

Wavelet-Based Transformations for Nonlinear Signal Processing

Robert D. Nowak[†] and Richard G. Baraniuk[‡]

[†]Department of Electrical Engineering
Michigan State University, East Lansing, MI 48824 USA

[‡] Department of Electrical and Computer Engineering
Rice University, Houston, TX 77005 USA

Abstract

Nonlinearities are often encountered in the analysis and processing of real-world signals. This paper develops new transformations for nonlinear signal processing. The theory of tensor norms is employed to show that wavelets provide an optimal basis for the new transformations. The results are applied to Volterra kernel identification.

1. Introduction

Nonlinear signal coupling, mixing, and interaction play an important role in the analysis and processing of many signals. For instance, harmonic distortions and intermodulations are indicative of nonlinear behavior in amplifiers and faults in rotating machinery. Nonlinearities also arise in speech and audio processing, imaging, and communications. Furthermore, nonlinear signal processing techniques are commonly used in signal estimation and detection, image enhancement and restoration, and filtering.

This paper describes two new structures for nonlinear signal processing. Both structures are represented by the nonlinear signal transformation depicted in Figure 1. A length m signal vector $\mathbf{x}$ is first decomposed by an orthonormal signal basis $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$. The outputs of this first stage, $\{\beta_1, \dots, \beta_m\}$, are the coefficients of the signal with respect to this basis. Next, the signal coefficients are combined in nonlinear processing nodes, denoted by η , that are simple p -th order polynomial operations on the signal coefficients. The nonlinear signal transformation (NST) depicted in Figure 1 is denoted by F_p . Concisely, $F_p : \mathbf{x} \mapsto \boldsymbol{\theta}$, where the $\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]^T$ are called p -th order nonlinear coefficients of the signal $\mathbf{x}$.

The objective of the NST is to generate all possible p -th order nonlinear interactions between the various signal components. The strengths of the interactions are reflected in the nonlinear signal coefficients $\boldsymbol{\theta}$. The nonlinear

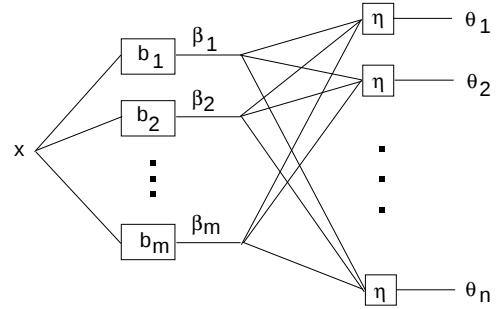


Figure 1. Nonlinear signal transformation F_p .

signal coefficients can be used for efficient nonlinear filter implementations, robust statistical estimation, and nonlinear signal analysis. Moreover, we will show that NSTs based on the *wavelet transform* are, in a certain sense, *optimal* for nonlinear processing. Applications of NSTs include Volterra filtering [3] and higher order statistical analysis.

Two types of nonlinear processing nodes distinguish the two structures considered in this paper. *Product nodes* form a nonlinear combination of the coefficients according to

$$\eta(\beta_1, \dots, \beta_m) = \beta_{i_1} \beta_{i_2} \dots \beta_{i_p}, \quad 1 \leq i_1 \leq \dots \leq i_p \leq m$$

Different p -fold products of the signal coefficients are computed at each product node. *Summing nodes* form a nonlinear combination of the coefficients according to

$$\eta(\beta_1, \dots, \beta_m) = \left(\sum_{i=1}^m a_i \beta_i \right)^p$$

where $\sum_i a_i \beta_i$ is a linear combination of the coefficients. Different linear combinations of the signal coefficients are raised to the p -th power at each summing node.

Although the outputs $\{\theta_k\}_{k=1}^n$ of the product and summing nodes are not equivalent, they both produce similar NSTs. In general, the outputs of the nonlinear processing nodes represent the p -fold interactions of various components of the original signal. For example, if $\mathbf{b}_1, \dots, \mathbf{b}_m$ are

This work was supported by the National Science Foundation, grant no. MIP-9457438, and the Office of Naval Research, grant no. N00014-95-1-0849.

Email: rnowak@egr.msu.edu, richb@rice.edu

Web: <http://www.egr.msu.edu/~rnowak>

<http://www-dsp.rice.edu>

chosen as Fourier basis elements, then the outputs represent p -fold frequency intermodulations. If $\mathbf{b}_1, \dots, \mathbf{b}_m$ are the canonical unit vectors and $\mathbf{x}$ is a vector of time-lagged samples, then the outputs represent p -fold products of different time lags of the signal. The number of nonlinear processing nodes n must be greater than the signal length m in order to completely represent all possible nonlinear interactions. In fact, it is shown in Section 2 that $\boldsymbol{\theta}$ represents all possible p -th order nonlinear interactions with $n = \binom{m+p-1}{p}$, where $\binom{m+p-1}{p}$ is the binomial coefficient.

Nonlinear signal analysis is traditionally carried out in the time or frequency domains. However, there are good reasons for considering alternative signal representations such as the wavelet domain [2]. From a practical perspective, studying nonlinear effects in the wavelet domain allows one to perform local nonlinear analysis and processing in both time and frequency. This can be advantageous in many non-stationary problems such as machinery monitoring [5] and image processing [6]. From a theoretical perspective, in Section 3 we show that the wavelet domain provides an optimal framework for studying nonlinear signals and systems. In Section 4, we apply the theoretical developments to a Volterra kernel identification problem.

2. Complete NSTs

2.1 Criterion for Completeness

In this section, we show that the transformation $F_p : \mathbf{x} \mapsto \boldsymbol{\theta}$ pictured in Figure 1 provides a complete representation of all p -th order nonlinear signal interactions. The notion of a complete transformation is defined as follows.

Definition 1. Let F_p be fixed. If for every signal $\mathbf{x} \in \mathbb{R}^m$ and every multidimensional array $h \in \underbrace{\mathbb{R}^m \times \dots \times \mathbb{R}^m}_{p\text{-times}}$

there exists a collection of real numbers $\{\alpha_k\}_{k=1}^n$ such that

$$\sum_{i_1, \dots, i_p=1}^m h_{i_1, \dots, i_p} x(i_1) \cdots x(i_p) = \sum_{k=1}^n \alpha_k \theta_k$$

where $\boldsymbol{\theta} = F_p(\mathbf{x})$, then the transformation F_p is said to be a complete p -th order signal transformation.

Definition 1 states that every p -th order multilinear functional of $\mathbf{x}$ may be computed by a linear functional of $\boldsymbol{\theta}$. Therefore, a complete p -th order nonlinear signal transformation allows us to study all possible p -th order nonlinear signal interactions of $\mathbf{x}$ in terms of simple linear operations on $\boldsymbol{\theta}$. This implies that a complete NST is capable of realizing every possible p -th order Volterra filter of $\mathbf{x}$. Furthermore, a complete NST captures all possible p -th order signal interactions necessary to compute higher order statistical quantities such as the moments and cumulants of $\mathbf{x}$.

We now show that both the product node and summing node transformations are complete. The NSTs can be interpreted as a linear mapping on an appropriate tensor space. Consequently, the notion of completeness can

be formulated as a spanning condition in a tensor space. Due to space limitations, we simply state the results. The details are given in [7]. The reader should keep in mind that the theory of tensor spaces plays an integral role in the analysis of NSTs, and we will discuss tensor spaces again in the next section.

2.2 Product Node Transformation

In the product node transformation, different p -fold products of the signal coefficients are computed at each node according to

$$\eta(\beta_1, \dots, \beta_m) = \beta_{i_1} \beta_{i_2} \cdots \beta_{i_p}$$

Completeness of this transformation is easily established by noting that this structure is related to a symmetric tensor space.

Theorem 2. Let $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ be a basis (orthonormal basis) for $\mathbb{R}^m$. Then the NST with $\binom{m+p-1}{p}$ product nodes forming all unique p -fold products of $\beta_1, \dots, \beta_m$ is complete.

2.3 Summing Node Transformation

Recall that the summing node nonlinearities raise linear combinations of the $\{\beta_1, \dots, \beta_m\}$ to the p -th power:

$$\eta(\beta_1, \dots, \beta_m) = \left(\sum_i a_i \beta_i \right)^p$$

We may view the $\{a_i\}$ as weights in the second stage of the structure pictured in Figure 1. A different linear combination feeds into each nonlinear processing node. Alternatively, we can interpret the summing node structure in the following way. Let $\{\alpha_{j_k}\}_{j=1}^m$ denote the weights to the k -th nonlinear node. Define

$$\mathbf{f}_k = \sum_{j=1}^m \alpha_{j_k} \mathbf{b}_j$$

Then the output of the k -th nonlinear node is given by

$$\theta_k = (\mathbf{f}_k^T \mathbf{x})^p$$

Hence, an alternative representation of the summing node structure is a bank of n linear filters $\{\mathbf{f}_k\}_{k=1}^n$ each followed by a simple monomial nonlinearity $(\cdot)^p$. The summing node transformation provides an extremely simple structure for generating complex nonlinear signal interactions. The key issue is how can the filters $\{\mathbf{f}_k\}_{k=1}^n$ be designed so that the resulting transformation is complete. A simple, constructive method to design a complete summing node structure is given next.

Let $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ be a basis for $\mathbb{R}^m$ and let $\{a_0, a_1, \dots, a_p\} \subset \mathbb{R}$. Next, let $n = \binom{m+p-1}{p}$ and define the set of vectors

$$\{\mathbf{a}_k\}_{k=1}^n = \left\{ [a_{k_1}, \dots, a_{k_m}]^T : \sum_{j=1}^m k_j = p, k_j \in \{0, \dots, p\} \right\} \quad (1)$$

Let $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_m]^T$ and define

$$\mathbf{f}_k = \mathbf{B} \mathbf{a}_k \quad (2)$$

$\mathbf{f}_k$ is a linear combination of the basis vectors $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ and $\mathbf{a}_k$ is a vector containing the coefficients of the combination. This construction of $\{\mathbf{f}_k\}$ leads to a complete p -th order transformation.

Theorem 3. Let $\rho \in \mathbb{R}$, $|\rho| \neq 1$, $\rho \neq 0$. Take $a_0 = 0$ and $a_j = \rho^j - 1$, $j = 1, \dots, p$, and form $\{\mathbf{f}_k\}_{k=1}^n$ according to equations (1) and (2). Then the summing node NST is complete.

The construction above basically generates a class of linear filters sufficiently rich so that their tensor products span all possible p -th order interactions of the basis vectors.

The summing node transformation is particularly interesting because it allows us to represent every p -th order Volterra filter as simple filter bank, as shown in Figure 2.

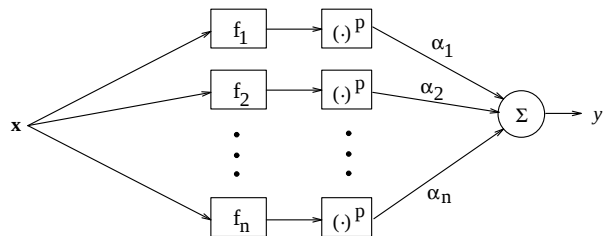


Figure 2. Volterra filter realization using summing node transformation. This realization conveniently separates the dynamics and nonlinearity of the filter.

Volterra filter realizations of this type are often referred to as *parallel-cascade realizations*. Previously studied parallel-cascade constructions rely on complicated numerical optimization techniques [3] and construct a set of *kernel-specific* linear filters. Hence, a separate parallel-cascade structure is required to represent each distinct Volterra filter. In contrast, the parallel-cascade realization based on the summing node transformation is capable of representing every p -th order Volterra filter by simply adjusting the output weights $\{\alpha_k\}_{k=1}^n$. The linear filters $\{\mathbf{f}_k\}_{k=1}^n$ of the summing node transformation are the same for every Volterra kernel. Hence, the summing node transformation is a *universal structure* capable of realizing every homogeneous Volterra filter. Nonhomogeneous Volterra filters can also be implemented with the summing nodes by following each linear filter with a degree p polynomial nonlinearity instead of the homogeneous p -th order monomial.

3. NSTs in the Wavelet Domain

The previous section has shown that complete nonlinear signal transformations can be derived from any orthonormal signal basis $\mathbf{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$. For example, $\mathbf{B}$ may be a time, Fourier, or wavelet domain basis. We

will now show that wavelet-based NSTs offer a significant theoretical advantage. The motivation for wavelet-based NSTs is developed for infinite dimensional (continuous) spaces (note that until now we have focused on finite dimensional signal spaces). The properties of wavelet-based NSTs in the infinite dimensional setting carry over to high-dimensional sampled spaces.

It has been shown that noise removal, compression, and signal recovery methods based on wavelet coefficient shrinkage or wavelet series truncation enjoy asymptotic minimax performance characteristics and do not introduce excessive artifacts in the signal reconstruction [1]. The theoretical justification for the exceptional performance of wavelet-based processing is the fact that wavelet bases are *unconditional bases* for many signal spaces.

It is well-known that wavelet bases derived from multiresolution transformations are unconditional bases for a diverse variety of signal spaces. However, for the NSTs of interest, tensor spaces are the natural framework to consider. Hence, we would like to establish the unconditionality of tensor product wavelet bases. It should be noted that the tensor wavelet basis, also referred to as the “rectangular wavelet decomposition” [2], is quite different from the usual multidimensional wavelet basis obtained via a multiresolution analysis.

The theorem below, proved in [7], shows that the tensor product of a wavelet basis is an unconditional basis for the tensor space $L_p(\mathbb{R}) \otimes_{\Delta_p} L_p(\mathbb{R})$, which is isometric to the space of 2D L_p functions.

Theorem 4. If $\{\phi_i\}$ is an unconditional wavelet basis for $L_p(\mathbb{R})$, $1 < p < \infty$, then $\{\phi_i \otimes \phi_j\}$ is an unconditional basis for $L_p(\mathbb{R}) \otimes_{\Delta_p} L_p(\mathbb{R})$.

This result can easily be extended to arbitrary p -th order tensor spaces, and shows that wavelet-based NSTs correspond to an unconditional basis expansion of the nonlinear signal coefficients. It should be possible to extend this result to more general spaces, including various smoothness spaces. One possible starting point for the general problem may be found in [4].

4. Application — Volterra Kernel Estimation and Approximation

To illustrate the estimation power of wavelet-based NSTs for Volterra filtering, consider the following Volterra kernel estimation problem. We observe the input and output of a quadratic interaction, given by

$$y(k) = \sum_{i,j=1}^m h(i,j) x(k-i)x(k-j)$$

We assume that the input signal $\{x(k)\}$ is i.i.d. zero-mean Gaussian. Hence, we have the model $Y = \sum_{i,j=1}^m h(i,j) X_i X_j$, where $\{X_i\}_{i=1}^m$ are i.i.d. zero-mean Gaussian with variance σ^2 . The kernel h is easily estimated, via correlation analysis, from independent observations of $\{X_i\}$ and Y . Let $s(i,j)$ denote the sample average estimate of $E[Y X_i X_j]$. This leads to the following

estimator of h :

$$\hat{h}(i, j) = \begin{cases} \frac{s(i, j) - (m+2)^{-1} \sum_k s(k, k)}{2 \sigma^4}, & i = j \\ \frac{s(i, j)}{2 \sigma^4}, & i \neq j \end{cases} \quad (3)$$

This simple correlation estimator converges to the true kernel as the number of observations increases. However, for a finite number of data the resulting estimate is typically very noisy. To improve our estimate, we suggest expanding the estimate in a tensor basis and applying a threshold to the coefficients of the expansion. The hard-threshold, at level τ , is given by

$$\tilde{\theta}_i = \begin{cases} \hat{\theta}_i, & \hat{\theta}_i \geq \tau, \\ 0, & \hat{\theta}_i < \tau. \end{cases} \quad (4)$$

We compare the performance of the wavelet and Fourier tensor bases in a simulated estimation problem. The kernel h is depicted in Figure 3 (a). This kernel is the actual quadratic Volterra kernel measured for an audio loudspeaker.¹ The raw estimate $\hat{h}$ obtained from the simulated estimation procedure above is pictured in Figure 3 (b). The noise in this raw estimate is reduced by thresholding the Fourier and wavelet kernel expansions at the level $\tau = [2 \log(n)]^{1/2} \sigma$, where σ standard deviation of the noise and $n = m^2 = 1024$, the dimension of the discretized kernel. This is the choice suggested in [1]. Figure 3 (c) and (d) show the resulting estimates using the Daubechies length 8 wavelet basis and Fourier basis, respectively.

5. Conclusions

We have developed two structures for computing p -th order nonlinear signal transformations. The transformations have an elegant interpretation in tensor spaces. It is shown in [7] that the product node transformation produces an orthogonal decomposition in the tensor space and is therefore very appropriate for estimation problems. The summing node transformation can be computed using a simple filter bank structure particularly well-suited to fast implementations. We have shown that the wavelet basis provides an optimal framework for NSTs. We have focused on the classical L_p tensor spaces, but extensions to more general settings may be possible using the results of [4]. Application of the nonlinear signal transformation to Volterra kernel identification demonstrate the utility our developments.

¹The authors wish to thank Dr. Walter Frank of Universität der Bundeswehr München for supplying the loudspeaker kernel measurements and Dr. Douglas Jones of the University of Illinois for discussing the use of the JAM in this context.

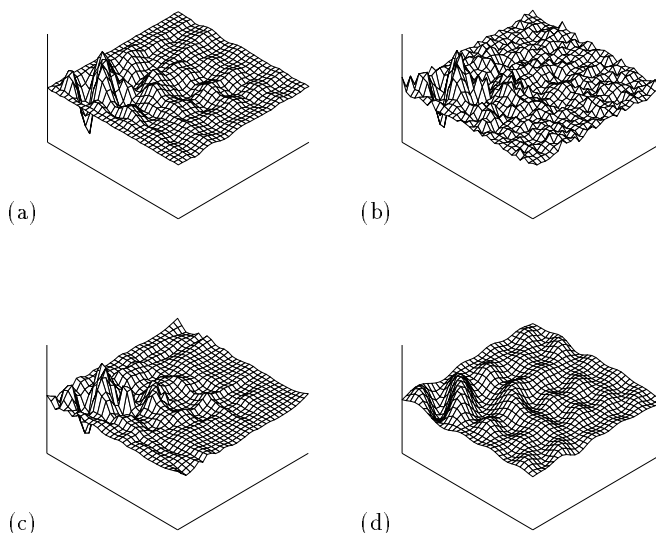


Figure 3. Estimates of quadratic kernel. (a) True kernel. (b) Raw estimate, MSE=0.20. (c) Thresholded wavelet expansion, MSE=0.15. (d) Thresholded Fourier expansion, MSE=0.40.

References

- [1] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *Applied and Computational Harmonic Analysis*, pp. 100-115, vol. 1, no. 1, December, 1993.
- [2] B. Jawerth and W. Sweldens, "An Overview of Wavelet Based Multiresolution Analysis," *SIAM Review*, vol.36, no. 3, pp. 377-412, 1994.
- [3] T. M. Panicker and V.J. Mathews, "Parallel-cascade realizations and approximations of truncated Volterra systems," *Proceedings of the 1996 IEEE Intl. Conf. on Acoustics, Speech, and Signal Proc.*, pp. 1589-1592, Atlanta, 1996.
- [4] W. Wang, "A discrete transform and Triebel-Lizorkin spaces on the bidisc," *Trans. AMS*, vol. 347, no. 4, April, 1996.
- [5] K. C. Chou and L. P. Heck, "A multiscale stochastic modelling approach to the monitoring of mechanical systems," *IEEE-SP Intl. Symp. on Time-frequency and Time-scale Analysis*, pp. 25-28, 1994.
- [6] R. D. Nowak and R. G. Baraniuk, "Optimal weighted highpass filters using multiscale analysis," submitted to *IEEE Trans. on Image Processing*, June, 1996.
- [7] R. D. Nowak and R. G. Baraniuk, "Wavelet-based transformations for nonlinear signal processing," submitted to *IEEE Trans. on Signal Processing*, January, 1997.