NORM: Compact Model Order Reduction of Weakly Nonlinear Systems

Peng Li and Lawrence T. Pileggi
Department of ECE, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA 15213
{pli, pileggi}@ece.cmu.edu

ABSTRACT

This paper presents a compact **N**onlinear model **O**rder **R**eduction **M**ethod (NORM) that is applicable for time-invariant and time-varying weakly nonlinear systems. NORM is suitable for reducing a class of weakly nonlinear systems that can be well characterized by low order Volterra functional series. Unlike existing projection based reduction methods [6]-[8], NORM begins with the general matrix-form Volterra nonlinear transfer functions to derive a set of *minimum* Krylov subspaces for order reduction. Direct moment matching of the nonlinear transfer functions by projection of the original system onto this set of minimum Krylov subspaces leads to a significant reduction of model size. As we will demonstrate as part of our comparison with existing methods, the efficacy of model order for weakly nonlinear systems is determined by the extend to which models can be reduced. Our results further indicate that a multiple-point version of NORM can substantially reduce the model size and approach the ultimate model compactness that is achievable for nonlinear system reduction. We demonstrate the practical utility of NORM for macromodeling weakly nonlinear RF circuits with time-varying behavior.

CATEGORIES AND SUBJECT DESCRIPTORS

I.6.5 [**Simulation and Modeling**]: Model Development

GENERAL TERMS

Algorithms, Design

KEYWORDS

Nonlinear Model Order Reduction, RF Circuit Modeling

1. INTRODUCTION

Over the past decade a large body of work on model order reduction of IC interconnect has emerged from the design automation community [1]-[5]. Compared to the success of model order reduction for linear time invariant (LTI) RLC networks, the problem of reducing nonlinear systems has been less understood and explored [6]-[8][13]. There are numerous applications, however, where abstracting weakly nonlinear effects into a compact macromodel is important. For instance, in RF communication IC design there is a growing interest in modeling circuit level nonlinearities as they impact system level distortion analyses. While these circuit blocks often exhibit only weak nonlinearities, the design specifications for linearity such as IIP3 and 1dB compression point are usually extremely important and very stringent. As depicted in Fig. 1, building compact blackbox type macromodels that capture accurately nonlinear input-output relationships not only facilitates efficient simulations of the circuit component being modeled, but also facilitates entire-system verification and design trade-off analyses that would otherwise be impossible.

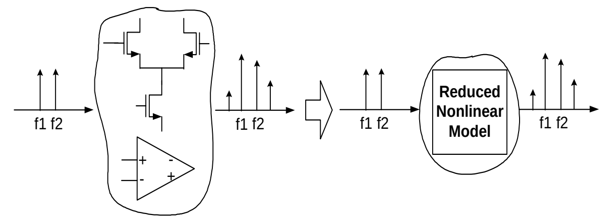

Fig. 1. Model order reduction of nonlinear analog circuits

Trajectory piecewise-linear approximation has been recently proposed for nonlinear system reduction, but it is limited by a system training input signal dependency [13]. In a different direction, Volterra functional series (e.g. [9][10]) have been used in symbolic block-diagram type system-level model generation [14], and for projection based nonlinear model order reduction [6]-[8]. One advantage of using Volterra series is that it provides a canonical characterization of weakly nonlinear systems in the form of Volterra kernels or nonlinear transfer functions. In particular, the method proposed in [6][8] makes automated nonlinear system reduction possible by extending the popular projection based techniques used for interconnects, such as PVL[2], Arnoldi method[3] and PRIMA[4].

Modeling a weakly nonlinear system as a set of linear networks via Volterra theory facilitates the direct application of the established projection based order reduction techniques. However, such an approach leaves several important issues unaddressed: How does the quality of each linear reduction problem impact the accuracy of the overall nonlinear model? What is the best strategy in carrying out each linear reduction for achieving good nonlinear model accuracy? Another interesting approach uses the bilinear form of a nonlinear system allowing an elegantly derived projection based moment-matching reduction scheme [7]. The critical issue associated with this approach is the explosion of state variables in a bilinear form; e.g., the bilinear form of a system with cubic nonlinearities has $O(N^3)$ state variables, where N is the number of state variables in the original state-equations.

In this paper, we further explore the projection based reduction framework laid out in [6]-[8]. As shown in the paper, model compactness is a particularly critical factor for effective reduction since it will become increasingly more costly to form the reduced model explicitly as the model size increases. Therefore, we propose an approach that begins with the most general matrix-form nonlinear transfer functions needed for model order reduction, and derive the expressions for nonlinear transfer function moments. This development leads to a deeper understanding on the interaction of Krylov subspace projection and the moment matching under nonlinear context, and thereby allows us to address some of the open questions created by the work in [6][8]. Most importantly, from this development we propose a new reduction scheme, NORM, which dramatically reduces the size of reduced order models and copes with the model growth problem for nonlinear system representation by using a set of minimum Krylov subspaces. We demonstrate the excellent accuracy of reduced-order models generated via NORM with orders of magnitude runtime speedup in harmonic balance simulations for several circuit examples.

2. EXISTING PROJECTION BASED APPROACHES

Under certain conditions, a weakly nonlinear system can be analyzed using Volterra functional series [9][10] in which the response $x(t)$ is expressed as a sum of responses at different orders: $x(t) = \sum_{n=1}^{\infty} x_n(t)$, where, x_n is the n-th order response. As an extension to the use of impulse response function in a LTI system, n-th order response can be related to *Volterra kernel of order n* $h_n(\tau_1, \ldots, \tau_n)$ by

$$x_n(t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) u(t-\tau_1) \ldots u(t-\tau_n) d\tau_1 \ldots d\tau_n . \quad (1)$$

The Laplace transform of the n-th order Volterra kernels $H_n(s_1, \ldots, s_n)$ is called as *nonlinear transfer function of order n.* Notice that Volterra kernels and nonlinear transfer functions are system properties independent of the system inputs, and capable of fully describing the weakly nonlinear behavior of the system.

For simplicity, consider the expanded state-equation of a weakly nonlinear single-input multiple-output (SIMO) system at its bias point (keeping only small-signal terms)

$$\frac{dx}{dt} = A_1 x + A_2(x \otimes x) + A_3(x \otimes x \otimes x) + \ldots + bu , \quad (2)$$

where Kronecker tensor products of state variable vector x are used. Applying Volterra functional analysis on (2), the first order through the third order nonlinear responses are recursively given by

$$\dot{x_1} = A_1 x_1 + bu , \quad (3)$$

$$\dot{x_2} = A_1 x_2 + A_2(x_1 \otimes x_1) , \quad (4)$$

$$\dot{x_3} = A_1 x_3 + A_2(x_1 \otimes x_2 + x_2 \otimes x_1) + A_3(x_1 \otimes x_1 \otimes x_1) . \quad (5)$$

The key idea in [6][8] is that (3)-(5) represent a set of LTI systems which can be in turn reduced by applying any existing projection methods for LTI systems [2]-[4]. The overall reduced nonlinear system can be expressed by a smaller set of nonlinear equations using the variable embedding $x = V \cdot \bar{x}$, where V is an orthonormal basis of $[V_1, V_2, V_3]$, V_1, V_2 and V_3 are the projections used for reducing (3)-(5), respectively. For instance, the third order matrix can be reduced to a matrix of a smaller dimension as $\overline{A_3} = VA_3(V \otimes V \otimes V)$. However, the reduced third order matrix is usually dense and has $O(q^4)$ entries, where q is the number of states of the reduced model. Hence, to void forming large reduced high order matrices, controlling the dimension of reduced model is crucial. Essentially, this method models a nonlinear system, possibly with small number of inputs, as several linear systems with many more inputs. Intuitively, this increase of degree of freedom leads to unnecessarily large reduced models. To asses the quality of this reduction approach and suggest strategies for optimizing model compactness, a careful moment analysis is required as shown in the following sections.

3. GENERAL MATRIX-FORM VOLTERRA NONLINEAR TRANSFER FUNCTIONS

Without loss of generality consider the Modified Nodal Analysis formulation of a SIMO weakly nonlinear system with n unknowns

$$f(x(t)) + \frac{d}{dt} q(x(t)) = bu(t), \; y(t) = d^T x(t) \quad (6)$$

where x and y are the vectors of circuit unknowns and outputs, u is the input, $f(\cdot)$ and $q(\cdot)$ are functions representing resistive and dynamic nonlinearities, b and d are the input and output matrices, respectively. First consider a small perturbation of (6) around a DC operating point x_0 (this can be generalized to a perturbation along a time-varying operating point). Expand $f(\cdot)$ and $q(\cdot)$ at the bias point and consider only small-signal quantities, we have

$$\frac{d}{dt}(C_1 x + C_2(x \otimes x) + C_3(x \otimes x \otimes x) + \ldots) \quad , \quad (7)$$
$$+ G_1 x_1 + G_2(x \otimes x) + G_3(x \otimes x \otimes x) + \ldots = bu(t)$$

where $G_i = \left.\frac{\partial^i f}{\partial x^i}\right|_{x = x_0} \in R^{n \times n^i}$, $C_i = \left.\frac{\partial^i q}{\partial x^i}\right|_{x = x_0} \in R^{n \times n^i}$ are the ith order conductance and capacitance matrices respectively. Before proceeding we first introduce the notations used throughout this paper. For matrices in (7) we define

$$A = -G_1^{-1} C_1, r_1 = G^{-1} b, r_2 \equiv r_1 \otimes r_1, r_3 \equiv r_1 \otimes r_1 \otimes r_1 , \quad (8)$$

and for an arbitrary matrix F define

$$F^{l \otimes m \otimes \ldots \otimes n} \equiv F^l \otimes F^m \otimes \ldots \otimes F^n, \overline{F^{l \otimes m}} \equiv \frac{1}{2}(F^{l \otimes m} + F^{m \otimes l})$$
$$\overline{F^{l \otimes m \otimes n}} \equiv \frac{1}{6}(F^{l \otimes m \otimes n} + F^{l \otimes n \otimes m} + F^{m \otimes l \otimes n} \quad (9)$$
$$+ F^{m \otimes n \otimes l} + F^{n \otimes l \otimes m} + F^{n \otimes m \otimes l})$$

Also define the Krylov subspace $K_m(A, p)$ corresponding to matrix A and vector (matrix) p as the space spanned by vectors $\{p, Ap, \ldots, A^{m-1}p\}$.

For the system in (7), the first order transfer function for state-variables x is simply the transfer function of the linearized system

$$(G_1 + sC_1)H_1(s) = b \text{, or } H_1(s) = (G_1 + sC_1)^{-1} b . \quad (10)$$

Defining $\bar{s} = s_1 + s_2$, the second order transfer function is given by

$$[G_1 + \bar{s}C_1]H_2(s_1, s_2) = -[G_2 + \bar{s}C_2] \cdot \overline{H_1(s_1) \otimes H_1(s_2)} , \quad (11)$$

where $\overline{H_1(s_1) \otimes H_1(s_2)} = \frac{1}{2}(H_1(s_1) \otimes H_1(s_2) + H_1(s_2) \otimes H_1(s_1))$.

Similarly define $\overline{H_1(s_1) \otimes H_1(s_2) \otimes H_1(s_3)}$ as the arithmetic average of terms of all possible permutations of frequency variables in the Kronecker product, and define $\tilde{s} = s_1 + s_2 + s_3$, then the third order nonlinear transfer function is given by the following equations

$$[G_1 + \tilde{s}C_1]H_3(s_1, s_2, s_3) = -[G_3 + \tilde{s}C_3] \cdot \quad (12)$$
$$\overline{H_1(s_1) \otimes H_1(s_2) \otimes H_1(s_3)} - 2 \cdot [G_2 + \tilde{s}C_2] \cdot \overline{H_1(s_1) \otimes H_2(s_2, s_3)}$$

$$\overline{H_1(s_1) \otimes H_2(s_2, s_3)} = \frac{1}{6}(H_1(s_1) \otimes H_2(s_2, s_3) + H_1(s_2) \otimes H_2(s_1, s_3)$$
$$+ H_1(s_3) \otimes H_2(s_1, s_2) + H_2(s_1, s_2) \otimes H_1(s_3) \quad (13)$$
$$+ H_2(s_1, s_3) \otimes H_1(s_2) + H_2(s_2, s_3) \otimes H_1(s_1))$$

Without loss of generality, expand (10) at the origin as a Mclaurin series

$$H_1(s) = \sum_{k=0}^{\infty} s^k A^k r_1 = \sum_{k=0}^{\infty} s^k M_{1,k} , \quad (14)$$

where $M_{1,k} = A^k r_1$ is the kth order moment of the first order transfer function. Expand $H_2(s_1, s_1)$ at the origin $(0, 0)$

$$H_2(s_1, s_2) = \sum_{k=0}^{\infty} \sum_{l=0}^{k} s_1^l s_2^{k-l} M_{2,k,l} , \quad (15)$$

where $M_{2,k,l}$ is a kth order moment for the second order transfer function. To derive the expressions for the moments of $H_2(s_1, s_2)$, substitute (14) into (11) and expand w.r.t. $\bar{s} = s_1 + s_2$

$$H_2(s_1, s_2) = -\sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \bar{s}^m s_1^k s_2^l A^m G_1^{-1} \cdot (G_2 + \bar{s}C_2)\overline{A^{k \otimes l}} r_2 \quad (16)$$

Comparing (15) with (16), we can express the moments of

$H_2(s_1, s_2)$ in (18) at the bottom of the page. Similarly, the third order transfer function can be expanded at the origin $(0, 0, 0)$ as

$$H_3(s_1, s_2, s_3) = \sum_{k=0}^{\infty}\sum_{l=0}^{k}\sum_{m=0}^{k-l} s_1^l s_2^m s_3^{k-l-m} M_{3,k,l,m}, \qquad (19)$$

where $M_{3,k,l,m}$ is a *k*th order moment. Due to the limit of space, we omit the expression of $M_{3,k,l,m}$. Also note that in (11) and (12), symmetrized nonlinear transfer functions (symmetric w.r.t frequency variable permutations) are used. This reduces the number of moments that need to be considered for H_2 and H_3 by approximately a factor of 2 and 6 respectively when expanding them at a point with equal coordinates such as the origin.

4. NORM

In this section we outline our proposed algorithm, NORM, but limit our discussion to SIMO time-invariant weakly nonlinear systems. We consider problems that can be well modeled by up to third order nonlinear transfer functions. Extensions to more general scenarios such as time-varying systems, can be derived in a very similar fashion. To assess the model order reduction quality from a moment matching perspective, we begin with the following definition:

Definition 1 A nonlinear reduced order model is a *k*th order model in $H_1(s)$ ($H_2(s_1, s_2)$ or $H_3(s_1, s_2, s_3)$) if and only if up to *k*th order moments $M_{1,l}$, $0 \le l \le k$, ($M_{2,l,m}$, $0 \le l \le k, 0 \le m \le l$ or $M_{3,l,m,n}$, $0 \le l \le k, 0 \le m \le l, 0 \le n \le l-m$) of the first (second or third) order transfer function of the original system defined in (14) ((15) or (19)) are preserved in the reduced model.

According to definition 1, a 2nd order reduced model in H_2 preserves moments of H_2 corresponding to the coefficients of terms $s_1^0 (s_2^0)$, s_1, s_2, $s_1 s_2$, s_1^2 and s_2^2 in the expansion.

4.1 Single-Point Expansion

To derive a set of minimum Krylov subspaces for the most compact order reduction, understanding the interaction between the moments of nonlinear transfer functions at different orders is extremely important. For example, for the moment matching of $H_2(s_1, s_2)$, this interaction is manifested in (16). A particular term corresponding to $s_1^p s_2^q$ in the expansion of H_2, where p, q are integers, is a consequence of the power series expansion of H_1 in (14) and the expansion of (11) w.r.t. $\bar{s} = s_1 + s_2$. As such, the final expression for $M_{2,l,m}$ is in the form as shown in (18). The expression for $M_{3,l,m,n}$ is derived similarly in a more complex form, which we omit here due to the space limitation. If we were to use a projection for order reduction, only Krylov subspaces containing the directions of moments to be matched are required. A close inspection of (18) reveals that the Krylov subspaces of matrix A given in Table 1 are the desired Krylov subspaces of the minimum total dimension for constructing a *k*th order model in H_2. In the table, both the starting vector and the dimension of each subspace are shown. For the last row, $m = \lfloor k/2 \rfloor$, $n = k - \lfloor k/2 \rfloor$, $p = \lfloor (k-1)/2 \rfloor$ and $q = k - 1 - \lfloor (k-1)/2 \rfloor$. We denote the union of Krylov subspaces in Table 1 as $K2(k) = \bigcup_i K_{m_i}(A, v_i)$, where k in the parenthesis indicates the order of H_2 moments up to which contained in the subspaces. In an analogous and somewhat more involved way, a set of minimum Krylov subspaces $K3(k)$ for moment-matching of H_3 up to *k*th order can be also derived. Using these subspaces we can rigorously prove the following:

Starting vector v_i	Subspace Dim m_i	Starting vector v_i	Subspace Dim m_i
$G_1^{-1}G_2\overline{A^{0\otimes 0}}r_2$	$k+1$	$G_1^{-1}C_2\overline{A^{0\otimes 0}}r_2$	k
$G_1^{-1}G_2\overline{A^{1\otimes 0}}r_2$	k	$G_1^{-1}C_2\overline{A^{1\otimes 0}}r_2$	$k-1$
…	…	…	…
$G_1^{-1}G_2\overline{A^{k\otimes 0}}r_2$	1	$G_1^{-1}C_2\overline{A^{(k-1)\otimes 0}}r_2$	1
$G_1^{-1}G_2\overline{A^{1\otimes 1}}r_2$	$k-1$	$G_1^{-1}C_2\overline{A^{1\otimes 1}}r_2$	$k-2$
$G_1^{-1}G_2\overline{A^{2\otimes 1}}r_2$	$k-2$	$G_1^{-1}C_2\overline{A^{2\otimes 1}}r_2$	$k-3$
…	…	…	…
$G_1^{-1}G_2\overline{A^{(k-1)\otimes 1}}r_2$	1	$G_1^{-1}C_2\overline{A^{(k-2)\otimes 1}}r_2$	1
…	…	…	…
$G_1^{-1}G_2\overline{A^{m\otimes n}}r_2$	1	$G_1^{-1}C_2\overline{A^{q\otimes p}}r_2$	1

Table 1. Krylov subspaces containing the moments of H_2

Theorem 1 If $V = qr([K_{k_1+1}(A, r_1), K2(k_2), K3(k_3)])$, where $k_1 \ge k_2 \ge k_3$, is an orthonormal basis for the union of subspaces $K_{k_1+1}(A, r_1)$, $K2(k_2)$ and $K3(k_3)$ then for (7) the reduced order model specified by the following system matrices

$$\tilde{G}_1 = V^T G_1 V,\ \tilde{C}_1 = V^T C_1 V,\ \tilde{b} = V^T b,\ \tilde{d} = V^T d$$
$$\tilde{G}_2 = V^T G_2 (V \otimes V),\ \tilde{C}_2 = V^T C_2 (V \otimes V),$$
$$\tilde{G}_3 = V^T G_3 (V \otimes V \otimes V),\ \tilde{C}_3 = V^T C_3 (V \otimes V \otimes V) \qquad (20)$$

is a k_1th order model in H_1, a k_2th order model in H_2 and a k_3th order model in H_3.

In theorem 1, $qr(.)$ refers to the QR procedure used for othonormalizing the input vectors. We omit the proof of the theorem 1 and the derivation of subspaces $K3(k)$, but outline the complete single point version of NORM algorithm (expanded at the origin) in Fig. 2. For step 3.1,

$$\overline{(A^l r_1) \otimes M_{2,(m,n)}} = \frac{1}{3}((A^l r_1) \otimes M_{2,m+n,n} + (A^m r_1) \otimes M_{2,l+n,n} + (A^n r_1) \otimes M_{2,l+m,l} + M_{2,m+n,n} \otimes (A^l r_1) + M_{2,l+n,n} \otimes (A^m r_1) + M_{2,l+m,l} \otimes (A^n r_1)) \qquad (21)$$

Note that in Fig. 2, to compute any Kronecker product term, one can exploit the original problem sparsity such that the computation takes only a linear time in the problem size. The requirement $k_1 \ge k_2 \ge k_3$ is due to the dependence of high order nonlinear transfer functions on transfer functions of lower orders. Provided this condition is satisfied, the order of moment matching for each nonlinear transfer function can be flexibly chosen to fit specific needs. For each of these choices, a reduced order model of minimum size is produced as described in the following section.

4.2 Size of Reduced Order Models

Without causing ambiguity, we define the size of a state-equation model in terms of its number of states. Consider the size of the SIMO based reduced order models generated using the method of [6][8] under definition 1. It can be shown that if the linear networks described by (3)-(5) are each reduced to a system preserving up to k_1, k_2 and k_3th order moments of the original system respectively, then the overall reduced nonlinear model is a k_1th order model in H_1, a $\overline{k_2}$th order model in H_2, and a $\overline{k_3}$th order

$$M_{2,k,l} = -\sum_{p=1}^{k} A^{p-1} \sum_{q=0,\, q\le l,\, p-q\le k-l}^{p} \binom{p}{q} G_1^{-1} C_2 \cdot \overline{A^{(l-q)\otimes(k-l-p+q)}} r_2 - \sum_{p=0}^{k} A^p \sum_{q=0,\, q\le l,\, p-q\le k-l}^{p} \binom{p}{q} G_1^{-1} G_2 \cdot \overline{A^{(l-q)\otimes(k-l-p+q)}} r_2 \qquad (18)$$

Input: $(G_1, C_1, G_2, C_2, G_3, C_3, b, d, k_1, k_2, k_3), k_1 \ge k_2 \ge k_3$

1. Compute a Krylov subspace of A to match up to k_1 th order moments of H_1: $V_1 \leftarrow qr(K_{k_1+1}(A, r_1))$.

2. Compute the following Krylov subspaces to match the moments of H_2 up to k_2 th order: $V_2 \leftarrow [\ \ ]$;

2.1. For each $m \ge 0, n \ge 0, m \le n, m + n \le k_2$:

$$v = G_1^{-1} G_2 \overline{A^{m \otimes n}} r_2, V_2 \leftarrow [V_2, K_{k_2 - m - n + 1}(A, v)]$$

2.2. For each $m \ge 0, n \ge 0, m \le n, m + n \le k_2 - 1$:

$$v = G_1^{-1} C_2 \overline{A^{m \otimes n}} r_2, V_2 \leftarrow [V_2, K_{k_2 - m - n}(A, v)]$$

3. To match the moments of H_3 up to k_3 th order compute the following Krylov spaces: $V_3 \leftarrow [\ \ ]$;

3.1. For each $m \ge 0, n \ge 0, l \ge 0, m \le n \le l, m + n + l \le k_3$:

$$v = G_1^{-1} G_3 \overline{A^{l \otimes m \otimes n}} r_3, V_3 \leftarrow [V_3, K_{k_3 - m - n - l + 1}(A, v)]$$

$$v = G_1^{-1} G_2 \overline{(A^l r_1) \otimes M_{2,(m,n)}}, V_3 \leftarrow [V_3, K_{k_3 - m - n - l + 1}(A, v)]$$

3.2. For each $m \ge 0, n \ge 0, l \ge 0, m \le n \le l, m + n + l \le k_3 - 1$:

$$v = G_1^{-1} C_3 \overline{A^{l \otimes m \otimes n}} r_3, V_3 \leftarrow [V_3, K_{k_3 - m - n - l}(A, v)]$$

$$v = G_1^{-1} C_2 \overline{(A^l r_1) \otimes M_{2,(m,n)}}, V_3 \leftarrow [V_3, K_{k_3 - m - n - l}(A, v)]$$

4. $V = qr([V_1, V_2, V_3])$, $\tilde{G}_1 = V^T G_1 V$, $\tilde{C}_1 = V^T C_1 V$, $\tilde{b} = V^T b$, $\tilde{d} = V^T d$, $\tilde{G}_2 = V^T G_2 (V \otimes V)$, $\tilde{C}_2 = V^T C_2 (V \otimes V)$, $\tilde{G}_3 = V^T G_3 (V \otimes V \otimes V)$, $\tilde{C}_3 = V^T C_3 (V \otimes V \otimes V)$.

Fig. 2. Single-Point NORM

model in H_3, where $\overline{k_2} = min(k_1, k_2)$, $\overline{k_3} = min(k_1, k_2, k_3)$. For instance, as can be seen from (11)(16), the k th order moments of H_2 depend on moments of H_1 with an order equal or less than k. Therefore, the lesser of k_1 and k_2 determines the order of moment matching for $H_2(s_1, s_2)$. It follows that for this method the most compact k th order model in H_2 with a size in $O(k^3)$ is achieved by choosing $k_1 = k_2 = k$. In other words, choosing $k_2 > k_1 = k$ does not necessarily increase the number of moments of H_2 matched in the reduced model. On the other hand, if one would like to have $k_1 > k_2$ to increase the accuracy of H_1, the best way is to only make use of the first $k_2 + 1$ moment directions of H_1 for reducing H_2, while the remaining H_1 moments are included in the projection only for matching H_1 itself. Similar strategies apply to reducing (5) for the moment matching of H_3 such as the optimal way to generate a k th order model in H_3 is to choose $k_1 = k_2 = k_3 = k$ with a resulting model size of $O(k^5)$. Note that these strategies are also employed in the single point NORM algorithm. To compare with the above "optimal" model sizes for the method of [6][8], using single-point NORM the sizes of a *k*th order model in H_2 and H_3 are in $O(k^3)$ and $O(k^4)$, respectively. The exact model sizes for several values of k is shown in the following section.

4.3 Multi-Point Expansions

To target a system's particular input frequency band, particularly for RF circuits, it might be desirable to expand both linear and nonlinear transfer functions at points other than the origin. It is crucial to note that under nonlinear context, a single expansion point may inherit multiple matrix factorizations due to the nonlinear frequency mixing effects. Suppose that the inband third order intermodulation of a nonlinear system around center frequency f_o is important to model. To build the most compact model, one would opt to expand $H_1(s)$ at $s = j2\pi f_0$, but to correctly perform moment matching for H_2 and H_3, the expansion points should be $(j2\pi f_0, j2\pi f_0)$ and $(j2\pi f_0, -j2\pi f_0)$ for H_2, and $(j2\pi f_0, j2\pi f_0, -j2\pi f_0)$ for H_3, respectively. Here the use of two expansion points for H_2 takes care of matching second order mixing effects in terms of both sum and difference frequencies around the center frequency, and also ensures the moment matching of the third order in-band intermodulations. This choice of expansion points requires the system matrix factorized at DC and $j4\pi f_0$ but not $j2\pi f_0$ in (11) or (4). Therefore, if every linear system in (3)-(5) is reduced by expanding at $s = j2\pi f_0$ using the method of [6][8], then there is no guarantee for matching any moments of H_2 and H_3.

In additional to the benefit of using specific expansion point, a multiple-point projection where several expansion points are used simultaneously has a unique efficiency advantage in terms of model compactness over the single-point method, i.e., adopting multiple-point methods can lead to significantly smaller reduced models. Although it is not surprising that the size of the nonlinear reduced order model grows faster than the order of moment matching, it is very revealing to recognize that the numbers of up to *k*th order moments of H_2 and H_3 are $k(k+1)/2$ and $O(k^3)$ respectively, i.e., the total dimension of the subspaces used in the single-point expansion version of NORM grows even faster than the number of moments matched[1]. However, to preserve the value of a nonlinear transfer function at a specific point (a zeroth-order moment), for example, only one vector needs to be included in the projection assuming the dependency between transfer functions of different orders is resolved properly, i.e., the reduced model size is the same as the number of moments matched.

Order K	3	4	5	6	7	8	9
[6][8]: H2	116	230	402	644	968	1386	1910
NORM(sp):H2	24	40	62	91	128	174	230
NORM(mp):H2	16	23	30	39	48	59	70
[6][8]: H2&3	3,700	11,480	28,914	63,070	123,848	224,460	381,910
NORM(sp):H2&3	66	118	194	301	446	636	880
NORM(mp):H2&3	37	56	78	108	141	182	229

Table 2. Comparison on the Reduced Model Sizes

This suggests that a multiple-point method based on low-order moment-matching at several expansion points can trade off computational cost with model compactness. The added computational cost due to more matrix factorizations in the multiple-point method might be alleviated by exploiting the idea of recycled Krylov-subspace vectors for time-varying systems [11][12]. It is also possible to use constant Jacobian iterations for solving the resulting linear problems for narrow band systems, such as those corresponding to certain RF applications. The LU factorization at one expansion point might be reused at another point not far apart as an approximated Jacobian in the iteration. Finally, for the state-equation form of (7), we compare the worst-case reduced model sizes generated by three methods in Table 2, where each method is used to match the moments of H_2 or $H_2 \& H_3$ up to *k*th order. In the table, the "optimal" strategies outlined in previous section are used for the method of [6][8], and NORM(mp) is the "equivalent" zeroth-order

1. Using the bilinear form can produce models with a size proportional to number of moments matched, however, this may be offset by the inflated problem size and the accuracy degradation for reducing a significantly larger system.

multiple-point method preserving the same total number of moments. As clearly shown, using NORM the model compactness has been significantly improved.

5. RESULTS

We compared the method of [6][8], single-point NORM (NORM-sp) and zeroth-order multiple-point NORM (NORM-mp) on two examples. The origin was chosen as the expansion point for the first two methods. Note that the optimal strategies presented in Section 4.2 were applied to the methods in [6][8], otherwise, significantly larger models resulted with little accuracy improvement. For all the three methods, SVD was used in a post-processing step to deflate the Krylov subspaces.

5.1 A Double-balanced Mixer

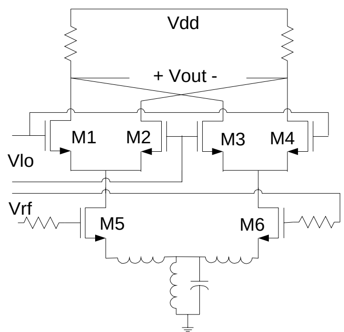

Fig. 3. A double-balanced mixer

A standard double-balanced mixer, as shown in Fig. 3, is modeled as a time-varying weakly nonlinear system with respect to the RF input. Circuit nonlinearities are modeled using third order polynomials around the time-varying operating point due to the large 830MHz LO. The full model has 2403 time-sampled circuit unknowns and characterized by time-varying Volterra series. The 60 state model generated by the method of [6][8] matches 4 moments of H_1, 2 moments of both H_2 and H_3. NORM-sp generates a model with 19 states matching 4 moments for all of H_1, H_2 and H_3. NORM-mp matches 4 moments of H_1 and H_3, 8 moments of H_2 resulting a model size of 14. As can be seen, smaller models generated by NORM can actually match more

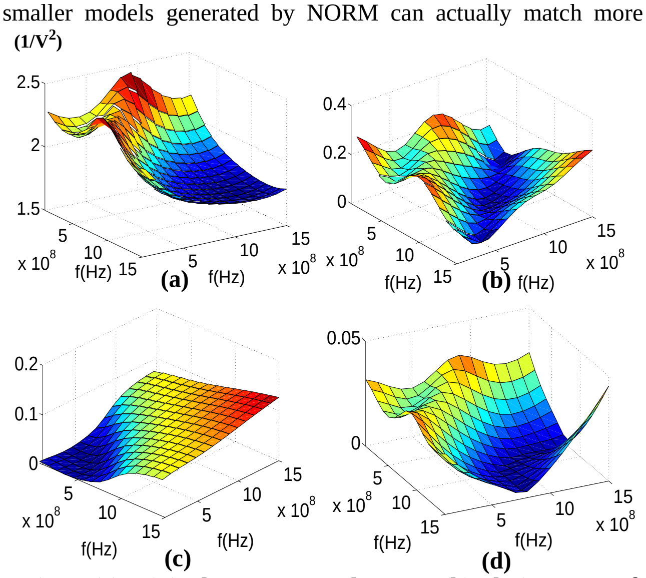

Fig. 4. (a)Original H3—2403 unknowns, (b)relative error of the method of [6][8]—60 states, (c)relative error of NORM-sp—19 states, and (d) relative error of NORM-mp—14 states

transfer function moments due to its proper selection of Krylov subspace vectors. Since the double-balanced mixer is fully symmetric, the second order transfer function H_2 is ideally zero (except for numerical noise). To see the third order intermodulation translated by one LO frequency, the corresponding harmonic of the third order time-varying nonlinear transfer function $H_3(1, j2\pi f_1, j2\pi f_2, j2\pi f_3)$ where $100MHz \le f_1, f_2 \le 1.5GHz$, $f_3 = -900MHz$ is plotted in Fig. 4(a). The maximum relative modeling errors are 27%, 13% and 4.5% respectively for the method of [6][8], NORM-sp, and NORM-mp. These models were also simulated for two-tone third order intermodulation tests using a harmonic-balance simulator as plotted in Fig. 5. We first fixed the RF input amplitude for both tones at 40mv while varying the frequency of one tone from 100MHz to 2GHz (the second tone was separated from the first one by 800KHz). Then, we fixed two tone frequencies at 600MHz and 600.8MHz respectively but varied the amplitude of the two tones from 20mv to 70mv. As can be seen from Fig. 5, the smallest model generated by NORM-mp is also the most accurate for both cases. The 60-state model generated by the method of [6][8] incurs apparent error for the first test. Also note that the amount of IM3 from the simulation is predicted accurately by the corresponding third order transfer functions.

Due to the resulting large third order matrices, the 60-state model of the method of [6][8] brought less than 5x runtime speedup over the full model for various input frequencies and amplitudes. However, for the much smaller models produced by NORM-sp and NORM-mp, significant runtime speedups of 350~380x and 720~840x were achieved respectively.

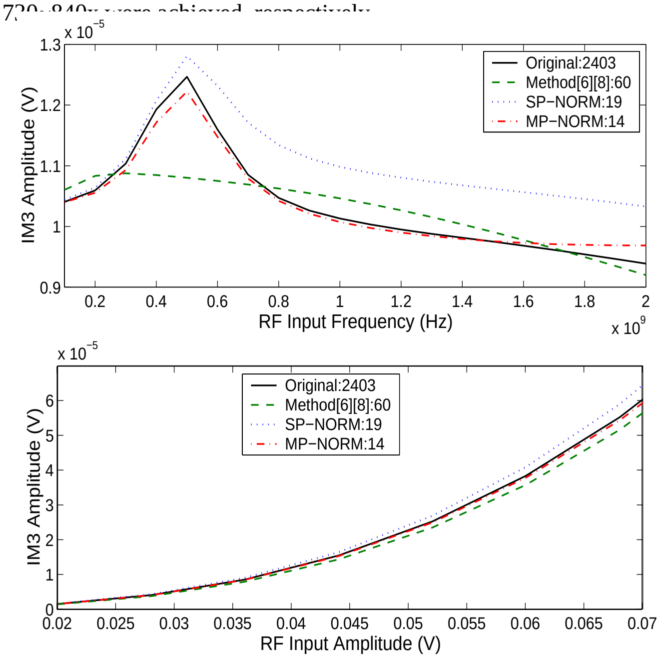

Fig. 5. IM3 as functions of input frequency and amplitude

5.2 A 2.4GHz Subharmonic Direct-conversion Mixer

A 2.4GHz subharmonic direct-conversion mixer used in WCDMA applications is shown in Fig. 6. It employs six phases of a LO signal at 800MHz to generate an equivalent LO at 2.4GHz. For direct-conversion mixers, second-order nonlinear effects are important, which exist when the perfect circuit symmetry is lost in a balanced architecture. For this example, we introduced about 2% transistor width mismatch in the circuit and applied the three methods to reduce the original time-varying system with 4130 time-sampled circuit unknowns, where each circuit nonlinearity is modeled

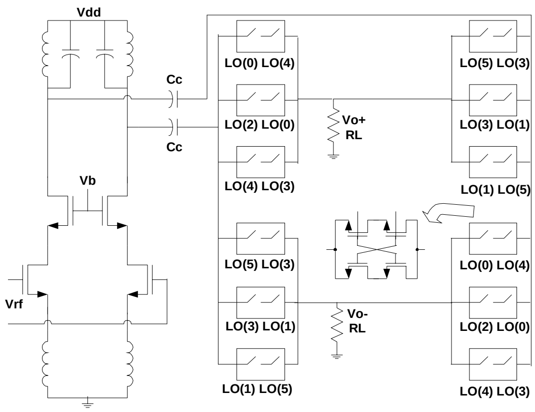

Fig. 6. A 2.4GHz subharmonic direct-conversion mixer

using a time-varying third order polynomial. The DC component of the time-varying H_2 specifying the mixing of two RF tones directly to the baseband and the corresponding relative modeling errors of the three methods are plotted in Fig. 7, where two RF frequencies vary from -2.6GHz to -2.2GHz, and from 2.2GHz to 2.6GHz respectively. The model produced by the method of [6][8] has 122 states and matches 4 moments of H_1, 6 moments of H_2, and 2 moments of H_3 with a maximum relative error about 700% or 16.9dB. NORM-sp produces a model with 34 states while matching 5 moments of H_1, 9 moments of H_2, and 2 moments of H_3. The maximum relative error for this model is about 14%. For both methods, the origin is used as the expansion point. The better accuracy obtained in the smaller model of NORM-sp can be explained by the fact that more moments are matched in the reduced model. We anticipate that both methods will generate more accurate models when the correct procedure is employed to expand transfer functions at a point close to the center frequency as outlined in Section 4.3 . Lastly, NORM-mp generates a compact 22-state model with the smallest maximum relative error of 4% while matching 6 moments of H_1, 12 moments of H_2 and 4 moments of H_3.

In a two-tone harmonic balance simulation, we applied two RF sinusoidal tones around 2.4GHz with 2mv amplitude, explicitly

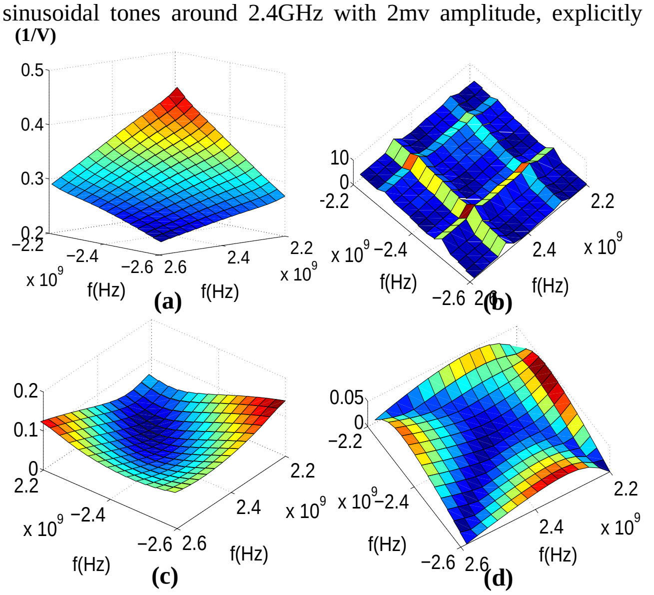

Fig. 7. (a)Original H2—4130 unknowns, (b)relative error of the method of [6][8]—122 states, (c)relative error of NORM-sp—34 states, and (d) relative error of NORM-mp—22 states

formed reduced models of NORM-sp and NORM-mp provided a runtime speedup of 237x and 1200x over the full model, respectively. Due to the large reduced third order system matrices, however, it becomes inefficient to explicitly form the 122-state model of the methods of [6][8]. Only the projection matrix was used to reduce the size of the linear problem solved at each simulation iteration. Consequently, the corresponding models did not provide a significant runtime speedup.

6. CONCLUSIONS

We have demonstrated that the rapid growth of reduced order models for nonlinear time-varying systems makes model order reduction much more difficult than for the case of linear time invariant system order reduction. The proposed nonlinear system order reduction algorithm, NORM, controls the model size growth by using a minimum set of Krylov subspace vectors. It was also shown that the use of multiple-point NORM further reduces the model size. The use of NORM algorithm allows us to tackle nonlinear reduction problems to which applying existing projection-based methods becomes ineffective.

7. ACKNOWLEDGMENTS

This work was supported by SRC under contract 2000-TJ-779. The authors would also like to thank Frank Liu and Emrah Acar of IBM Austin Research Lab for helpful discussions on using constant-Jacobian iterations for AC analysis.

8. REFERENCES

[1] L. Pillage and R. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. CAD*, vol. 9, pp. 352-366, April 1990.

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

[3] L. Silveira, M. Kamon and J. White, "Efficient reduced-order modeling of frequency-dependent coupling inductances associated with 3-d interconnect structures," *Proc. of ACM/IEEE DAC*, 1995.

[4] A. Odabasioglu, M. Celik and L. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. CAD*, vol. 17, no. 8, pp. 645-653, August 1998.

[5] J. Phillips, L. Daniel and M. Silveira, "Guaranteed passive balancing transformations for model order reduction," *Proc. of ACM/IEEE DAC*, 2002.

[6] J. Roychowdhury, "Reduced-order modeling of time-varying systems," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 10,Oct., 1999.

[7] J. Phillips, "Projection frameworks for model reduction of weakly nonlinear systems," *Proc. of ACM/IEEE DAC*, 2000.

[8] J. Phillips, "Automated extraction of nonlinear circuit macromodels," *Proc. of IEEE CICC*, 2000.

[9] D. Wiener and J. Spina, *Sinusoidal analysis and modeling of weakly nonlinear circuits*, Van Nostrand Reinhold, 1980.

[10] S. A. Maas, *Nonlinear Microwave Circuits*, Artech House, Norwood, Massachusetts, 1988.

[11] R. Telichevesky, K. Kundert and J. White, "Efficient AC and noise analysis of two-tone RF circuits," *Proc. of ACM/IEEE DAC*, 1996.

[12] J. Phillips, "Model reduction of time-varying linear systems using approximate multipoint krylov-subspace projectors," *Proc. of ACM/IEEE ICCAD*, 1998.

[13] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," *Proc. of ACM/IEEE ICCAD*, 2001.

[14] P. Wambacq, P. Dobrovolny, S. Donnay, M. Engels and I. Bolsens, "Compact modeling of nonlinear distortion in analog communication circuits," *Proc. of ACM/IEEE DATE*, 2000.