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 Nonlinear model Order Reduction Method (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 macro-modeling 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.

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 $n$ $h_n(t_1, \ldots, t_n)$ by

$$x_n(t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_n(t_1, \ldots, t_n) u(t-t_1) \ldots u(t-t_n) dt_1 \ldots dt_n. \tag{1}$$

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)

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

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

$$x_1 = A_1 x_1 + bu, \tag{3}$$

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

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

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 $x = \mathbf{V} \cdot x$, where $\mathbf{V}$ is an orthonormal basis of $\{V_1, V_2, V_3, \ldots, V_n\}$ and $V_k$ 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 $A_3 = \mathbf{V} A_3 \mathbf{V}^t \otimes \mathbf{V} \otimes \mathbf{V}$. However, the reduced third order matrix is usually dense and has $O(q^3)$ 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 model 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 assess 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

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

where $x$ and $y$ are the vectors of circuit unknowns and outputs, $u$ is the input, $f(\cdot)$ and $g(\cdot)$ 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 $g(\cdot)$ at the bias point and consider only small-signal quantities, we have

$$\frac{d}{dt}C_1 x + C_2 (x \otimes x) + C_3 (x \otimes x \otimes x) + \ldots, \tag{7}$$

$$+ G_1 x_1 + G_2 (x \otimes x) + G_3 (x \otimes x \otimes x) + \ldots = bu(t),$$

where $G_i = \frac{\partial f_i}{\partial x_i} \in \mathbb{R}^{n \times d}, C_i = \frac{\partial g_i}{\partial x_i} \in \mathbb{R}^{n \times n'}$ are the $i$th order conductance and capacitance matrices respectively. Before proceeding we first introduce the notions used throughout this paper. For matrices in (7) we define

$$A = -G_1^{-1} C_1, r_1 = G_2^{-1} b, r_2 = r_1 \otimes r_1, r_3 = r_1 \otimes r_1 \otimes r_1, \tag{8}$$

and for an arbitrary matrix $F$ define

$$F^{\otimes m} = F^\otimes F^\otimes \ldots, F^{\otimes 0} = \mathbb{1} \tag{9}$$

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

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

$$(G_1 + s C_1)H_1(x) = b, \text{ or } H_1(x) = (G_1 + s C_1)^{-1} b. \tag{10}$$

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

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

Similarly define $H_3(s_1) \otimes H_2(s_2) \otimes H_1(s_3)$ 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

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

$$H_1(s_1) \otimes H_1(s_2) \otimes H_1(s_3) = 2H_1(s_1) \otimes H_2(s_2, s_3) \tag{13}$$

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

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

where $M_{1,k}$ is the $k$th order moment of the first order transfer function. Expand $H_2(s_1, s_2)$ at the origin (0, 0)

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

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

$$H_2(s_1, s_2) = \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} M_{2,k,l} A^{m} G_1^{-1} \cdot (G_2 + s C_2) A^{n-1} r_2 \tag{16}$$
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

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

where \(M_{k, l, m}\) is a kth order moment. Due to the limit of space, we omit the expression of \(M_{k, l, m}\). 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 kth 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 kth order moments \(M_{1, l, m}, 0 \leq l \leq k, (M_{2, l, m}, 0 \leq l \leq k, 0 \leq m \leq l)\) or \(M_{3, l, m, n}, 0 \leq l \leq k, 0 \leq m \leq l, 0 \leq n \leq l - m\) ) 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 \(s_1^1, s_1^2, s_1^3, s_2^1, s_2^2, s_2^3, s_3^1, s_3^2\) 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. \(\tau = s_1 + s_2 + s_3\). As such, the final expression for \(M_{2, l, m}\) 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 4th 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 \(K_2(k) = \bigcup K_2(A, v_i)\), 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 \(K_3(k)\) for moment-matching of \(H_2\) up to 4th order can be also derived. Using these subspaces we can rigorously prove the following:

Theorem 1 If \(V = qr(K_{2,1}(A, r_1), K_2(k_2), K_3(k_3))\), where \(k_1 \geq k_2 \geq k_3\) is an orthonormal basis for the union of subspaces \(K_{2,1}(A, r_1), K_2(k_2)\) and \(K_3(k_3)\) then for (7) the reduced order model specified by the following system matrices

$$\tilde{G}_1 = \tilde{V}^T G_1 V, \quad \tilde{C}_1 = \tilde{V}^T C_1 V, \quad \tilde{b} = \tilde{V}^T b, \quad \tilde{d} = \tilde{V}^T d$$

$$\tilde{G}_2 = \tilde{V}^T G_2 (V \otimes V), \quad \tilde{C}_2 = \tilde{V}^T C_2 (V \otimes V)$$

$$\tilde{G}_3 = \tilde{V}^T G_3 (V \otimes V \otimes V), \quad \tilde{C}_3 = \tilde{V}^T C_3 (V \otimes V \otimes V)$$

is a \(k_1\) th order model in \(H_1\), a \(k_2\) th order model in \(H_2\), and a \(k_3\) th order model in \(H_3\).

In theorem 1, \(qr()\) refers to the QR procedure used for orthonormalizing the input vectors. We omit the proof of the theorem 1 and the derivation of subspaces \(K_3(k)\), but outline the complete single point version of NORM algorithm (expanded at the origin) in Fig. 2. For step 3.1,

$$(A^r_{1} r_1) \otimes M_{2, l, m, n, n} + (A^w_{1} r_1) \otimes M_{2, l, m, n, n}$$

$$+ (A^r_{1} r_1) \otimes M_{2, l, m, n, n} + (A^w_{1} r_1) \otimes M_{2, l, m, n, n}$$

$$+ (A^r_{1} r_1) \otimes M_{2, l, m, n, n} + (A^w_{1} r_1) \otimes M_{2, l, m, n, n}$$

$$+ (A^r_{1} r_1) \otimes M_{2, l, m, n, n} + (A^w_{1} r_1) \otimes M_{2, l, m, n, n}$$

where \(k, m, n, n\) is an orthonormal basis for the union of subspaces \(K_{2,1}(A, r_1), K_2(k_2)\) and \(K_3(k_3)\) then for (7) the reduced order model specified by the following system matrices

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 \(k_1 \geq k_2 \geq k_3\) 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 \(k_3\) th order moments of the original system respectively, then the overall reduced nonlinear model is a \(k_1\) th order model in \(H_1\), a \(k_2\) th order model in \(H_2\), and a \(k_3\) th order model in \(H_3\).

Table 1. Krylov subspaces containing the moments of \(H_2\)
for RF circuits, it might be desirable to expand both linear and non-
linear transfer functions at points other than the origin. It is crucial
to note that under nonlinear context, a single expansion point may
inherite multiple matrix factorizations due to the nonlinear frequency
mixing effects. Suppose that the inband third order intermodula-
tion of a nonlinear system around center frequency \( f_c \) is important
to model. To build the most compact model, one would opt to ex-
and \( H_3 \) 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 en-
sures the moment matching of the third order in-band intermodula-
tions. 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 addition 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 or-
der model grows faster than the order of moment matching, it is
very revealing to recognize that the numbers of up to \( k \)th order mo-
ments 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 expan-
sion version of NORM grows even faster than the number of mo-
mments matched. 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 as-
suming the dependency between transfer functions of different or-
ders is resolved properly, i.e., the reduced model size is the same
as the number of moments matched.

### Table 2. Comparison on the Reduced Model Sizes

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

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 non-
linear transfer functions at points other than the origin. It is crucial

### Table 2. Comparison on the Reduced Model Sizes

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

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 \( H_1 \), \( H_2 \), \( H_3 \). NORM-mp matches 4 moments of \( H_1 \) and \( H_2 \), 8 moments of \( H_3 \) generating a model size of 14. As can be seen, smaller models generated by NORM can actually match more 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 is ideally zero (except for numerical noise). To see the third order intermodulation translated by one LO frequency, the corresponding harmonic of the second-order nonlinearity transfer function

\[
\begin{align*}
\omega_3 & = -900\text{MHz} \\
\omega_1 & = f_1 \\
\omega_2 & = f_2
\end{align*}
\]

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 600.5MHz 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 730~840x were achieved, respectively.

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...
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