SPARSE VOLTERRA SYSTEMS: THEORY AND PRACTICE

Andrew Bolstad and Benjamin A. Miller

MIT Lincoln Laboratory
Lexington MA 02420

ABSTRACT

Nonlinear effects limit analog circuit performance, causing both in-band and out-of-band distortion. The classical Volterra series provides an accurate model of many nonlinear systems, but the number of parameters grows extremely quickly as the memory depth and polynomial order are increased. Recently, concepts from compressed sensing have been applied to nonlinear system modeling in order to address this issue. This work investigates the theory and practice of applying compressed sensing techniques to nonlinear system identification under the constraints of typical radio frequency (RF) laboratories. The main theoretical result shows that these techniques are capable of identifying sparse Memory Polynomials using only single-tone training signals rather than pseudorandom noise. Empirical results using laboratory measurements of an RF receiver show that sparse Generalized Memory Polynomials can also be recovered from two-tone signals.

Index Terms— Nonlinear system identification, nonlinear equalization, sparse modeling, compressed sensing, compressive sensing

1. INTRODUCTION

Nonlinear characteristics of analog circuits can degrade system performance by introducing both in-band and out-of-band distortion. This distortion can be supressed by nonlinear digital equalization provided that the nonlinearities are invertible and can be identified. The classical model for a nonlinear system is the Volterra series [1]:

$$y[n] = \sum_{p=0}^{P} \sum_{m_1,\dots,m_p=0}^{M-1} h_{VS}^{(p)}(m_1,\dots,m_p) \prod_{l=1}^{p} x[n-m_l].$$

Due to the rapid growth in the number of Volterra series coefficients with memory depth M and polynomial order P, alternative, "pruned" Volterra approaches have been proposed. Two well-known choices include the Memory Polynomial (MP) [2]:

$$y[n] = \sum_{p=0}^{P} \sum_{m=0}^{M-1} h_{MP}^{(p)}[m] x^p[n-m]$$

and Generalized Memory Polynomial (GMP) [3]:

$$y[n] = \sum_{p=0}^{P} \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} h_{GMP}^{(p)}[m_1, m_2] \times x[n-m_1] x^{p-1}[n-m_1-m_2]. \tag{1}$$

While these simplifications of the Votlerra series reduce the number of coefficients considerably, they can still contain large numbers of parameters. Moreover, estimating coefficeints $h^{(p)}$ is not always straightforward. The output is linear in the unknowns, but correlation between different polynomial combinations of the input requires special consideration (usually involving pseudorandom input signals) to achieve an invertible linear system [1, 4].

2. RELATION TO PRIOR WORK

The complexity of the standard Volterra series approach to nonlinear system modeling makes identification and compensation of these systems challenging. Thus, a sparse modeling approach is particularly attractive. Recently, ideas from the fields of sparse signal processing and compressed sensing have been applied to nonlinear system identification [5, 6, 7]. In these works, it is shown that certain sparse nonlinear systems can be identified using "short" random training sequences. The use of random training sequences permits, in some cases, the construction of matrices which satisfy a restricted isometry property (RIP), thus providing theoretical guarantees to compressed sensing approaches to nonlinear system identification.

While it is possible to generate excititory pseudorandom signals in an RF laboratory, it can be difficult to produce "clean" signals; i.e., signals whose amplitude can be precisely known, yet random at a given set of time points. On the other hand, signal generators, passive filters and passive combiners capable of producing signals consisting of several sinusoids

This work is sponsored by the Assistant Secretary of Defense for Research and Engineering under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.

with 90+ dB spur-free dynamic range (SFDR) are commonplace. The goal of this work is to show that sparse model identification can be applied to nonlinear systems using only sinusoidal input signals.

3. TESTBED LIMITATIONS

The theoretical and practical contributions of this work are motivated by the realities of RF test laboratories. In particular, the training signals used to stimulate a nonlinear system under test must be easily generated in a typical RF laboratory using common equipment. The following capabilities of the assumed testbed capture the key factors motivating the approach:

1. Clean tones can be generated; i.e. harmonics from the signal generators have been pushed below the noise floor via passive filters.
2. Tones can be generated anywhere in the band of interest, which does not include dc.
3. The analog-to-digital converter (ADC) does not introduce any additional nonlinearities into the signal.
4. The "resolution bandwidth" is sufficient to capture all harmonics of interest (these may be aliased).

The key factor differentiating this work from prior uses of compressive sensing techniques in nonlinear systems is that capability 1 is easily achieved in practice whereas generating clean pseudorandom signals is significantly more challenging.

4. RECOVERY OF SPARSE MEMORY POLYNOMIALS

This section presents the main theoretical result of the paper. Namely, that a nonlinear system described by a sparse Memory Polynomial can be identified by a small number of single-tone stimuli. Consider first the non-sparse case.

The output of the p^{th}-order kernel of a discrete time Memory Polynomial with memory depth M is given by [2]:

$$y_p[n] = \sum_{m=0}^{M-1} h_{MP}^{(p)}[m] x^p[n-m]. \tag{2}$$

Setting the input signal to $x[n] = \cos\left(\frac{2\pi k}{N}n + \theta\right)$ results in (letting $\omega = \frac{2\pi k}{N}$):

$$\begin{aligned} x^p[n-m] &= \frac{1}{2^p}\left(e^{j(\omega(n-m)+\theta)} + e^{-j(\omega(n-m)+\theta)}\right)^p \\ &= \frac{1}{2^p}\sum_{k=0}^{p}\binom{p}{k} e^{j(\omega(n-m)+\theta)(p-2k)} \\ &= \frac{1}{2^p}\left(e^{jp\theta}e^{-jp\omega m}e^{jp\omega n} + \ldots + e^{-jp\theta}e^{jp\omega m}e^{-jp\omega n}\right) \end{aligned}$$

From which the Discrete Time Fourier Series coefficient of Eq. 2 at $p\omega$ can be found:

$$Y_p[p\omega] = \sum_{m=0}^{M-1} 2^{-p} e^{jp\theta} e^{-jp\omega m} h_{MP}^{(p)}[m]. \tag{3}$$

Collecting measurements over a set of $M-1$ test tones at frequencies $\omega_i = \frac{2\pi}{N}k_i$ leads to:

$$2^p \begin{bmatrix} Y[p\omega_0] \\ Y[p\omega_1] \\ \vdots \\ Y[p\omega_{M-1}] \end{bmatrix} = \begin{bmatrix} e^{jp\theta_0} & 0 & \ldots & 0 \\ 0 & e^{jp\theta_1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & e^{jp\theta_{M-1}} \end{bmatrix} \times \tag{4}$$
$$\begin{bmatrix} 1 & e^{-jp\omega_0} & \ldots & e^{-jp(M-1)\omega_0} \\ 1 & e^{-jp\omega_1} & \ldots & e^{-jp(M-1)\omega_1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-jp\omega_{M-1}} & \ldots & e^{-jp(M-1)\omega_{M-1}} \end{bmatrix} \begin{bmatrix} h_{MP}^{(p)}[0] \\ h_{MP}^{(p)}[1] \\ \vdots \\ h_{MP}^{(p)}[M-1] \end{bmatrix}$$

The diagonal matrix in Eq. 4 is invertible since each entry is nonzero. The $M \times M$ Vandermonde matrix is invertible whenever $k_i \neq k_j$ modulo $\frac{N}{p}$. This results in the following:

Theorem 1: The coefficients of the p^{th}-order kernel of a Memory Polynomial of depth M are the solution to an invertible linear system defined by M sinusoidal training signals via Eq. 3.

This intermediate result suggests that recovery of sparse Memory Polynomial kernels is possible from a subset of the measurements needed to recover all M MP coefficients. This is in fact the case under certain condtions. For a given polynomial order p, select $N = lp$ for some integer $l \geq M$. Suppose only R frequencies are selected uniformly at random according to $\omega_r = \frac{2\pi}{N}k_r = \frac{2\pi}{N}(k_0 + i_r)$ with $i_r \in \{0, 1, \ldots, \frac{N}{p}\}$. In this case the full matrix of Eq. 4 can be written as:

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & e^{-jp\frac{2\pi}{N}k_1} & \ldots & e^{-jp(M-1)\frac{2\pi}{N}k_1} \\ 1 & e^{-jp\frac{2\pi}{N}k_2} & \ldots & e^{-jp(M-1)\frac{2\pi}{N}k_2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-jp\frac{2\pi}{N}k_R} & \ldots & e^{-jp(M-1)\frac{2\pi}{N}k_R} \end{bmatrix} \\ &= \begin{bmatrix} 1 & e^{-j\frac{2\pi}{N}pi_1} & \ldots & e^{-j\frac{2\pi}{N}p(M-1)i_1} \\ 1 & e^{-j\frac{2\pi}{N}pi_2} & \ldots & e^{-j\frac{2\pi}{N}p(M-1)i_2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\frac{2\pi}{N}pi_R} & \ldots & e^{-j\frac{2\pi}{N}p(M-1)i_R} \end{bmatrix} \times \\ &\quad \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & e^{-jp\frac{2\pi}{N}k_0} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & e^{-jp(M-1)\frac{2\pi}{N}k_0} \end{bmatrix} \end{aligned} \tag{5}$$

Denote this decomposition $\mathbf{A} = \mathbf{F}_{N/p}\mathbf{D}$. Since N is divisible by p and $\frac{N}{p} \geq M$, $\mathbf{F}_{N/p}$ is a sub-matrix of an $\frac{N}{p} = l$ point

DFT matrix consisting of R random rows and the the first M columns. It is well known in compressive sensing theory that such matrices are amenable to sparse reconstruction. For example, applying the main theorem of [8] gives:

Theorem 2: A p^{th} order Memory Polynomial kernel with K nonzero coefficients can be uniquely recovered with high probability from $\mathcal{O}(K\text{polylog}N)$ sinusoidal test signals by minimizing $\sum_m \left|h_{MP}^{(p)}[m]\right|$ subject to Eq. 3.

Remark: Alternatively, if N and p are relatively prime (with $N \geq p$), then the M columns of $\mathbf{F}_{N/p}$ are unique and it is again an appropriate sub-matrix of a DFT matrix.

5. SPARSE GENERALIZED MEMORY POLYNOMIALS

The Generalized Memory Polynomial (GMP) (Eq. 1) provides a more general and more broadly applicable nonlinear system model than MP [3]. Empirical results suggest that sparse GMP models can be identified using two-tone calibration signals and applied to improve the SFDR of RF receivers (see Sec. 6). Theoretical conditions establishing when this is possible (analogous to Thm. 2) are the subject of ongoing research. In the remainder of this section, a procedure for estimating sparse GMP systems is described.

Nonlinear effects in RF receivers can be compensated in the digital domain by (approximately) inverting the nonlinear system which describes the analog circuit behavior. Rather than approximating the system of interest with a Volterra series and subsequently finding its inverse, the (sparse) Volterra series representing the inverse can be found directly (see Fig. 1).

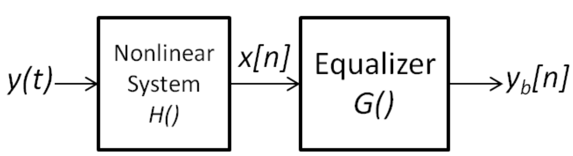


Fig. 1. Rather than identify nonlinear system H, equalizer G can be identified directly.

Consider nonlinear system H described by $x[n] = H(y(t))$. The goal of nonlinear equalizer G is to remove nonlinear distortions from the digitized signal $x[n]$ so that $G(x[n]) = y_b[n] = y_b(nT)$, where T is the sampling interval and y_b refers to signal y at baseband. The response of system H to a two-tone stimulus consists of the original two-tones (amplified, delayed, and converted to baseband) as well as intermodulation and harmonic distortions caused by nonlinear mixing of the two tones (see Fig. 3). Identifying a sparse GMP-based equalizer amounts to finding a small subset of terms from Eq. 1 that removes all distortion spurs while preserving the original two-tone signal of interest.

To this end, the nonlinear system H can be probed with multi-tone training signals. The desired output $y_b[n]$ can be written as a sum of the distorted output signal $x[n]$ and a linear combination of its products of the form $x[n-m_1]x^{(p-1)}[n-m_1-m_2]$:

$$\begin{aligned} y_b[n] &= x[n] + \sum_{p=2}^{P}\sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_{GMP}^{(p)}[m_1,m_2] \times \\ &\quad x[n-m_1]x^{p-1}[n-m_1-m_2] \qquad (6) \\ &= x[n] + \tilde{G}(x[n]). \end{aligned}$$

Since multi-tone training signals are used, the spectrum of $x[n]$ will consist of spikes at the frequencies of the input tones as well as spurious components at harmonic and intermodulation frequencies. The frequencies of spurious components can be easily determined from the tone frequencies and the nonlinear orders considered ($p = 2, 3, \ldots, P$). The frequency components of $\tilde{G}(x[n])$ can be found in a similar fashion, although spurious components not found in $x[n]$ are often negligably small. Considering only frequencies containing spurious activity allows Eq. 6 to be expressed in the frequency domain. The resulting equation at frequency ω_i is:

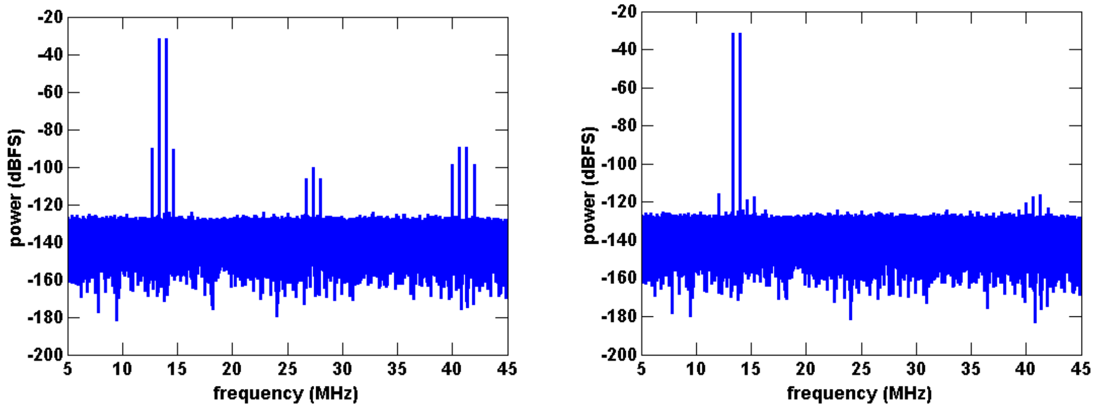
$$\begin{aligned} Y_b(\omega_i) = X(\omega_i) + \big[f_{\omega_i}(x^2[n]) \quad f_{\omega_i}(x[n]x[n-1]) \quad \ldots \\ f_{\omega_i}(x[n-M+1]x^{P-1}[n-2M+2])\big]\,\mathbf{h_{GMP}} \end{aligned}$$

where $f_{\omega_i}(u)$ is the DFT coefficient of u at ω_i.

Collecting measurements at all significant spurious frequencies over a number of multi-tone training signals leads to a linear inverse problem whose solution is a set of GMP coefficients describing the desired inverse nonlinear system G. Sec. 4 showed that a similar linear system (for the Memory Polynomial case) can admit the correct sparse solution or be invertible for an appropriate choice of training signals. The addition of a second delay m_2 in the GMP model complicates this analysis, but empirical results suggest that recovery of sparse GMP systems is possible.

6. MEASURED RESULTS

The sparse GMP approach described in Sec. 5 was applied to a low power homodyne RF receiver developed at MIT Lincoln Laboratory [9, 10]. Two-tone signals with baseband equivalent frequencies between 5 and 45 MHz were created at RF using Agilent E8257D signal generators, passive combiners, and a passive bandpass filter to eliminate nonlinear spurs generated by the signal generators. These signals were injected into the homodyne receiver whose output was digitized using an LTC2209 16-bit ADC from Linear Technology. The ADC was operated below full scale to avoid introducing additional nonlinearties into the data. The measured SFDR from the ADC alone at this input power was greater than 95 dBc.



(a) Two-tone signal before equalization.

(b) Two-tone signal after equalization.

Fig. 3. A sparse GMP equalizer using 30 coefficients improves the SFDR of a two-tone validation signal from 58 to 85 dBc.

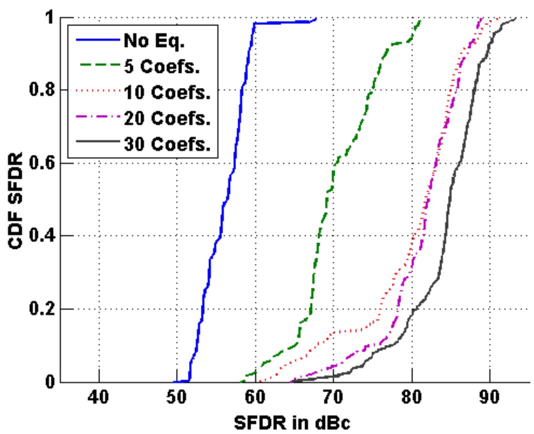


Fig. 2. Cumulative distribution function of validation signal SFDR. Median SFDR improves from 56 dBc with no equalization to 69 dBc using 5 GMP filter taps, 82 dBc using 10 or 20 taps, and 85 dBc using 30 filter taps.

The 752 two-tone signals collected in this manner were divided into training and validation groups of 526 and 226 test signals, respectively. As described in Sec. 5, the training signals were used to create a system of linear equations corresponding to a Volterra series with second-, third-, and fifth-order kernels. Delays m_1 and m_2 ranged from -3 to 4 (a negative delay corresponds to an advance). The resulting linear system $\mathbf{y} = \mathbf{X}\mathbf{h}_{\mathbf{GMP}}$ was approximately solved for sparse vector $\mathbf{h}_{\mathbf{GMP}}$ via the orthogonal matching pursuit (OMP) algorithm [11]. The columns of matrix $\mathbf{X}$ were normalized before applying OMP.

The result of applying the sparse GMP equalizer to the validation data is shown in Fig. 2. The circuit achieves about 56 dBc SFDR without equalization. With only five GMP coefficients, the SFDR increases by about 13 dB. Adding more GMP coefficients improves performance, but most improvement is aleardy achieved with 10 coefficients. With thirty coefficients, the SFDR improves by nearly 30 dB to 85 dBc.

Figure 2 illustrates some interesting characteristics of the sparse GMP equalizer. When 30 coefficients are used, about 77% of the validation data achieved 82 dBc or better, but the remaining 23% range between 65 and 81 dBc. One possible explanation for this wide range is that the sparse GMP model might not be sufficient near band edges. Increasing memory depth or selecting from a full Volterra series may improve this, but the sparse approximation problem becomes more difficult as the number of possible coefficients increases.

The result of applying the sparse GMP equalizer to a single two-tone validation signal is shown in Fig. 3. In this example SFDR improves from 58 to 85 dBc. Second- and third-order harmonics as well as third-order intermodulation distortions are significantly reduced by the equalizer.

7. SUMMARY

This work provides theoretical and practical results of applying compressed sensing techniques to nonlinear system identification using typical RF laboratory equipment. The main theoretical result proves that sparse Memory Polynomials can be recovered by applying a small number of single-tone signals to the nonlinear system. Expanding this theoretical result to sparse Generalized Memory Polynomials and sparse Volterra series is the subject of ongoing research. In the GMP case, laboratory tests suggest that two-tone signals are sufficient to estimate a nonlinear equalizer which boosts SFDR from 58 dBc to 85 dBc in an RF receiver.

8. ACKNOWLEDGEMENT

The authors would like to thank Michael Ericson, Karen Gettings, Merlin Green, Helen Kim, and all of the REDSoC team.

9. REFERENCES

[1] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, Wiley, 2000.

[2] L. Ding et al., “A robust digital baseband predistorter constructed using memory polynomials,” *IEEE Trans. Communications*, vol. 52, pp. 159–165, 2004.

[3] D.R. Morgan et al., “A generalized memory polynomial for digital predistortion of RF power amplifiers,” *IEEE Trans. on Signal Processing.*, vol. 54, pp. 3852–3860, 2006.

[4] R. D. Nowak and B. D. Van Veen, “Random and pseudorandom inputs for Volterra filter identification,” *IEEE Trans. on Signal Processing.*, vol. 42, pp. 2124–2135, 1994.

[5] B. Nazer and R. Nowak, “Sparse interactions: Identifying high-dimensional multilinear systems via compressed sensing,” in *Allerton Conference on Communication, Control and Computing*, September 2010.

[6] V. Kekatos and G. B. Giannakis, “Sparse Volterra and polynomial regression models: Recoverability and estimation,” *IEEE Trans. on Signal Processing.*, vol. 59, no. 12, pp. 5907–5920, December 2011.

[7] S. I. Adalbjörnsson, G. Glentis, and A. Jakobsson, “Efficient block and time-recursive estimation of sparse Volterra systems,” in *Proceedings of the IEEE Statistical Signal Processing Workshop*, Ann Arbor, MI, USA, August 2012, pp. 173–176.

[8] M. Rudelson and R. Vershynin, “On sparse reconstruction from Fourier and Gaussian measurements,” *Communications in Pure and Applied Mathematics*, vol. 61, pp. 1025–1045, 2008.

[9] H. H. Kim, M. Green, B. A. Miller, A. Bolstad, and D. D. Santiago, “An active filter achieving 43.6dBm OIP3,” in *Proceedings of the IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, Baltimore, MD, USA, June 2011.

[10] K. Gettings, A. Bolstad, S. S. Chen, M. Ericson, B. A. Miller, and M. Vai, “Low power sparse polynomial equalization (SPEq) for nonlinear digital compensation of an active anti-alias filter,” in *Proceedings of the IEEE Workshop on Signal Processing Systems (SiPS)*, Quebec City, Quebec, Canada, October 2012, pp. 250–253.

[11] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, “Orthogonal matching pursuit : Recursive function approximation with application to wavelet decomposition,” in *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, Monterey, CA, USA, 1993.