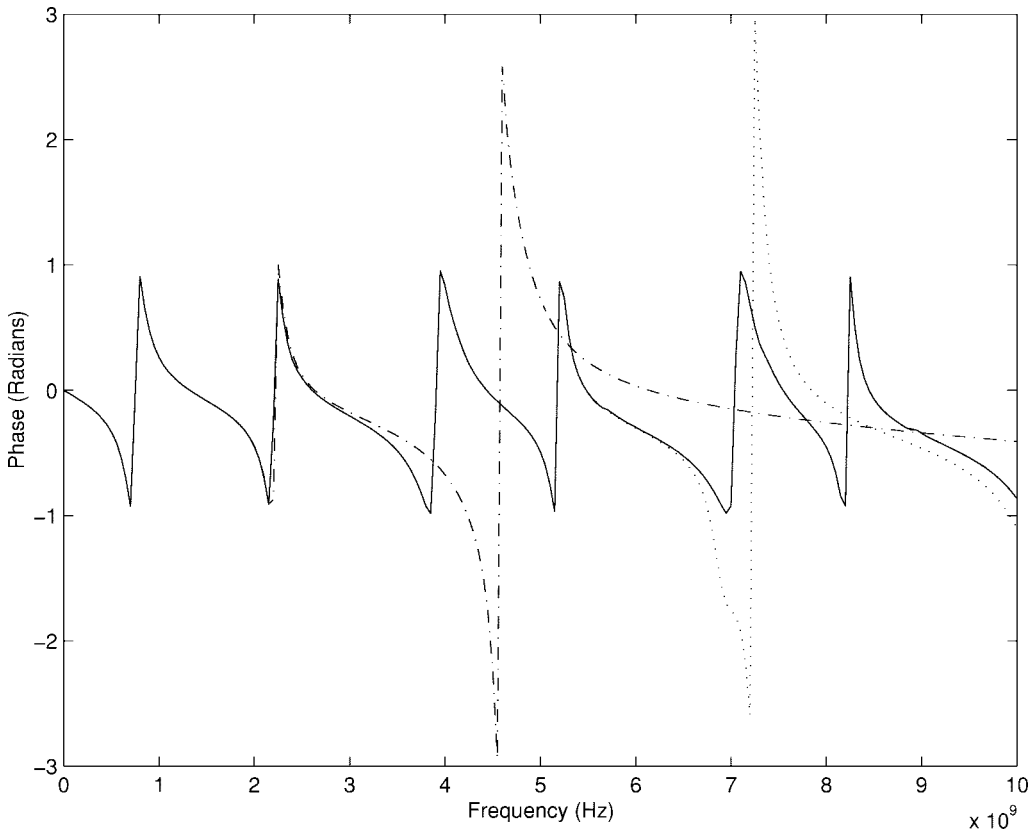

Fig. 4. Phase of the frequency responses. Solid line: $\bar{H}(s)$; dash–dot line: $H_n(s)$ for *tol* $= 10^{-4}$ at 1 GHz; dot–dot line: $H_n(s)$ for *tol* $= 10^{-4}$ at 5 GHz; dash–dash line: $H_n(s)$ for *tol* $= 10^{-4}$ at 10 GHz.

error estimate approximation. In this study, however, (18) was shown to be efficient and useful in terminating the PVL algorithm as shown in Fig. 3 despite violation of the inequality in the upper frequency regions.

V. Summary and Conclusions

In this paper, a new matrix-based derivation of the error between the original circuit transfer function and the reduced-order transfer function generated using the PVL technique was presented. This derivation gave rise to a computationally inexpensive error bound of the approximation. It was shown how this error bound could be implemented as a stopping criterion for the PVL algorithm. Several practical implementation issues were discussed. Numerical simulation results were presented to illustrate the combined PVL-WEB method.

Matrix PVL for a multiple-input, multiple-output system was introduced in [9]. The error-estimation scheme presented in this paper can be extended to multiple-input, multiple-output systems. There is also a good theoretical understanding of the convergence of the PVL approximation. In addition, by using proper matrix balancing, it is also possible to reduce $\|A\|$ so that the simple error bound holds for the higher frequency regions, $|\sigma| \geq 1/\|A\|$. For discussion of these issues, refer to [1] and references therein.

Acknowledgment

The authors would like to thank Dr. T. Jerse for providing the framework for the code to calculate the potential coefficients and partial inductances of the PEEC model. They also would like to thank the referees for their valuable comments on the manuscript. R. D. Slone would also like to thank T. Kowalski for many helpful discussions.

References

[1] Z. Bai and Q. Ye, "Error estimation of the Padé approximation of transfer function via the Lanczos process," *Elec. Trans. Numer. Anal.*, vol. 7, pp. 1–17, 1998.

[2] G. A. Baker, Jr., and P. Graves-Morris, *Padé Approximants, Part I: Basic Theory*. Reading, MA: Addison-Wesley, 1981.

[3] J. E. Bracken, V. Raghavan, and R. A. Rohrer, "Interconnect simulation with asymptotic waveform evaluation (AWE)," *IEEE Trans. Circuits Syst.*, vol. 39, pp. 869–878, Nov. 1992.

[4] M. Celik and A. C. Cangellaris, "Simulation of dispersive multiconductor transmission lines by Padé approximation via the Lanczos process," *IEEE Trans. Microwave Theory Tech.*, vol. 44, pp. 2525–2535, Dec. 1996.

[5] E. Chiprout and M. S. Nakhla, "Analysis of interconnect networks using complex frequency hopping (CFH)," *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 186–200, Feb. 1995.

[6] ———, *Asymptotic Waveform Evaluation and Moment Matching for Interconnect Analysis*. Boston, MA: Kluwer Academic, 1994.

[7] S. K. Das and W. T. Smith, "Application of asymptotic waveform evaluation for analysis of skin effect in lossy interconnects," *IEEE Trans. Electromag. Compat.*, to be published.

[8] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Padé approximation via the Lanczos process," *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 639–649, May 1995.

[9] ———, "Reduced-order modeling of large linear subcircuits via a block Lanczos algorithm," in *Proc. 32nd Design Automation Conf.*, June 1995.

[10] K. Gallivan, E. Grimme, and P. Van Dooren, "Asymptotic waveform evaluation via a Lanczos method," *Appl. Math. Lett.*, vol. 7, no. 5, pp. 75–80, 1994.

[11] W. B. Gragg, "Matrix interpretations and applications of the continued fraction algorithm," *Rocky Mountain J. Math.*, vol. 4, pp. 213–225, 1994.

[12] E. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1997.
[13] H. Heeb, A. E. Ruehli, J. E. Bracken, and R. A. Rohrer, "Three dimensional circuit oriented electromagnetic modeling for VLSI interconnects," in *Proc. IEEE Int. Conf. Computer Design*, 1992, pp. 218–221.
[14] N. J. Higham, "FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation," *ACM Trans. Math. Soft.*, vol. 14, no. 4, pp. 381–396, Dec. 1988.
[15] C. W. Ho, A. E. Ruehli, and P. A. Brennan, "The modified nodal approach to network analysis," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 504–509, June 1975.
[16] R. Khazaka, E. Chiprout, M. S. Nakhla, and Q. J. Zhang, "Analysis of high-speed interconnects with frequency dependent parameters," in *Proc. 11th Int. Zurich Symp. Technical Exhibition on Electromagnetic Compatibility*, Mar. 1995, pp. 203–208.
[17] T. V. Nguyen, J. Li, and Z. Bai, "Dispersive coupled transmission line simulation using an adaptive block Lanczos algorithm," in *Proc. IEEE Conf. Custom Integrated Circuits*, 1996.
[18] L. T. Pillage and R. A. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. Computer Aided-Design*, vol. 9, pp. 352–366, Apr. 1990.
[19] A. E. Ruehli, "Equivalent circuit models for three-dimensional multiconductor systems," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-22, pp. 216–221, Mar. 1974.
[20] R. D. Slone, "A computationally efficient method for solving electromagnetic interconnect problems: The Padé approximation via the Lanczos process with an error bound," Master's thesis, University of Kentucky, Lexington, 1997.
[21] T. K. Tang and M. S. Nakhla, "Analysis of high-speed VLSI interconnects using the asymptotic waveform evaluation technique," *IEEE Trans. Computer-Aided Design*, vol. 11, pp. 341–352, Mar. 1992.
[22] D. Xie and M. S. Nakhla, "Delay and crosstalk simulation of high-speed VLSI interconnects with nonlinear terminations," *IEEE Trans. Computer Aided Design*, vol. 12, pp. 1198–1211, Nov. 1993.

Zhaojun Bai is an Associate Professor in the Department of Mathematics at the University of Kentucky, Lexington. His major research interests include numerical linear algebra, parallel scientific computing, and software development. He was involved in the design and implementation of the numerical linear algebra software package LAPACK and is a coauthor of the LAPACK user's guide.

Rodney D. Slone (S'93) received the B.S. degree in electrical engineering and mathematics and the M.S. degree in electrical engineering from the University of Kentucky, Lexington, in 1996 and 1997, respectively. He currently is pursuing the Ph.D. degree in electrical engineering from The Ohio State University, Columbus.

His areas of interest are computational electromagnetics, numerical analysis, and electromagnetic compatibility.

Mr. Slone was a National Merit Scholar, a National Science Scholar, and a Robert C. Byrd Scholar. He has received scholarships and fellowships from the University of Kentucky and a fellowship from The Ohio State University.

William T. Smith (S'88–M'90) received the B.S.E.E. degree from the University of Kentucky, Lexington, in 1980. He received the M.S.E.E. and Ph.D.E.E. degrees from Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, in 1986 and 1990, respectively.

He is an Associate Professor in the Department of Electrical Engineering at the University of Kentucky. From 1981 to 1984, he was an RF engineer for Harris Corp., Melbourne, FL. He was with the Satellite Communications Group while at Virginia Tech. He joined the Faculty of the University of Kentucky in 1990. During the summers of 1991 and 1992, he was a NASA/ASEE Summer Faculty Fellow at NASA Langley Research Center, Hampton, VA.

Dr. Smith is a member of the American Society for Engineering Education, Eta Kappa Nu, Tau Beta Pi, and Omicron Delta Kappa.

Qiang Ye is an Associate Professor in the Department of Applied Mathematics, University of Manitoba, Canada. His research interests include numerical analysis, scientific computing, and applied linear algebra. He is presently involved in research on numerical methods for large matrices and their applications.