Title	Fast nonlinear model order reduction via associated transforms of high-order volterra transfer functions
Author(s)	Zhang, Y; Liu, H; Wang, Q; Fong, N; Wong, N
Citation	The 49th ACM/EDAC/IEEE Design Automation Conference (DAC 2012), San Francisco, CA., 3-7 June 2012. In ACM/IEEE Design Automation Conference Proceedings, 2012, p. 289-294
Issued Date	2012
URL	http://hdl.handle.net/10722/165276
Rights	ACM/IEEE Design Automation Conference Proceedings. Copyright © IEEE Computer Society.

Fast Nonlinear Model Order Reduction via Associated Transforms of High-Order Volterra Transfer Functions

Yang Zhang, Haotian Liu, Qing Wang, Neric Fong and Ngai Wong
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong
{yzhang,htliu,wangqing,nfong,nwong}@eee.hku.hk

ABSTRACT

We present a new and fast way of computing the projection matrices serving high-order Volterra transfer functions in the context of (weakly and strongly) nonlinear model order reduction. The novelty is to perform, for the first time, the association of multivariate (Laplace) variables in high-order multiple-input multiple-output (MIMO) transfer functions to generate the standard single-s transfer functions. The consequence is obvious: instead of finding projection subspaces about every s_i, only that about a single s is required. This translates into drastic saving in computation and memory, and much more compact reduced-order nonlinear models, without compromising any accuracy.

Categories and Subject Descriptors

B.7.2 [**Hardware**]: Design Aids—*Simulation, Verification*; I.6.5 [**Computing Methodologies**]: Model Development—*Modeling methodologies*; J.6 [**Computer Applications**]: Computer-Aided Engineering—*Computer-aided design (CAD)*

General Terms

Algorithms, Design, Theory, Verification

Keywords

Association of variables, Model order reduction (MOR), Nonlinear system, Analog/RF circuits

1. INTRODUCTION

Simulation techniques for VLSI circuits at the system level are strongly demanded. The analog and radio-frequency (RF) modules, though occupying a small part of a typical mixed-signal chip, are critical while hard to simulate and design due to their nonlinearities. Indeed, model order reduction (MOR) for complex nonlinear systems is often needed whereby the reduced-order model (ROM) inherits the dominant dynamics of the original system while featuring a much smaller dimension for simulation.

DAC 2012, June 3-7, 2012, San Francisco, California, USA.

So far, MOR techniques have been studied extensively for linear time-invariant (LTI) systems, such as explicit moment matching based asymptotic waveform evaluation (AWE) [12] and projection-based implicit moment matching [9]. The underlying workhorse appears to be the Krylov subspace method (e.g. [11, 9]), which projects the original system onto a ROM via a projection matrix from matching the moments of the original system. Although these methods have made a great success on the LTI systems, the situation becomes complicated in the nonlinear scenario.

As a powerful analysis tool for nonlinear systems, Volterra theory has been studied for decades [15] and successfully applied to nonlinear MOR (NMOR) for years, e.g., [10, 7, 6]. Although it provides an analytical and systematic approach, deployment of the Krylov subspace method (e.g., the NORM algorithm proposed in [7, 6]) suffers from exponentially growing subspace dimensions due to the multiple frequency axes in high-order Volterra transfer functions [15]. Subsequently, the order of the Volterra series used for matching the moments in NMOR is severely limited, rendering the Volterra approach applicable essentially to weakly nonlinear systems. Other NMOR methods, such as the trajectory piecewise-linear (TPWL) approximation [14], which can deal with strongly nonlinear systems, also suffer from training input sequence dependence.

Recently, the MOR of strongly nonlinear systems is transformed into the MOR problem of the quadratic-linear differential algebraic equations (QLDAEs) [4, 5], which are obtained by adding extra states related to strong nonlinearities such as the sine/cosine or exponential (diode-type) curves:

$$C\dot{x} = G_1 x + G_2 x \otimes x + D_1 xu + D_2 x \otimes xu + bu, \qquad (1)$$

where $x \in \mathbb{R}^n$ is the state vector and $\otimes$ denotes the Kronecker product. All other matrices are of compatible dimensions and a scalar input u is assumed for notational ease whose multi-input multi-output (MIMO) generalization is possible. The key advantage of QLDAE is that it keeps the strongly nonlinear functions only in quadratic-linear format instead of cubic or higher-order terms.

In the following, we work with a trimmed version of (1) by assuming an invertible C (called a regular system borrowing from the linear system terminology) so that it can be replaced with an identity matrix. Also, we leave out the D_2 term as it seldom appears in electrical circuits of interest. Subsequently, (1) turns into

$$\dot{x} = G_1 x + G_2 x \otimes x + D_1 xu + bu. \qquad (2)$$

These assumptions are made mainly for the ease of notation.

For instance, a singular C is analogous to a linear descriptor system whereby the regular (nonsingular) part can be extracted via the canonical projector or a Weierstrass form transformation [16], and the impulsive (singular) part is often immaterial or related algebraically to the regular subsystem. Moreover, all results in this paper are extensible to include the D_2 component and a multi-column B (instead of the vector input matrix b).

The key contribution of this paper is to break the "dimensionality curse" of Krylov subspace method in the NMOR context. Strong nonlinearity is accommodated with the use of QLDAE [4, 5]. Whereas the true computation bottleneck, viz. the transfer matrix moment expansion at multiple frequency axes, is completely avoided by the *first-time proposed* use of association of variables in multivariate MIMO high-order Volterra transfer functions. The consequence is obvious: instead of finding projection subspaces about every s_i which leads to an exponential growth in the overall subspace dimension, only that about a single s is required. This translates into drastic saving in computation and memory, and much more compact nonlinear ROMs, without compromising any accuracy.

This paper is organized as follows. Section 2 briefly reviews the nonlinear system description by Volterra theory, as well as the *association of variables* and the corresponding *associated transform*. Then, the conventional single-input single-output (SISO) association of variables method is extended to its MIMO counterpart through two important theorems. The Krylov subspace generation in associated transfer functions, with practical considerations, is described. In Section 3, the proposed scheme is verified through examples. Section 4 discusses some important remarks. Finally, Section 5 draws the conclusion.

2. ASSOCIATED TRANSFORM IN NMOR

We present the main results of the paper in this section. A succinct account of association of frequency-domain (Laplace) variables is first given, which is mainly applied to SISO systems in the literature, see e.g., [1, 2, 8, 13, 15]. Then, important theorems are devised which allow the natural utilization of associated transform in MIMO scenario to facilitate projection-based NMOR.

2.1 Volterra theory and association of variables

To begin with, the state vector of a Volterra system is progressively approximated with high-order responses, namely,

$$x(t) = x_1(t) + x_2(t) + x_3(t) + \cdots,$$

where

$$x_n(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n) \cdot u(t-\tau_1) \cdots u(t-\tau_n) d\tau_1 \cdots d\tau_n, \quad (3)$$

and $h_n(\tau_1, \cdots, \tau_n)$ is the n-th order Volterra kernel. In particular, x_1 is the usual first-order convolution having its Laplace domain representation $X_1(s) = H_1(s)U(s)$ where $H_1(s) = \int_{-\infty}^{\infty} h_1(\tau) e^{-s\tau} d\tau$ is the impulse response transfer function form. Analogously, the (nonlinear) high-order transfer function counterparts are defined as

$$H_n(s_1, \cdots, s_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n) \cdot e^{-s_1\tau_1} \cdots e^{-s_n\tau_n} d\tau_1 \cdots d\tau_n. \quad (4)$$

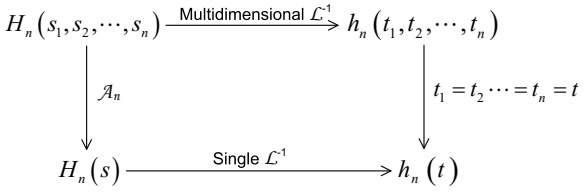


Figure 1: Association of variables for finding $H_n(s)$ from $H_n(s_1, s_2 \cdots, s_n)$.

Nonetheless, unlike the first-order case, there is no direct counterpart in the (multivariate) Laplace domain except if we replace the single time axis in the product of u in (3) by multiple axes as $u(t_1 - \tau_1) \cdots u(t_n - \tau_n)$, yielding [15]

$$X_n(s_1, \cdots, s_n) = H_n(s_1, \cdots, s_n) U(s_1) \cdots U(s_n). \quad (5)$$

The (multidimensional) inverse Laplace transform of (4) is a generalization of the univariate formula, given by

$$h_n(t_1, \cdots, t_n) = \mathscr{L}^{-1}\left(H_n(s_1, \cdots, s_n)\right) = \frac{1}{(2\pi j)^n} \int_{\sigma_n - j\infty}^{\sigma_n + j\infty} \cdots \int_{\sigma_1 - j\infty}^{\sigma_1 + j\infty} H_n(s_1, \cdots, s_n) \cdot e^{s_1 t_1} \cdots e^{s_n t_n} ds_1 \cdots ds_n. \quad (6)$$

To restore the required $h_n(t)$, one then evaluates along the *diagonal line* in the multi-time hyperplane, i.e., $h_n(t) = h_n(t_1, \cdots, t_n)|_{t_1 = t_2 = \cdots = t_n = t}$. Of course, the same procedure is used to obtain $x_n(t)$ directly from $X_n(s_1, \cdots, x_n)$ in (5) if the input $u(t)$ or $U(s)$ is known.

The above approach unifies the time variables in the time domain. An alternative is to carry out the unification first in the multi-frequency domain. So its time-domain counterpart is automatically a single-time function. This process is termed the *association of variables* and the corresponding frequency function the *associated transform*, denoted as $H_n(s) = \mathscr{A}_n\left(H_n(s_1, \cdots, s_n)\right)$, from which $h_n(t)$ can be derived from the conventional inverse Laplace transform of $H_n(s)$. Fig. 1 depicts the relationship between these time- and frequency-domain operations.

A closed-form expression for the association of variables also follows from (6) by setting $t_1 = \cdots = t_n = t$ [13, 15]:

$$H_n(s) = \mathscr{A}_n\left(H_n(s_1, \cdots, s_n)\right) = \frac{1}{(2\pi j)^{n-1}} \int_{\sigma_n - j\infty}^{\sigma_n + j\infty} \cdots \int_{\sigma_2 - j\infty}^{\sigma_2 + j\infty} H_n(s - s_2 - \cdots - s_n, s_2, \cdots, s_n) ds_2 \cdots ds_n. \quad (7)$$

Moreover, certain factored forms in $H_n(s_1, \cdots, s_n)$ allow the direct use of (6) and/or (7) to produce useful theorems, e.g., see Chapter 2 of [15]. (Henceforth, the dimensional subscripts in the transfer functions or association operator are sometimes omitted when they are obvious from context.) In particular, an interesting property [1] is that if $H(s_1, \cdots, s_n)$ can be written as $F(s_1 + \cdots + s_n) G(s_1, \cdots, s_n)$, then by (7)

$$H(s) = \mathscr{A}\left(F(s_1 + \cdots + s_n) G(s_1, \cdots, s_n)\right) = F(s) \mathscr{A}\left(G(s_1, \cdots, s_n)\right) = F(s) G(s). \quad (8)$$

Nonetheless, to our best knowledge, existing works on association of variables mainly deal with SISO (viz. scalar) systems (see e.g., [1, 2, 8, 13]), even though the formulas in (3)–(7) do not distinguish between SISO or MIMO cases.

2.2 MIMO extension

In the following, we propose two important theorems which facilitate the use of the association of variables in the NMOR of general MIMO systems. To start with, two properties of Kronecker product ($\otimes$) and Kronecker sum ($\oplus$), to be used in later proofs, are recalled: i) for compatible dimensions, $(M_1 \otimes M_2)(N_1 \otimes N_2) = (M_1N_1) \otimes (M_2N_2)$ and ii) $e^{M\oplus N} = e^M \otimes e^N$, wherein $M \in R^{n_M \times n_M}$, $N \in R^{n_N \times n_N}$ and $M \oplus N = M \otimes I_{n_N} + I_{n_M} \otimes N$. We also use two handy shorthands for multiple Kronecker product or sum of a matrix M, namely, $M \otimes M = M^{②}$ and $M \oplus M = ②M$ etc.

THEOREM 1. *For two square matrices $A_1 \in \mathbb{R}^{n_1 \times n_1}$ and $A_2 \in \mathbb{R}^{n_2 \times n_2}$, associating the Kronecker product of their resolvent matrices in variables s_1 and s_2 is given by*

$$\begin{aligned}\mathscr{A}_2\left((s_1 I_{n_1} - A_1)^{-1} \otimes (s_2 I_{n_2} - A_2)^{-1}\right)\\ = (sI_{n_1 n_2} - (A_1 \oplus A_2))^{-1}.\end{aligned} \tag{9}$$

PROOF. We apply (6) to find the associated time function of the Kronecker product, namely,

$$\begin{aligned}&\frac{1}{(2\pi j)^2}\int_{\sigma_2-j\infty}^{\sigma_2+j\infty}\int_{\sigma_1-j\infty}^{\sigma_1+j\infty}(s_1 I_{n_1} - A_1)^{-1} \otimes (s_2 I_{n_2} - A_2)^{-1}\cdot\\ &\qquad e^{s_1 t_1} e^{s_2 t_2} ds_1 ds_2\\ =&\left(\frac{1}{2\pi j}\int_{\sigma_1-j\infty}^{\sigma_1+j\infty}(s_1 I_{n_1} - A_1)^{-1} e^{s_1 t_1} ds_1\right) \otimes\\ &\qquad \left(\frac{1}{2\pi j}\int_{\sigma_2-j\infty}^{\sigma_2+j\infty}(s_2 I_{n_2} - A_2)^{-1} e^{s_2 t_2} ds_2\right)\\ =&\, e^{A_1 t_1} \otimes e^{A_2 t_2} = e^{A_1 t_1 \oplus A_2 t_2}.\end{aligned} \tag{10}$$

Setting $t_1 = t_2 = t$ in (10) the proof is complete. □

COROLLARY 1. *Repeatedly using Theorem 1 we get the general result*

$$\begin{aligned}\mathscr{A}_k\left((s_1 I_{n_1} - A_1)^{-1} \otimes \cdots \otimes (s_k I_{n_k} - A_k)^{-1}\right)\\ = \left(sI_{n_1 n_2 \cdots n_k} - \oplus_{i=1}^{k}(A_i)\right)^{-1}.\end{aligned} \tag{11}$$

THEOREM 2. *The two-variable association of the univariate transfer function $(s_1 I - A)^{-1} b$ is simply b, or*

$$\mathscr{A}_2\left((s_1 I - A)^{-1} b\right) = b. \tag{12}$$

PROOF. We apply (6) to find the associated time function

$$\begin{aligned}&\frac{1}{(2\pi j)^2}\int_{\sigma_2-j\infty}^{\sigma_2+j\infty}\int_{\sigma_1-j\infty}^{\sigma_1+j\infty}(s_1 I - A)^{-1} b e^{s_1 t_1} e^{s_2 t_2} ds_1 ds_2\\ =&\left(\frac{1}{2\pi j}\int_{\sigma_1-j\infty}^{\sigma_1+j\infty}(s_1 I - A)^{-1} b e^{s_1 t_1} ds_1\right)\cdot\\ &\qquad \left(\frac{1}{2\pi j}\int_{\sigma_2-j\infty}^{\sigma_2+j\infty} e^{s_2 t_2} ds_2\right) = e^{At_1} b\delta(t_2).\end{aligned} \tag{13}$$

Setting $t_1 = t_2 = t$ and taking Laplace transform again, the proof follows from the sieving property of the delta function. □

With the above properties in place, we are ready to derive the key results of this paper. Using the growing exponential (also called harmonic probing) method [15], the first three transfer functions of the QLDAE (2) are

$$H_1(s) = (sI - G_1)^{-1} b, \tag{14a}$$

$$\begin{aligned}H_2(s_1, s_2) = \frac{1}{2}\left((s_1 + s_2)I - G_1\right)^{-1}\{G_2\left[H_1(s_1) \otimes H_1(s_2)\right.\\ \left.+H_1(s_2) \otimes H_1(s_1)\right] + D_1\left(H_1(s_1) + H_1(s_2)\right)\},\end{aligned} \tag{14b}$$

$$\begin{aligned}H_3(s_1, s_2, s_3) = \frac{1}{3}\left((s_1 + s_2 + s_3)I - G_1\right)^{-1}\cdot\\ \{G_2\left[H_1(s_1) \otimes H_2(s_2, s_3) + H_2(s_2, s_3) \otimes H_1(s_1)\right.\\ + H_1(s_2) \otimes H_2(s_1, s_3) + H_2(s_1, s_3) \otimes H_1(s_2)\\ \left.+H_1(s_3) \otimes H_2(s_1, s_2) + H_2(s_1, s_2) \otimes H_1(s_3)\right]\\ +D_1\left[H_2(s_1, s_2) + H_2(s_1, s_3) + H_2(s_2, s_3)\right]\}.\end{aligned} \tag{14c}$$

Taking the G_2 part of (14b) as an example and using the property in (8) and Theorem 1, we immediately get

$$\begin{aligned}&\mathscr{A}_2\left(\left[(s_1 + s_2)I - G_1\right]^{-1} G_2\left[H_1(s_1) \otimes H_1(s_2)\right.\right.\\ &\qquad \left.\left.+H_1(s_2) \otimes H_1(s_1)\right]/2\right)\\ =&(sI - G_1)^{-1} G_2\left(sI - ②G_1\right)^{-1} b^{②}.\end{aligned} \tag{15}$$

Next, using Theorem 2 the D_1 part is easily checked to be

$$\begin{aligned}&\mathscr{A}_2\left(\left[(s_1 + s_2)I - G_1\right]^{-1} D_1\left[H_1(s_1) + H_1(s_2)\right]/2\right)\\ =&(sI - G_1)^{-1} D_1 b.\end{aligned} \tag{16}$$

Subsequently, using the often used transfer function notation $\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = C(sI - A)^{-1}B + D$ to combine (15) and (16),

$$\begin{aligned}\mathscr{A}_2(H_2(s_1, s_2)) &= (sI - G_1)^{-1}\left(G_2\left(sI - ②G_1\right)^{-1} b^{②} + D_1 b\right)\\ &= \left[\begin{array}{c|c} G_1 & I_n \\ \hline I_n & 0 \end{array}\right]\cdot\left[\begin{array}{c|c} ②G_1 & b^{②} \\ \hline G_2 & D_1 b \end{array}\right]\\ &= \left[\begin{array}{cc|c} G_1 & G_2 & D_1 b \\ 0 & ②G_1 & b^{②} \\ \hline I_n & 0 & 0 \end{array}\right]\\ &= \left[\begin{array}{c|c} \tilde{G}_2 & \tilde{b}_2 \\ \hline \tilde{c}_2 & 0 \end{array}\right],\end{aligned} \tag{17}$$

where obviously $\mathscr{A}_2(H_2)$ is recast into a higher order $(n+n^2)$ linear state space. Using similar mechanism, $\mathscr{A}_3(H_3)$ can be carefully derived to be

$$\mathscr{A}_3(H_3) = (sI - G_1)^{-1}\left(G_2\tilde{H}_3(s) + D_1^2 b\right)$$

where

$$\begin{aligned}\tilde{H}_3(s) =&(I_n \otimes \tilde{c}_2)\left(sI - G_1 \oplus \tilde{G}_2\right)^{-1}(b \otimes \tilde{b}_2)\\ &+ (\tilde{c}_2 \otimes I_n)\left(sI - \tilde{G}_2 \oplus G_1\right)^{-1}(\tilde{b}_2 \otimes b),\end{aligned}$$

which can again be put into a linear state space as in (17).

2.3 Krylov subspace for NMOR

We give concise exposition regarding efficient computer implementation. To construct the projection matrix spanning the moment space of the associated $H_1(s)$, $H_2(s) =$

$\mathscr{A}_2(H_2(s_1,s_2))$ etc. (note that these are all $n \times 1$ vectors), one often resorts to finding the Krylov subspace defined as

$$\mathcal{K}_p(G_1, b) = span\left(b, G_1 b, G_1^2 b, \cdots, G_1^{p-1} b\right).$$

The subspace basis construction is popularly done through the Arnoldi iteration, e.g. [3, 9], which then results in a lower-order orthogonal NMOR projection matrix for matching the transfer function moments [7, 6]. Nonetheless, in all practical cases, it is important to respect and exploit the structures pertinent to the state-space matrices to achieve fast speed and high accuracy.

For instance, expanding (14a) differently at $s = \infty$ and $s = 0$ would invoke $K_p(G_1, b)$ and $K_p(G_1^{-1}, G_1^{-1}b)$, respectively. Not surprisingly, the latter is more accurate in matching low-pass responses, though at the expense of computing the matrix factorization (e.g., LU) of G_1 for once. Another implementation issue is in the efficient computation of Krylov subspace projector. Referring to the second last equality in (17), the direct Arnoldi process requires multiplying the large 2×2 upper triangular block matrix (or its inverse as discussed above) with a tall matrix, consuming expensive $O((n+n^2)^2)$ work. Then at the end $[I_n \quad 0]$ is left-multiplied onto the terminated iterate to reduce it to n rows. Apparently, such brute force realization results in poor algorithmic scalability.

To reduce the computational cost, an important insight is to perform an eigenspace decomposition by applying a one-time similarity transform to the state space in (17),

$$\begin{bmatrix} G_1 & G_2 \\ 0 & ②G_1 \end{bmatrix} \begin{bmatrix} I_n & \Pi \\ 0 & I_{n^2} \end{bmatrix} = \begin{bmatrix} I_n & \Pi \\ 0 & I_{n^2} \end{bmatrix} \begin{bmatrix} G_1 & 0 \\ 0 & ②G_1 \end{bmatrix}$$

where Π is solved through the Sylvester equation

$$G_1\Pi + G_2 = \Pi ②G_1$$

which is always solvable when $\lambda_i(G_1)+\lambda_j(G_1)+\lambda_k(G_1) \neq 0$, $i,j,k = 1, \cdots, n$, where $\lambda_i(\circ)$ denotes the eigenvalue. This is always true, e.g., when G_1 is stable. Subsequently, $H_2(s)$ can be put into

$$\left[\begin{array}{cc|c} G_1 & 0 & D_1 b - \Pi b^{②} \\ 0 & ②G_1 & b^{②} \\ \hline I_n & \Pi & 0 \end{array}\right]$$
$$= (sI - G_1)^{-1}(D_1 b - \Pi b^{②}) + \Pi(sI - ②G_1)^{-1} b^{②}. \quad (18)$$

Now it becomes obvious that the Krylov subspace for $H_2(s)$ can be found from each of these subsystems. In general, a similar procedure produces k subsystems in $H_k(s)$, implying that *parallelization* is feasible for such Krylov subspace generation from distinct subsystems.

A final note is on accelerating the matrix inversion, say, in the multiplication of $(②G_1)^{-1}$ onto a vector when computing the Krylov subspace expanded at $s = 0$ in (18). The trick is to first factor G_1 into a convenient form. For example, suppose the Schur form of $G_1 = QRQ^T$ whereby Q is unitary and R is quasi (upper) triangular [3], then $②G_1 = Q^{②}(②R)(Q^{②})^T$ so that every inversion requires essentially a backward solve as $②R$ is quasi triangular, too.

Further results on optimized computer implementation, however, are beyond the scope of this work and would be reported elsewhere.

3. NUMERICAL EXPERIMENTS

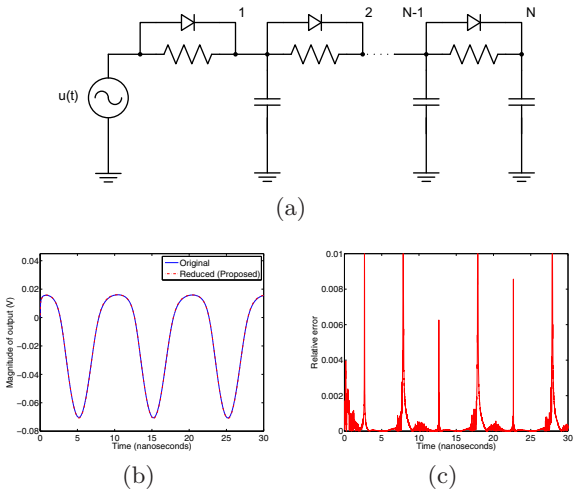


Figure 2: A nonlinear transmission line circuit with voltage source. (a) Circuit schematic. (b) Transient responses. (c) Relative errors.

In this section, the proposed associated transform-based NMOR method is applied to the following cases: QLDAEs with and without the D_1 term, multi-input single-output (MISO) QLDAE and ODE with a cubic term. All experiments are performed on a platform of Intel Pentium 4 with 2.8GHz CPU and 2GB RAM.

3.1 QLDAE with D_1 term

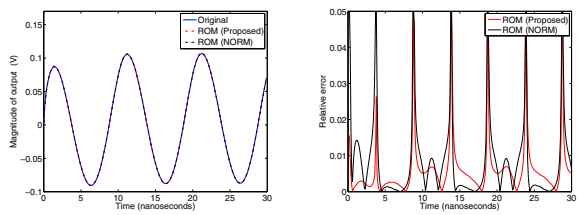
We first try the nonlinear transmission line circuit common to many NMOR papers, with a minor modification made to the signal source. As shown in Fig. 2(a), a voltage source is injected into the circuit consisting of 100 stages. All resistors and capacitors are set to 1. The I-V characteristic of the diodes is $i_D = e^{40v_D} - 1$, which has been quadratic-linearized. Using modified nodal analysis (MNA), the circuit can be characterized by a QLDAE of (2). The full model is reduced to a 13th-order ROM by the proposed associated transform approach, with moment matching up to 6 moments of $H_1(s)$, 3 moments of $H_2(s)$ and 2 moments of $H_3(s)$ (by similar order selection as in [7, 6]). The transient simulations of the full model and ROM are shown in Fig. 2(b), and the relative errors in Fig. 2(c). This immediately validates the accuracy of the proposed association of variables NMOR scheme.

3.2 QLDAE without D_1 term

If the above circuit is injected with a current source instead of the voltage source, the resulting QLDAE equation does not have the D_1 term and the final characteristic equation has a form of

$$C\dot{x} = G_1 x + G_2 x^{②} + u(t)$$

with $x \in \mathbb{R}^{70}$. Compared with NORM [7, 6] which results in a ROM of order 20, the proposed NMOR method requires only 9 to match the same number of moments. Though the NMOR time is longer in the proposed scheme due to the larger-size matrix-vector multiplication in Arnoldi iteration, such MOR process is done *only for once* and the more compact nonlinear ROM (*to be used repeatedly*) brings about a 61% reduction in simulation time compared to the

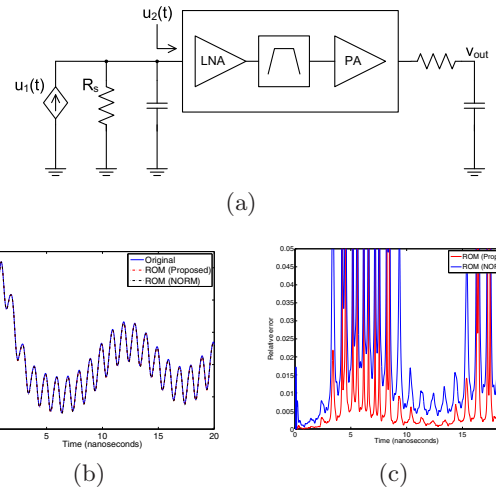


(a) (b)

Figure 3: A nonlinear transmission line circuit with current source. (a) Transient responses. (b) Relative errors.

Table 1: Runtime comparison between the proposed method and NORM

	Original	Reduced (Proposed)	Reduced (NORM)
Sect. 3.2 Ex.			
Arnoldi	—	268s	88s
ODE solve	2723s	649s	1663s
Sect. 3.3 Ex.			
Arnoldi	—	159s	72s
ODE solve	1876s	182s	381s



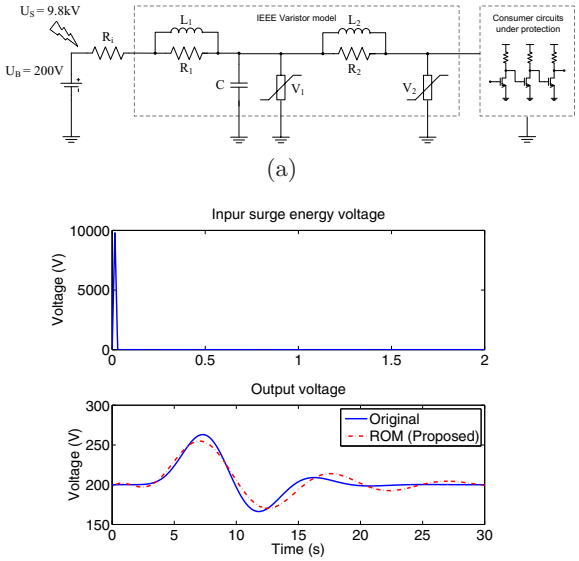
(a) (b) (c)

Figure 4: An example of an RF receiver system. (a) Block diagram. (b) Transient responses. (c) Relative errors.

NORM-reduced ROM, as recorded in Table 1. The transient responses are plotted in Fig. 3(a) and the nonlinear ROM from association of variables has almost the same accuracy as that produced by NORM.

3.3 MISO QLDAE

In Fig. 4(a), an RF receiver system with an input signal $u_1(t)$ is interfered by a noise signal $u_2(t)$ coupled from external environment. Considering both the signal and noise sources, the model of this system can be described in an MISO QLDAE form of (2) with $D_1 = 0$. In this experiment,



(a) (b)

Figure 5: An example of a ZnO varistor protection circuit. (a) Equivalent circuit. (b) Transient responses.

with the same moment matching orders, the original full system has 173 voltage/current unknowns and is reduced to 14 and 27 states with the proposed and NORM methods, respectively. Transient responses and relative errors of original model and ROMs are shown in Figs. 4(b) and 4(c). Timing data are also listed in Table 1, with similar observations as in the previous example.

3.4 ODE with cubic terms

The proposed NMOR method not only applies to the QLDAEs, but is also applicable to other forms of ODEs. In this example, another industrial example verifies the feasibility of the proposed NMOR framework for a nonlinear system with a cubic term. In Fig. 5(a), a surge protection circuit by ZnO varistors is described by an ODE system with a cubic Kronecker product

$$C\dot{x} + G_1 x + G_3 x^{③} = u.$$

The ODE has 102 states which is reduced to only 8 by the proposed method. A sudden, high voltage pulse is fed into the system and generates the dynamic response as shown in Fig. 5(b). Again, a close match in the responses is obtained even with such a low-order ROM.

4. DISCUSSION AND REMARKS

- Compared to the classical Krylov-based NORM approaches [7, 6, 4, 5], the proposed NMOR method via association of variables enjoys a much more compact ROM while matching the same order of moments. For example, for a ROM to preserve up to k_1, k_2 and k_3th-order moments in the first-, second- and third-order transfer functions, the size of the projection matrix of the proposed scheme has a dimension of $\mathcal{O}(k_1 + k_2 + k_3)$, in contrast to the much higher $\mathcal{O}(k_1 + k_2^3 + k_3^4)$

in NORM. Moreover, automatic selection of moment numbers in $H_1(s)$, $H_2(s)$, $H_3(s)$ etc. can utilize the Hankel singular values or similar measure inherent to linear MOR [11], again in contrast to the *ad hoc* order choice in NORM.

- As briefly mentioned, a singular C in (1) can proceed with the regular part extraction (viz. the "differential" or ODE part) with respect to the tuple (C, G_1). This can be done by Weierstrass canonical transform or the descriptor-system projector technique [16] taking advantages of circuit structures. In physical circuits, the decoupled "algebraic" part can often be easily handled as they are either immaterial or proportionally related to the regular subsystem.

- Non-DC or multipoint frequency expansion for moment matching is particularly straightforward with this associated transform approach. The resultant Volterra transfer functions all contain one single s and thereby practice from linear system theory follows. Moreover, the decomposition of $H_2(s)$ in (17) or (18) into the cascade of two LTI systems, and $H_3(s)$ into three etc., allows insightful interpretation of stability and passivity of the original nonlinear model. To our knowledge, this kind of simplicity has not appeared in the NMOR literature before.

Results along the 2nd and 3rd bullet points, however, are outside the focus of this paper and would be reported in a separate work.

5. CONCLUSION

This paper has presented an elegant approach for highly efficient Volterra-based NMOR. For the first time, the associated transform is extended to MIMO transfer functions to reduce the high-order multi-frequency-parameter transfer functions into standard single-frequency linear state spaces. Subsequently, linear MOR techniques can be directly utilized for NMOR. Such approach gives rise to remarkable savings in computation and memory, and produces much more compact models than existing NMOR schemes.

6. REFERENCES

[1] C. F. Chen and R. F. Chiu. New theorems of association of variables in multiple dimensional Laplace transform. *Int. J. Systems Sci.*, 4(4):647–664, 1973.

[2] J. Debnath and N. C. Debnath. Associated transforms for solution of nonlinear equations. *Intl. J. Math. & Math. Sci.*, 14(1):177–190, 1991.

[3] G. Golub and C. V. Loan. *Matrix Computations*. JohnsHopkins Univ. Press, Baltimore, 3rd edition, 1989.

[4] C. Gu. QLMOR: a new projection-based approach for nonlinear model order reduction. In *Proc. Int. Conf. Computer Aided Design*, pages 389–396, Nov. 2009.

[5] C. Gu. QLMOR: a projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 30(9):1307–1320, Sept. 2011.

[6] P. Li and L. Pileggi. Compact reduced-order modeling of weakly nonlinear analog and RF circuits. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 23(2):184–203, Feb. 2005.

[7] P. Li and L. T. Pileggi. NORM: compact model order reduction of weakly nonlinear systems. In *DAC*, pages 472–477, 2003.

[8] J. K. Lubbock and V. S. Bansal. Multidimensional Laplace transforms for solution of nonlinear equations. *Proc. IEE*, 116(12):2075–2082, Dec. 1969.

[9] A. Odabasioglu, M. Celik, and L. T. Pileggi. PRIMA: Passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 17(8):645–654, Aug. 1998.

[10] J. R. Phillips. Projection-based approaches for model reduction of weakly nonlinear, time-varying systems. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 22(2):171–187, Feb. 2003.

[11] J. R. Phillips, L. Daniel, and L. M. Silveira. Guaranteed passive balancing transformations for model order reduction. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 22(8):1027–1041, Aug. 2003.

[12] L. Pillage and R. Rohrer. Asymptotic waveform evaluation for timing analysis. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 9:352–366, Apr. 1990.

[13] D. C. Reddy and N. C. Jagan. Multidimensional transforms: new technique for the association of variables. *Electron. Lett.*, 7(10):278–279, May 1971.

[14] M. Rewienski and J. White. A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 22(2):155–170, Feb. 2003.

[15] W. Rugh. *Nonlinear System Theory – The Volterra-Wiener Approach*. Baltimore, MD: Johns Hopkins Univ. Press, 1981.

[16] Z. Zhang and N. Wong. An efficient projector-based passivity test for descriptor systems. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 29(8):1034–1042, Aug. 2010.