Wavelet-Based Transformations for Nonlinear Signal Processing

Robert D. Nowak

Department of Electrical Engineering
Michigan State University
East Lansing, MI 48824–1226

Fax: (517) 353–1980
Email: rnowak@egr.msu.edu
Web: www.egr.msu.edu/spc/

Richard G. Baraniuk *

Department of Electrical and Computer Engineering
Rice University
Houston, TX 77005–1892

Fax: (713) 524–5237
Email: richb@rice.edu
Web: www.dsp.rice.edu

Submitted to *IEEE Transactions on Signal Processing*, February 1997
Revised June 1998
EDICS Numbers: SP-2.4.4 Wavelets and filter banks, SP-2.7.2 Nonlinear analysis

Abstract

Nonlinearities are often encountered in the analysis and processing of real-world signals. In this paper, we introduce two new structures for nonlinear signal processing. The new structures simplify the analysis, design, and implementation of nonlinear filters and can be applied to obtain more reliable estimates of higher-order statistics. Both structures are based on a two-step decomposition consisting of a linear orthogonal signal expansion followed by scalar polynomial transformations of the resulting signal coefficients. Most existing approaches to nonlinear signal processing characterize the nonlinearity in the time domain or frequency domain; in our framework any orthogonal signal expansion can be employed. In fact, there are good reasons for characterizing nonlinearity using more general signal representations like the wavelet expansion. Wavelet expansions often provide very concise signal representation and thereby can simplify subsequent nonlinear analysis and processing. Wavelets also enable local nonlinear analysis and processing in both time and frequency, which can be advantageous in non-stationary problems. Moreover, we show that the wavelet domain offers significant theoretical advantages over classical time or frequency domain approaches to nonlinear signal analysis and processing.

1 Introduction

Nonlinear signal coupling, mixing, and interaction play an important rôle in the analysis and processing of signals and images. For instance, harmonic distortions and intermodulations indicate nonlinear behavior in amplifiers and faulty behavior in rotating machinery. Nonlinearities also arise in speech and audio processing, imaging, and communications. Nonlinear signal processing techniques are commonly applied in signal detection and estimation, image enhancement and restoration, and filtering.

In this paper, we develop a new approach to nonlinear signal processing based on the *nonlinear signal transformation* (NST) depicted in Figure 1. Here, a length-m signal vector $\mathbf{x}$ is first expanded in an orthonormal signal basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ to produce the vector of coefficients $[\beta_1, \ldots, \beta_m]^T$. These signal coefficients are then combined in nonlinear processing nodes η, which are simple n-th order polynomial operations, to form the *n-th order nonlinear coefficients* of the signal $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^T$. Concisely, we denote the NST of Figure 1 by the operator $F_n : \mathbf{x} \mapsto \boldsymbol{\theta}$.

*This work was supported by the National Science Foundation, grant nos. MIP–9701692 and MIP–9457438, the Office of Naval Research, grant no. N00014–95–1–0849, and DARPA/AFOSR, grant no. F49620-97-1-0513.

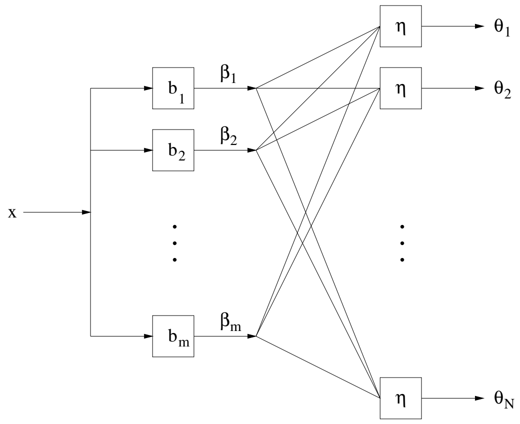


Figure 1: *Nonlinear signal transformation (NST) $F_n : \mathbf{x} \mapsto \boldsymbol{\theta}$. The front end processing (expansion in terms of the basis $\{\mathbf{b}_j\}$) is linear; the back end processing (by η from* (1) *or* (2)*) is nonlinear.*

The NST framework encompasses two new structures, each corresponding to a different choice for the scalar processing nodes η in Figure 1. *Product nodes* compute different n-fold products of the signal coefficients at each node:

$$\eta\left(\beta_1, \ldots, \beta_m\right) \;=\; \beta_{i_1}\beta_{i_2}\cdots\beta_{i_n}. \tag{1}$$

Summing nodes raise linear combinations of the coefficients to the n-th power:

$$\eta\left(\beta_1, \ldots, \beta_m\right) \;=\; \left(\sum_{j=1}^{m} a_j\, \beta_j\right)^n. \tag{2}$$

(Although the outputs of the product and summing nodes are not equivalent, we will see that they both produce similar NSTs.)

We will prove that an NST architecture with $\binom{m+n-1}{n}$ processing nodes can generate *all possible* n-th order nonlinear interactions between the various signal components, with the strengths of these interactions reflected in the nonlinear signal coefficients $\boldsymbol{\theta}$. Therefore, these coefficients can be used for efficient nonlinear filter implementations, robust statistical estimation, and nonlinear signal analysis.

The NST framework is flexible, because it does not rely on a particular choice of basis $\{\mathbf{b}_j\}$. Traditionally, nonlinear signal analysis has been carried out in the time or frequency domains. For example, if the $\{\mathbf{b}_j\}$ are the canonical unit vectors, or delta basis, then the components of $\boldsymbol{\theta}$ represent n-th order interactions between different time lags of the signal $\mathbf{x}$ (see Figure 3(a)). If the $\{\mathbf{b}_j\}$ make up the Fourier basis, then $\boldsymbol{\theta}$ represents the n-th order frequency intermodulations (see Figure 3(b)). In this paper, we will emphasize the *wavelet basis* [7], whose elements are localized in both time and frequency. Wavelet-based NSTs represent the local n-th order interactions between signal components at different times *and* frequencies (see Figure 3(c)). ¿From a practical perspective, this can be advantageous in problems involving non-stationary data, such as machinery monitoring [6] and image processing [20]. From a theoretical perspective, we will show that the wavelet domain provides an optimal framework for studying nonlinear signals and systems.

We will consider several applications of NSTs in this paper. NSTs provide an elegant structure for the *Volterra filter* that simplifies filter analysis, design, and implementation. Applications of Volterra filters include signal detection and estimation, adaptive filtering, and system identification [15, 25]. The output of a Volterra filter applied to a signal $\mathbf{x}$ consists of a polynomial combination of the samples of $\mathbf{x}$. We will show that every n-th order Volterra filter can be represented by simple linear combinations of the nonlinear signal coefficients $\boldsymbol{\theta}$. NSTs are also naturally suited for performing *higher-order statistical signal analysis* [17]. For example, in the time or frequency domains, the expected values of the nonlinear signal coefficients $\boldsymbol{\theta}$ are simply values of a higher-order moment or higher-order spectrum. We will argue that the wavelet domain provides an alternative, and optimal, representation for higher-order statistical analysis.

The paper is organized as follows. First, we introduce the NST framework. Second, we investigate the advantages offered by a wavelet basis formulation instead of classical time or frequency domain formulations. Specifically, in Section 2, we provide a brief introduction to the theory of tensor spaces, which are central to the NST and its analysis. In Section 3, we show that both the product and summing node NSTs provide a complete representation of all possible n-th order nonlinear signal interactions. Then using the theory of tensor norms and Gordon-Lewis spaces, we examine the issue of choosing a signal basis for NSTs. In particular, we exploit the special properties of the wavelet basis to show in Section 4 that wavelet bases are, in a certain sense, optimal for nonlinear signal analysis and processing. Section 5 applies the theory to three nonlinear signal processing applications. Section 6 offers a discussion and conclusions.

2 Tensor Spaces

In this Section, we provide a brief introduction to the theory of tensor spaces, which provide an elegant and powerful framework for analyzing NSTs. The theory of tensor spaces will be used to establish the completeness of NSTs and to assess the merits of different basis transformations.

2.1 Finite-dimensional tensor spaces

First, some notation for $\mathbb{R}^m$ (we will deal exclusively with real-valued signals in this paper). All vectors will be assumed to be columns and will be denoted using bold lowercase letters; for example, $\mathbf{v} = [v_1, \ldots, v_m]^T$. Bold uppercase letters will denote matrices. Define the inner product $\langle \mathbf{u}, \mathbf{v} \rangle \triangleq \mathbf{u}^T \mathbf{v}$.

Given a collection of m-dimensional, real-valued vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, with $\mathbf{v}_k = [v_{1,k}, \ldots, v_{m,k}]^T$, the n-fold *tensor* or *Kronecker product* [4, 28] $\boldsymbol{\tau} = \bigotimes_{j=1}^{n} \mathbf{v}_j$ produces a vector composed of all possible n-fold cross-products of the elements in $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. We can also interpret the tensor $\boldsymbol{\tau}$ as an amorphous n-dimensional array with elements $\tau_{i_1,\ldots,i_n} = v_{i_1,1} \cdots v_{i_n,n}$. The n-fold tensor product of a vector $\mathbf{v}$ with itself is denoted by $\mathbf{v}^{(n)}$ and contains all n-fold cross-products of the elements in $\mathbf{v}$.

The span of all n-th order tensors generates the n-th order *tensor space* $T^n(\mathbb{R}^m)$ [28]. For example, if $n = 2$, then

$$T^2(\mathbb{R}^m) \triangleq \left\{ \sum_{j=1}^{L} \mathbf{u}_j \otimes \mathbf{v}_j : \mathbf{u}_j, \mathbf{v}_j \in \mathbb{R}^m, L \geq 1 \right\}. \tag{3}$$

Practically speaking, $T^n(\mathbb{R}^m)$ is simply the space $\mathbb{R}^{m^n}$.

A tensor $\boldsymbol{\tau} \in T^n(\mathbb{R}^m)$ is *symmetric* [28] if for every set of indices $\{i_1, \ldots, i_n\}$ and for every permutation $\{\omega(1), \ldots, \omega(n)\}$ from the set Ω of permutations of $\{1, \ldots, n\}$ we have

$$\tau_{i_1,\ldots,i_n} \;=\; \tau_{i_{\omega(1)},\ldots,i_{\omega(n)}}. \tag{4}$$

Any tensor $\boldsymbol{\tau} \in T^n(\mathbb{R}^m)$ can be symmetrized by averaging over all possible permutations of the indices, forming

$$\mathcal{S}(\boldsymbol{\tau}) \;\triangleq\; \frac{1}{n!} \sum_{\omega \in \Omega} \tau_{i_{\omega(1)},\ldots,i_{\omega(n)}}. \tag{5}$$

The subspace of $T^n(\mathbb{R}^m)$ containing all n-th order tensors satisfying (4) is termed the n-th order *symmetric tensor space* $S^n(\mathbb{R}^m)$. The dimension of $S^n(\mathbb{R}^m)$ is $\binom{m+n-1}{n}$, the number of n-selections from an m element set. Throughout the sequel, we will set $N = \binom{m+n-1}{n}$.

2.2 Example

To illustrate the above ideas, consider the tensor space $T^2(\mathbb{R}^2)$ and the symmetric tensor space $S^2(\mathbb{R}^2)$. For example, let $\mathbf{u} = [u_1,\, u_2]^T, \mathbf{v} = [v_1,\, v_2]^T \in \mathbb{R}^2$. Then $\boldsymbol{\tau} \triangleq \mathbf{u} \otimes \mathbf{v} = [u_1v_1,\, u_1v_2,\, u_2v_1,\, u_2v_2]^T \in T^2(\mathbb{R}^2)$. We can also interpret $\boldsymbol{\tau}$ as a 2-d array:

$$\boldsymbol{\tau} \;=\; \mathbf{u}\mathbf{v}^T \;=\; \begin{bmatrix} u_1v_1 & u_1v_2 \\ u_2v_1 & u_2v_2 \end{bmatrix}. \tag{6}$$

The symmetrized tensor $\mathcal{S}(\tau) \in S^2(\mathbb{R}^2)$ is given by

$$\mathcal{S}(\boldsymbol{\tau}) \;=\; (\mathbf{u}\mathbf{v}^T + \mathbf{v}\mathbf{u}^T)/2 \;=\; \begin{bmatrix} u_1v_1 & (u_1v_2 + u_2v_1)/2 \\ (u_1v_2 + u_2v_1)/2 & u_2v_2 \end{bmatrix}. \tag{7}$$

2.3 Continuous-time tensor spaces

In practice, we work with the finite-dimensional tensor spaces associated with finite duration, discrete-time signals. However, in order to assess the merits of various signal bases (Fourier versus wavelet, for example) it is useful to consider the situation in continuous-time (infinite-dimensional) signal spaces. We will see that here the wavelet basis offers a significant advantage over the Fourier basis. Hence, we may infer that these advantages carry over into high sample rate discrete-time signal spaces.

We now consider the construction of continuous-time tensor spaces. Let $\mathcal{X}$ be a signal space. The n-th order tensor space $T^n(\mathcal{X})$ is the space generated by the span of all n-fold tensor products of signals in $\mathcal{X}$ [8]. For example, if $n = 2$, then

$$T^2(\mathcal{X}) \;\triangleq\; \left\{ \sum_{j=1}^{L} x_j \otimes y_j \,:\; x_j, y_j \in \mathcal{X}, L \geq 1 \right\}. \tag{8}$$

If $x, y \in \mathcal{X}$ are one-dimensional functions of a parameter t, then $x \otimes y$ is canonically identified with the two-dimensional function $z(t_1, t_2) = x(t_1)\, y(t_2)$.

To rigorously study continuous-time tensor spaces, we must equip $T^n(\mathcal{X})$ with a *tensor norm* [8]. First, we assume that the space $\mathcal{X}$ is itself equipped with a norm — for example, $\mathcal{X} = L_p(\mathbb{R})$. The norm on

$\mathcal{X}$ can induce a norm on $T^n(\mathcal{X})$ in a number of ways. Focusing on L_p spaces, consider the *natural tensor norm* Δ_p, which is generated by the standard one-dimensional L_p norm. We equip the algebraic tensor space $L_p(\mathbb{R}) \otimes L_p(\mathbb{R})$ with Δ_p and let $L_p(\mathbb{R}) \otimes_{\Delta_p} L_p(\mathbb{R})$ denote the completion of this space. Roughly speaking, Δ_p is a tensor norm that acts like the standard two-dimensional L_p norm. In fact, the normed tensor space $L_p(\mathbb{R}) \otimes_{\Delta_p} L_p(\mathbb{R})$ is isometric to the space of p-integrable two-dimensional functions $L_p(\mathbb{R} \times \mathbb{R})$. We will rejoin continuous-time tensor spaces in Section 4, where we study the performance of tensor wavelet bases from an approximation-theoretic perspective.

3 Complete NSTs

In this section, we show that the transformation $F_n : \mathbf{x} \mapsto \boldsymbol{\theta}$, pictured in Figure 1, provides a complete representation of all possible n-th order nonlinear signal interactions. More precisely, every n-th order multilinear functional of the samples of the signal $\mathbf{x}$ is expressible as a linear functional of the nonlinear signal coefficients $\boldsymbol{\theta}$. Practical implications of completeness are that an n-th order NST is capable of realizing every possible n-th order Volterra filter of $\mathbf{x}$ and can capture all possible n-th order signal interactions necessary to compute higher-order statistical quantities such as the moments and cumulants of $\mathbf{x}$. We focus our attention primarily on sampled, finite duration signals. Using the theory of finite-dimensional tensor spaces, we equate the completeness of the NSTs to a spanning condition in a tensor space.

3.1 Criterion for completeness

Definition 1 *Let $F_n : \mathbf{x} \mapsto \boldsymbol{\theta}$ be fixed. If for every signal $\mathbf{x} \in \mathbb{R}^m$ and tensor $h \in T^n(\mathbb{R}^m)$ there exists a collection of real numbers $\{\alpha_k\}_{k=1}^N$, $N = \binom{m+n-1}{n}$, such that*

$$\sum_{1 \le i_1 \le \ldots \le i_n \le m} h_{i_1,\ldots,i_n}\, x_{i_1} \cdots x_{i_n} \;=\; \sum_{k=1}^{n} \alpha_k\, \theta_k, \tag{9}$$

then the transformation F_n is a **complete n-th order NST**.

In words, a complete NST can represent every n-th order multilinear functional of the signal samples as a linear functional of the nonlinear signal coefficients $\boldsymbol{\theta}$.

Using the theory of tensor spaces, the completeness property is easily described. Note that the tensor $\mathbf{x}^{(n)}$ contains every product of the form

$$x_{i_1} \cdots x_{i_n}, \qquad 1 \le i_1, \ldots, i_n \le m. \tag{10}$$

In tensorial notation, we can rewrite the multilinear function on the left side of (9) as the inner product

$$\sum_{1 \le i_1 \le \ldots \le i_n \le m} h_{i_1,\ldots,i_n}\, x_{i_1} \cdots x_{i_n} \tag{11}$$

Furthermore, since $\mathbf{x}^{(n)}$ is a symmetric tensor, we can assume without loss of generality that $\mathbf{h} \in S^n(\mathbb{R}^m)$. We now show that both the product node and summing node NSTs are complete.

3.2 Product node transformation

The product node NST is computed as follows. The coefficients $\beta_1, \ldots, \beta_m$ of the orthogonal expansion are simply the inner products of the basis vectors $\{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ with the signal vector $\mathbf{x}$; that is, $\beta_j = \langle \mathbf{b}_j, \mathbf{x} \rangle$. The coefficients $\boldsymbol{\theta}$ output at the second, nonlinear stage are given by all n-fold products of the $\{\beta_j\}_{j=1}^m$ (see (1)). The output of the product node NST F_n is thus

$$\{\theta_k\}_{k=1}^N \;=\; \{\beta_{i_1} \cdots \beta_{i_n} \;:\; 1 \le i_1 \le \ldots \le i_n \le m\}\,. \tag{12}$$

Tensor products simplify the description of the product node NST. First note that products of the form $\beta_{i_1} \cdots \beta_{i_n}$ in (12) can be expressed, using standard tensor product identities [4], as

$$\beta_{i_1} \cdots \beta_{i_n} \;=\; \langle \mathbf{b}_{i_1}, \mathbf{x} \rangle \cdots \langle \mathbf{b}_{i_n}, \mathbf{x} \rangle \;=\; \left\langle \bigotimes_{j=1}^n \mathbf{b}_{i_j}, \mathbf{x}^{(n)} \right\rangle. \tag{13}$$

Next, since the ordering of the $i_1, \ldots, i_n$ does not affect the product value, we can symmetrize (13)

$$\beta_{i_1} \cdots \beta_{i_n} \;=\; \left\langle \mathcal{S}\left(\bigotimes_{j=1}^n \mathbf{b}_{i_j} \right),\; \mathbf{x}^{(n)} \right\rangle. \tag{14}$$

Now consider the collection of symmetric tensors

$$\left\{ \mathcal{S}\left(\bigotimes_{j=1}^n \mathbf{b}_{i_j} \right) \;:\; 1 \le i_1 \le \ldots \le i_n \le m \right\}. \tag{15}$$

Applying each of these tensors to the signal tensor $\mathbf{x}^{(n)}$ produces the $\{\theta_k\}_{k=1}^N$ defined in (12). Hence, the linear combination $\sum_{k=1}^N \alpha_k\, \theta_k$ of Definition 1 is given by

$$\sum_{1 \le i_1 \le \ldots \le i_n \le m} \alpha_{i_1,\ldots,i_n} \; \left\langle \mathcal{S}\left(\bigotimes_{j=1}^n \mathbf{b}_{i_j} \right),\; \mathbf{x}^{(n)} \right\rangle, \tag{16}$$

where we have used a multi-indexing scheme on the $\{\alpha_k\}$ for notational convenience. Comparing this expression to (9) and (11), we make the identification

$$\mathbf{h} \;=\; \sum_{1 \le i_1 \le \ldots \le i_n \le m} \alpha_{i_1,\ldots,i_n}\; \mathcal{S}\left(\bigotimes_{j=1}^n \mathbf{b}_{i_j} \right). \tag{17}$$

It follows from (17) and Definition 1 that the product node NST is complete if the following condition is satisfied:

$$\mathrm{Span}\left\{ \mathcal{S}\left(\bigotimes_{j=1}^n \mathbf{b}_{i_j} \right) \;:\; 1 \le i_1 \le \ldots \le i_n \le m \right\} \;=\; S^n(\mathbb{R}^m). \tag{18}$$

This is in fact the case.

Theorem 1 [28] *Let $\{\mathbf{b}_j\}_{j=1}^m$ be a basis (orthonormal basis) for $\mathbb{R}^m$. Then the $N = \binom{m+n-1}{n}$ symmetric tensors* (15) *form a basis (orthonormal basis) for $S^n(\mathbb{R}^m)$.*

Thus, the product node structure affords a complete NST, provided $\{\mathbf{b}_j\}_{j=1}^m$ is a basis for $\mathbb{R}^m$.

3.3 Summing node transformation

Recall that the summing node nonlinearities (2) raise linear combinations of the $\{\beta_1, \ldots, \beta_m\}$ to the n-th power. For the k-th output θ_k, we can write

$$\theta_k \triangleq \left(\sum_{j=1}^{m} a_{j,k} \, \beta_j \right)^n = \left(\sum_{j=1}^{m} a_{j,k} \, \langle \mathbf{b}_j, \mathbf{x} \rangle \right)^n, \qquad k = 1, \ldots, N. \tag{19}$$

We can interpret (19) as weighting the connection between the j-th basis element and the k-th summing node with the gain $a_{j,k}$ (see Figure 1).

We can also write (19) as

$$\theta_k = \left\langle \sum_{j=1}^{m} a_{i,k} \, \mathbf{b}_i \, , \, \mathbf{x} \right\rangle^n = \langle \mathbf{f}_k, \mathbf{x} \rangle^n \, , \tag{20}$$

with

$$\mathbf{f}_k \triangleq \sum_{j=1}^{m} a_{i,k} \, \mathbf{b}_i, \qquad k = 1, \ldots, N \tag{21}$$

a linear combination of the original basis vectors. Equivalently, by collecting the basis (column) vectors into the matrix $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_m]$ and defining $\mathbf{a}_k = [a_{1,k}, \ldots, a_{m,k}]^T$, we can write

$$\mathbf{f}_k = \mathbf{B} \, \mathbf{a}_k, \qquad k = 1, \ldots, N. \tag{22}$$

If the basis vectors $\{\mathbf{b}_i\}$ are viewed as functions with a single "bump" (for example, the delta basis in the time domain, the Fourier basis in the frequency domain, or the wavelet basis in either domain — see Figure 3), then the vectors $\{\mathbf{f}_k\}$ will be functions with multiple "bumps." In this alternative representation, the summing node NST provides an extremely simple structure for generating arbitrary n-th order nonlinear signal interactions. As we see from Figure 2, this representation consists of two decoupled subsystems:

1. an overcomplete set of $N = \binom{m+n-1}{n}$ linear filters $\{\mathbf{f}_k\}_{k=1}^{N}$ that control both the dynamics and component mixing, followed by
2. a set of trivial monomial nonlinearities $(\cdot)^n$.

In Section 5.2, we will apply this representation of the summing node NST to the Volterra filter implementation problem. The filter bank representation not only leads to a simple and effective representation for the computation of a filter output, but also provides insight into the dynamics of the filter.

We now show that the summing node NST is complete. Using tensorial notation, we can write (20) as $\theta_k = \left\langle \mathbf{f}_k^{(n)}, \mathbf{x}^{(n)} \right\rangle$. Following Definition 1, the linear combination $\sum_{k=1}^{N} \alpha_k \, \theta_k = \sum_{k=1}^{N} \alpha_k \left\langle \mathbf{f}_k^{(n)}, \mathbf{x}^{(n)} \right\rangle$. Comparing this expression to (11), we make the identification

$$\mathbf{h} = \sum_{k=1}^{N} \alpha_k \, \mathbf{f}_k^{(n)}, \tag{23}$$

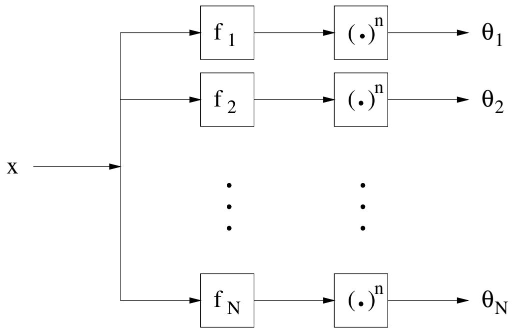


Figure 2: *Filter bank realization of the summing node NST. By combining the basis vectors as in* (21)*, we can decompose an arbitrary summing node NST into a parallel cascade of a redundant set of linear filters* $\{\mathbf{f}_k\}_{k=1}^N$*, each followed by a simple monomial nonlinearity* $(\cdot)^n$*.*

and it follows that this NST is complete if

$$\text{Span}\left\{\mathbf{f}_k^{(n)}\right\}_{k=1}^N \;=\; S^n(\mathbb{R}^m). \tag{24}$$

We will provide three different constructions for complete summing node NSTs. The first is valid for arbitrary nonlinear order n. (For the proof, see Appendix A.)

Theorem 2 *Fix* $\rho \in \mathbb{R}$*,* $|\rho| \neq 1$*,* $\rho \neq 0$*. Set* $\gamma_r = \rho^r$*,* $r = 0, \ldots, n$*. Form the collection of* $N = \binom{m+n-1}{n}$ *length-*m *vectors* $\{\mathbf{a}_k\}_{k=1}^N$ *according to*

$$\{\mathbf{a}_k\}_{k=1}^n \;=\; \left\{[\gamma_{l_1}, \ldots, \gamma_{l_m}]^T :\; \sum_{j=1}^m l_j = n,\;\; l_j \in \{0, \ldots, n\}\right\}. \tag{25}$$

Then, with $\{\mathbf{a}_k\}_{k=1}^N$ *employed in* (19) *or* (22)*, the condition* (24) *holds, and the corresponding summing node NST is complete.*

This construction generates a class of filters $\left\{\mathbf{f}_k^{(n)}\right\}_{k=1}^N$ sufficiently rich for their tensor products to generate all possible n-th order interactions of the basis vectors. While the definition of the combination vectors $\{\mathbf{a}_k\}_{k=1}^N$ in (25) is a notational nightmare, their structure is actually quite simple. Consider an example with $m = 3$, $n = 2$, and $\rho = 2$. Since $n = 2$, the multi-index l_j can take the values $\{0, 1, 2\}$, with corresponding γ_{l_j} values $\{1, 2, 4\}$. The $\{l_j\}_{j=1}^m$ in each $\mathbf{a}_k$ vector must sum to $n = 2$, so the entries in each $\mathbf{a}_k$ will consist of all 1s except for either the single value 4 or a pair of 2s. There are $n = \binom{4}{2} = 6$ combinations of 3-vectors with these nonzero coefficients:

$$\mathbf{a}_1 = [4\;1\;1]^T, \qquad \mathbf{a}_2 = [1\;4\;1]^T, \qquad \mathbf{a}_3 = [1\;1\;4]^T$$
$$\mathbf{a}_4 = [2\;2\;1]^T, \qquad \mathbf{a}_5 = [2\;1\;2]^T, \qquad \mathbf{a}_6 = [1\;2\;2]^T.$$

These coefficients can be interpreted either as 18 weights $a_{j,k}$ to be employed in (19) and Figure 1 or as the combination factors in (22) that generate six different filters for use in Figure 2. In either case, a complete NST results. In Section 5, we consider a cubic example with $n = 3$.

Since Theorem 2 generates $\mathbf{a}_k$ vectors with no zero entries, each $\mathbf{f}_k$ filter will have m "bumps." Larger values of the ρ parameter, however, lead to a simple interpretation of the $\{\mathbf{f}_k\}$. For example, choosing $\rho = 10$ in the $m = 3$, $n = 2$ construction above yields

$$\mathbf{a}_1 = [100\ \ 1\ \ 1]^T, \qquad \mathbf{a}_2 = [1\ \ 100\ \ 1]^T, \qquad \mathbf{a}_3 = [1\ \ 1\ \ 100]^T$$
$$\mathbf{a}_4 = [10\ \ 10\ \ 1]^T, \qquad \mathbf{a}_5 = [10\ \ 1\ \ 10]^T, \qquad \mathbf{a}_6 = [1\ \ 10\ \ 10]^T.$$

Thus, the $\mathbf{f}_1$ channel in Figure 2 will create a quadratic interaction between the signal component lying primarily in the $\mathbf{b}_1$ direction and itself, while the $\mathbf{f}_4$ channel will create a quadratic interaction between signal components lying primarily the $\mathbf{b}_1$ and $\mathbf{b}_2$ directions. This reasoning cannot be carried on ad infinitum, since in the limit as $\rho \to \infty$, a numerically ill-conditioned system results. It could also be tempting to simply subtract 1 from each weight vector above; however, this destroys an important symmetry condition used to prove Theorem 2.

For quadratic summing node NSTs ($n = 2$), we have a very simple alternative construction that clearly reveals the underlying dynamical interaction. In this construction, each filter $\mathbf{f}_k$ equals either a single basis vector or a combination two basis vectors, and the squared output of each filter generates all necessary coupling between different basis elements. The following result is proved in Appendix B.

Theorem 3 *Set $n = 2$ and form the collection of $N = \binom{m+1}{2}$ length-m vectors $\{\mathbf{a}_k\}_{k=1}^N$ according to*

$$\{\mathbf{a}_k\}_{k=1}^N \;=\; \left\{ [\gamma_1, \ldots, \gamma_m]^T \;:\; \sum_{i=1}^m \gamma_i \leq 2,\;\; \gamma_i \in \{0, 1\} \right\}. \tag{26}$$

(Each $\mathbf{a}_k$ is an m-vector with entries of 1 or 0, and each has at most 2 non-zero entries.) Then, with $\{\mathbf{a}_k\}_{k=1}^N$ employed in (19) *or* (22)*, the condition* (24) *holds, and the corresponding second-order summing node NST is complete.*

To complete our study of the summing node NST, we provide a direct construction of a complete set of filters $\{\mathbf{f}_k\}_{k=1}^N$ that bypasses the choice of basis $\{\mathbf{b}_i\}$. Interestingly, randomly generating the filters $\{\mathbf{f}_k\}_{k=1}^N$ produces a complete summing node NST. For the proof, see Appendix C.

Theorem 4 *Let $\{\mathbf{f}_k\}_{k=1}^N$ be a collection of $N = \binom{m+n-1}{n}$ independent and identically distributed observations from an $\mathbb{R}^m$–valued probability density. Then, with probability one,* (24) *holds and the corresponding summing node NST is complete.*

Finally, note that the above constructions for the filters $\{\mathbf{f}_k\}_{k=1}^N$ do not depend on the signal length m. Hence, these constructions can be extended to separable continuous-time spaces.

3.4 Relating the product and summing node structures

It should be noted that the summing node transformation is different from the product node transformation. While both transformations are complete, under the conditions stated previously in this Section, the non-linear signal coefficients $\boldsymbol{\theta}$ are, in general, different for the two structures. However, the coefficients of the two structures can be related by a simple linear transformation. Form $\mathbf{F}^T = \left[\mathbf{f}_1^{(n)}, \ldots, \mathbf{f}_N^{(n)}\right]$ and let $\mathbf{P}^T$ be a matrix whose columns are the $N = \binom{m+n-1}{n}$ tensors $\left\{\mathcal{S}\left(\bigotimes_{j=1}^n \mathbf{b}_{i_j}\right) : 1 \leq i_1 \leq \ldots \leq i_n \leq m\right\}$. The summing node nonlinear signal coefficients are given by

$$\boldsymbol{\theta}_{\text{sum}} = \mathbf{F}\mathbf{x}^{(n)}, \tag{27}$$

while the product node coefficients are given by

$$\boldsymbol{\theta}_{\text{prod}} = \mathbf{P}\mathbf{x}^{(n)}. \tag{28}$$

Since both of these representations are complete, the Moore-Penrose pseudoinverses [1] $\mathbf{F}^\dagger$ and $\mathbf{P}^\dagger$ exist and satisfy $\mathbf{x}^{(n)} = \mathbf{F}^\dagger\, \boldsymbol{\theta}_{\text{sum}} = \mathbf{P}^\dagger\, \boldsymbol{\theta}_{\text{prod}}$. Thus, the vectors $\boldsymbol{\theta}_{\text{sum}}$ and $\boldsymbol{\theta}_{\text{prod}}$ are related by

$$\boldsymbol{\theta}_{\text{prod}} = \mathbf{P}\mathbf{F}^\dagger\, \boldsymbol{\theta}_{\text{sum}}, \tag{29}$$

$$\boldsymbol{\theta}_{\text{sum}} = \mathbf{F}\mathbf{P}^\dagger\, \boldsymbol{\theta}_{\text{prod}}. \tag{30}$$

One advantage of the product node structure is that it produces an orthogonal transformation in the symmetric tensor space, whereas the summing node transformation is never orthogonal. While the product node structure may provide a more efficient representation, the summing node structure has a much simpler and elegant implementation in terms of a redundant filter bank. In Section 5, we will see that this is useful in certain problems.

4 NSTs in the Wavelet Domain

4.1 Choice of NST Basis

The previous Section has shown that complete NSTs can be derived from any orthonormal signal basis $\mathbf{B} = \{\mathbf{b}_j\}_{j=1}^m$. For example, $\mathbf{B}$ may be a delta, Fourier, or wavelet basis [7]. Figure 3 illustrates the relationships and differences between these three different choices.

Classical approaches to nonlinear signal processing and analysis are based in the time or Fourier domain. However, in this section we argue that there are significant advantages to wavelet-based methods. Heuristically speaking, one expects that wavelet-based approaches may provide more robust tools for nonlinear signal processing. This expectation is partly based on the well known compression and regularity properties of the wavelet transform [26]. Furthermore, as illustrated in Figure 3, the wavelet transform provides a joint time-frequency signal analysis, providing added flexibility in comparison to strictly time or frequency based approaches. Unfortunately, it is difficult to quantify benefits of wavelets in a discrete-time setting.

In order to assess the potential advantages of wavelet-based nonlinear signal processing, we will compare the characteristics of NST designed with time, frequency, and wavelet bases in continuous-time (infinite-dimensional) signal spaces. In this setting, we will show that the wavelet basis offers significant advantages over the classical signal bases for nonlinear signal processing. It can be inferred that these advantages carry over into high sample rate discrete-time signal spaces.

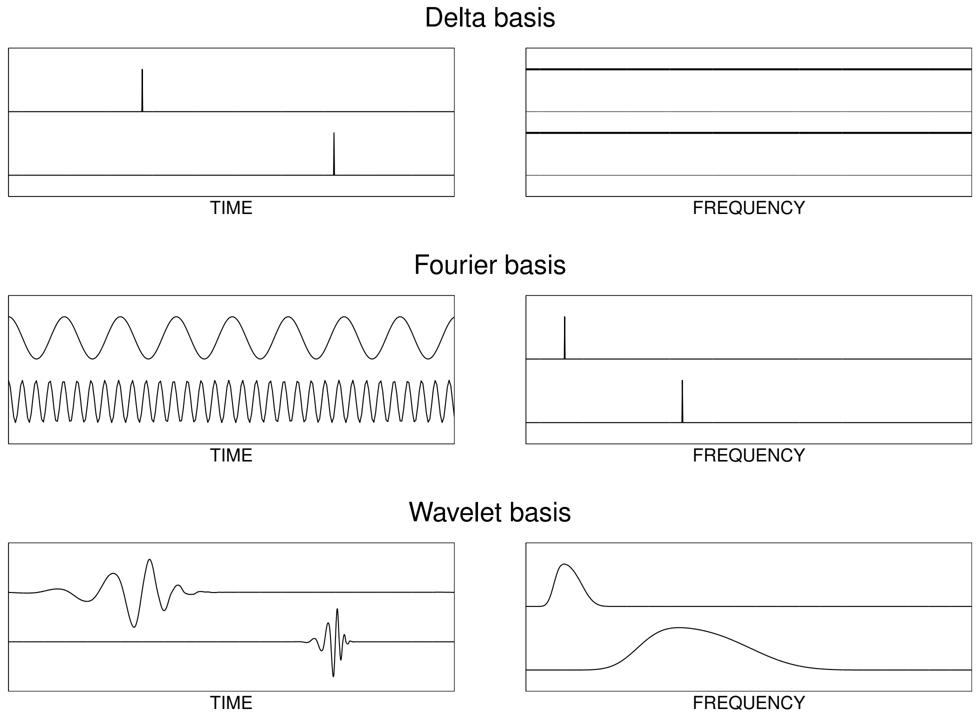


Figure 3: *Comparison of different bases* $\{\mathbf{b}_j\}$ *for nonlinear signal processing. The choice of basis employed in the linear front end of the NST of Figure 1 determines in which domain we represent signal interactions. Consider a second-order NST, which generates squares* β_j^2 *and cross-products* $\beta_i\beta_j$ *of the signal coefficients. We illustrate two basis elements* $\mathbf{b}_i$ *and* $\mathbf{b}_j$ *from three different bases, in both the time domain and the frequency (squared magnitude) domain. In the delta basis, each* $\mathbf{b}_j$ *is a unit pulse, so* β_j *is simply a sample of the signal. The corresponding NST represents coupling between different time lags of the signal. In the Fourier basis, each* $\mathbf{b}_j$ *is a sinusoid, so* β_j *is a Fourier coefficient of the signal. The corresponding NST represents intermodulations between different frequencies. In the wavelet basis, each* $\mathbf{b}_j$ *is localized in both time and frequency simultaneously, so* β_j *measures the time-frequency content of the signal. The corresponding NST represents coupling between different localized wavelet atoms.*

4.2 The wavelet transform

The *wavelet transform* is an atomic decomposition that represents a real-valued continuous-time signal $x(t)$ in terms of shifted and dilated versions of a prototype bandpass wavelet function $\psi(t)$ and lowpass scaling function $\phi(t)$ [7, 16]. For special choices of the wavelet and scaling function, the atoms

$$\psi_{j,k}(t) \;\triangleq\; 2^{-j/2}\,\psi\!\left(2^{-j}t-k\right), \qquad j,k\in\mathbb{Z},\; j\le J \tag{31}$$

$$\phi_{J,k}(t) \;\triangleq\; 2^{-J/2}\,\phi\!\left(2^{-J}t-k\right) \tag{32}$$

form an orthonormal basis, and we have the signal representation [7, 16]

$$x(t) \;=\; \sum_k u_k\,\phi_{J,k}(t) \;+\; \sum_{j=-\infty}^{J}\sum_k w_{j,k}\,\psi_{j,k}(t), \tag{33}$$

with $w_{j,k} \triangleq \int x(t)\,\psi_{j,k}(t)\,dt$ and $u_k \triangleq \int x(t)\,\phi_{J,k}(t)\,dt$. The *wavelet coefficients* $\{w_{j,k}\}$ and *scaling coefficients* $\{u_k\}$ comprise the wavelet transform. For a wavelet centered at time zero and frequency f_0, $w_{j,k}$ measures the content of the signal around the time $2^j k$ and frequency $2^{-j}f_0$ (equivalently, scale j). Wavelet transforms of sampled signals can be computed extremely efficiently using multirate filter bank structures [7, 16].

Recently, 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, moreover, do not introduce excessive artifacts in the signal reconstruction [10]. The explanation for this exceptional performance lies in the fact that wavelet bases are *unconditional bases* for many signal spaces.

A basis $\{z_i\}$ for a Banach space $\mathcal{X}$ is unconditional if there exists a constant $C<\infty$ such that

$$\left\| \sum_{i=1}^{L} \epsilon_i\,c_i\,z_i \right\|_{\mathcal{X}} \;\le\; C \left\| \sum_{i=1}^{L} \epsilon_i\,c_i\,z_i \right\|_{\mathcal{X}}, \tag{34}$$

for every finite set of coefficients $\{c_1,\ldots,c_L\}$ and every set of multipliers $\{\epsilon_1,\ldots,\epsilon_L\}$ of ±1. The notion of unconditionality is important in *transform domain* signal processing for the following reason. If we represent a signal $x\in\mathcal{X}$ in terms of the basis $\{z_i\}$, *i.e.,* $x=\sum_i c_i z_i$, then one can perform transform domain processing by applying attenuation coefficients $\{m_i\}$ to the coefficients $\{c_i\}$:

$$\widetilde{x} \;=\; \sum_i (m_i\,c_i)\,z_i, \qquad |m_i|\le 1. \tag{35}$$

If the basis is unconditional for the space $\mathcal{X}$, then, using (34), for any set of attenuation coefficients the norm of the processed signal $\widetilde{x}$ can be bounded in terms of the norm of the original signal

$$\|\widetilde{x}\|_{\mathcal{X}} \;\le\; C\,\|x\|_{\mathcal{X}}. \tag{36}$$

The unconditional nature of the wavelet basis is crucial to wavelet-domain processing, because it guarantees that the norm of the processed signal will not "blow up" when wavelet coefficients are discarded or reduced in magnitude. Because the wavelet basis is an unconditional basis for many signal spaces, including the L_p, Sobolev, Bounded Variation, Besov, and Triebel spaces [16], this guarantee holds under a wide variety of different signal norms. (The same guarantee does not hold for the Fourier basis, for example.) Obviously, this result has significant implications for signal processing.

The attractive properties of the continuous-time wavelet basis carry over to high-dimensional sampled signal spaces as well. Even though all bases for finite-dimensional signal spaces are unconditional, including Fourier and wavelet bases, and all finite-dimensional norms are equivalent, the constants that relate different finite-dimensional norms are extremely dependent on the dimension. These constants can, in general, grow in an unwieldy manner as we move to higher and higher sample rates (dimensions). The fact that the underlying infinite-dimensional basis is unconditional limits how large the constants grow and consequently guarantees that practical, finite-dimensional wavelet domain processing algorithms will be well behaved under a wide variety of performance measures (all finite-dimensional l_p norms, $1 < p < \infty$, for example).

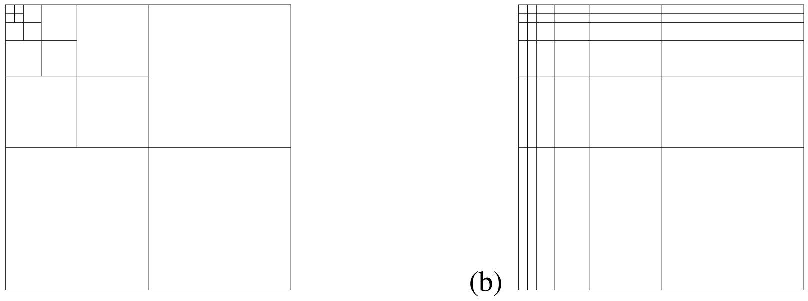
As mentioned above, wavelets form unconditional bases for a diverse variety of signal spaces. However, for NSTs, tensor spaces are the natural framework to consider. Hence, we wish to establish the unconditionality of tensor product wavelet bases. Using the theory of tensor norms and a result from the theory of Gordon-Lewis spaces, we will show that the tensor product of an unconditional basis is again an unconditional basis for a tensor space equipped with an appropriate L_p norm. This result proves that the tensor product of a wavelet basis is an unconditional basis for many tensor spaces of interest. Hence, wavelet-based NSTs inherit the remarkable properties associated with wavelet domain processing. To the authors' knowledge, this is a new result.

It should be noted that the tensor wavelet basis is quite different from the multidimensional wavelet basis obtained via multiresolution analysis [7, 14, 16]. To illustrate the differences, consider the case for functions $x(t_1, t_2)$ of two dimensions. Given a one-dimensional wavelet basis $\{\phi_{J,k}(t)\}_k \cup \{\psi_{j,k}(t)\}_{j \leq J,k}$, the two-dimensional tensor wavelet basis consists of products of all possible pairs of wavelets and scaling functions:

$$\mathbf{B}_{\text{tensor}} = \Big(\{\phi_{J,k}(t_1)\}_k \cup \{\psi_{j,k}(t_1)\}_{j\leq J,k}\Big) \otimes \Big(\{\phi_{J,k}(t_2)\}_k \cup \{\psi_{j,k}(t_2)\}_{j\leq J,k}\Big) \tag{37}$$

$$\begin{aligned} &= \{\phi_{J,k_1}(t_1)\,\phi_{J,k_2}(t_2)\}_{k_1,k_2} \cup \left(\bigcup_{j_1\leq J}\{\psi_{j_1,k_1}(t_1)\,\phi_{J,k_2}(t_2)\}_{k_1,k_2}\right) \\ &\quad \cup \left(\bigcup_{j_2\leq J}\{\phi_{J,k_1}(t_1)\,\psi_{j_2,k_2}(t_2)\}_{k_1,k_2}\right) \cup \left(\bigcup_{j_1,j_2\leq J}\{\psi_{j_1,k_1}(t_1)\,\psi_{j_2,k_2}(t_2)\}_{k_1,k_2}\right). \end{aligned} \tag{38}$$

The tensor basis contains, for example, elements measuring coarse scale (low frequency) information in one direction and fine scale (high frequency) information in the other. To compute the tensor wavelet expansion of a multidimensional function, we simply operate on each coordinate axis separately using a one-dimensional wavelet transform. Neumann and von Sachs have shown that tensor wavelet bases are natural for multidimensional signal estimation applications in signal spaces having differing degrees of smoothness in different directions [18]. In contrast, a multiresolution wavelet basis consists of products of



(a) (b)

Figure 4: *Graphical depiction of the elements of the two-dimensional multiresolution and tensor wavelet bases. The aspect ratios of the tiles correspond roughly to the size of the "regions of support" of the basis elements. (a) The multiresolution wavelet basis consists only of products of pairs of wavelets and scaling functions from the same scale; hence all tiles have the same aspect ratio. (b) The tensor wavelet basis consists of products of pairs of wavelets and scaling functions from all possible scales; hence many different aspect ratios result. (Strictly speaking, these mosaics illustrate the organization of the* coefficients *obtained upon expanding onto these bases. Nevertheless, there is a direct correspondence between the size of a coefficient tile and the size of the region of support of the associated basis elements: Basis functions broad in one direction result in fewer expansion coefficients in that direction and hence a narrower tile.)*

all possible pairs of wavelet and scaling functions *at the same scale:*

$$\begin{aligned} \mathbf{B}_{\text{multi}} &= \{\phi_{J,k_1}(t_1)\,\phi_{J,k_2}(t_2)\}_{k_1,k_2} \\ &\cup \bigcup_{j\leq J} \{\psi_{j,k_1}(t_1)\,\phi_{j,k_2}(t_2),\ \ \phi_{j,k_1}(t_1)\,\psi_{j,k_2}(t_2),\ \ \psi_{j,k_1}(t_1)\,\psi_{j,k_2}(t_2)\}_{k_1,k_2}. \end{aligned} \tag{39}$$

In Figure 4 we illustrate the differences between these bases graphically.

4.3 Unconditional bases for L_p tensor spaces

Let $\{z_i\}$ be a basis for $L_p(\mathbf{T})$, with $\mathbf{T} \subseteq \mathbb{R}$. It follows from the classical result of Gelbaum and Gil de Lamadrid [13] that the tensor basis $\{z_i \otimes z_j\}$ is a basis for the tensor space $L_p(\mathbf{T}) \otimes_{\Delta_p} L_p(\mathbf{T})$, with Δ_p the natural norm. However, this does not guarantee that the tensor product of an unconditional basis is an unconditional basis for the tensor space. We now show that this is indeed the case. (We will work only with second-order tensor spaces for notational convenience; the extension to n-th order tensor spaces $L_p(\mathbf{T}) \otimes_{\Delta_p} \cdots \otimes_{\Delta_p} L_p(\mathbf{T})$ is straightforward.)

First we state a result due to Pisier [24]. Let $\mathcal{Y}$ and $\mathcal{Z}$ be Banach spaces with unconditional bases $\{y_i\}$ and $\{z_i\}$ respectively. Let α be a norm on the tensor space $\mathcal{Y} \otimes \mathcal{Z}$ such that given any two linear operators $U : \mathcal{Y} \to \mathcal{Y}$ and $V : \mathcal{Z} \to \mathcal{Z}$, the tensor product $U \otimes V$ is a bounded linear operator on $\mathcal{Y} \otimes \mathcal{Z}$ equipped with norm α. If this condition holds, then α is called a *uniform norm*. Let $\mathcal{Y} \otimes_\alpha \mathcal{Z}$ denote the completion of $\mathcal{Y} \otimes \mathcal{Z}$ with respect to α.

Theorem 5 [24] *Let $\{y_i\}$ and $\{z_i\}$ be unconditional bases for the Banach spaces $\mathcal{Y}$ and $\mathcal{Z}$, respectively. Let α be a uniform norm for the tensor space $\mathcal{Y} \otimes_\alpha \mathcal{Z}$. Then $\{y_i \otimes z_j\}$ is an unconditional basis for $\mathcal{Y} \otimes_\alpha \mathcal{Z}$ if and only if $\mathcal{Y} \otimes_\alpha \mathcal{Z}$ is a Gordon-Lewis (GL) space.*

Before we can apply this result to $L_p(\mathbf{T}) \otimes_{\Delta_p} L_p(\mathbf{T})$, we must ensure that Δ_p is a uniform tensor nown. To this end, we employ a result due to Beckner.

Theorem 6 [3] *Let $U \otimes V$ be a linear mapping from $[L_q(\mathbf{T}) \otimes_{\Delta_q} L_q(\mathbf{T})]$ to $[L_p(\mathbf{T}) \otimes_{\Delta_p} L_p(\mathbf{T})]$. If $1 \leq q \leq p \leq \infty$, then $\|U \otimes V\| = \|U\| \, \|V\|$.*

It remains only to verify that $L_p(\mathbf{T}) \otimes_{\Delta_p} L_p(\mathbf{T})$ is a GL space. For our purposes it suffices to note the following [9]:

$$L_p(\mathbf{T} \times \mathbf{T}) \text{ is a GL space, for } 1 \leq p \leq \infty. \tag{40}$$

Since $L_p(\mathbb{R}) \otimes_{\Delta_p} L_p(\mathbb{R})$ is isometric to $L_p(\mathbb{R} \times \mathbb{R})$, it follows that $L_p(\mathbf{T}) \otimes_{\Delta_p} L_p(\mathbf{T})$ is also a GL space. Combining these results, we have shown the following:

Theorem 7 *Let $\{z_i\}$ be an unconditional basis for $L_p(\mathbf{T})$, $1 < p < \infty$. Then $\{z_i \otimes z_j\}$ is an unconditional basis for $L_p(\mathbf{T}) \otimes_{\Delta_p} L_p(\mathbf{T})$.*

We have excluded the case $p = 1$, since $L_1(\mathbf{T})$ does not admit unconditional bases [16]. However, more can be said for the subspace of $L_1(\mathbf{T})$ having unconditional wavelet expansions — the Hardy space $H_1(\mathbf{T})$. It follows easily from Theorem 7 that the tensor product of an unconditional basis for $H_1(\mathbf{T})$ is an unconditional basis for the product space $H_1(\mathbf{T} \times \mathbf{T})$. This fact is well-known [16]. (Also, recall that $H_p(\mathbf{T}) = L_p(\mathbf{T})$, $1 < p < \infty$). There are many other tensor spaces of interest, including tensor spaces constructed from Sobolev, Besov, and Triebel spaces. Ongoing work is aimed at assessing the performance of tensor wavelet bases in such spaces.

5 Applications

In this section, we study three applications of NSTs. We first investigate NST-based estimation of correlation functions using the product node architecture and the wavelet basis. Wavelet domain representations of higher-order correlations can be much more efficient than Fourier or time domain representations. In the second application, we demonstrate that the summing node NST is capable of realizing arbitrary Volterra filters. Finally, we examine the potential of truncated wavelet expansions for nonlinear system identification.

5.1 Correlation analysis

The product node NST is well-suited for correlation and higher-order statistical analysis. The n-th order correlation of a random vector $\mathbf{x}$ are given by $\mathrm{E}\left[\mathbf{x}^{(n)}\right]$ [2]. If $\mathbf{x}$ is zero-mean, then the second-order correlation $\mathrm{E}\left[\mathbf{x}^{(2)}\right]$ is simply a vectorized version of the covariance matrix of $\mathbf{x}$, while the third-order correlation $\mathrm{E}\left[\mathbf{x}^{(3)}\right]$ is a vectorized version of the third-order cumulant of $\mathbf{x}$.

It is often advantageous to study the higher-order signal correlations in domains other than time. For example, the n-th order spectrum results from applying the Fourier transform, denoted by $\mathbf{F}$, to $\mathbf{x}$ and

(a) (b)

Figure 5: *Data from an acoustic emission experiment. (a) Emission from a typical trial. (b) Overlay of data from twenty trials.*

computing $\mathrm{E}\left[(\mathbf{F}\mathbf{x})^{(n)}\right]$. The n-th order spectrum measures n-fold correlations between different sinusoidal components of the signal.

If $\mathbf{W}$ denotes the wavelet transform, then $\mathrm{E}\left[(\mathbf{W}\mathbf{x})^{(n)}\right]$ represents the n-th order correlations in the wavelet domain. Because wavelets better match many real-world signals, wavelet domain representations of higher order correlations can be much more efficient — concentrating the dominant correlations in fewer coefficients — than Fourier or time domain representations. This claim is supported by the fact that tensor products of wavelet bases provide unconditional bases for a wide variety of tensor spaces (as shown in Section 4.3).

Now let us examine the product node NST. Let $\mathbf{B}$ denote the orthonormal basis used in the first stage of the structure. The output $\boldsymbol{\theta}$ of the product node transformation of a random vector $\mathbf{x}$ produces all possible n-th order interactions of this vector in the $\mathbf{B}$ domain. If follows that the expected value of the nonlinear signal coefficients $\boldsymbol{\theta}$ produces the n-th order correlations of the process $\mathbf{x}$ in the $\mathbf{B}$ domain. In fact, $\mathrm{E}[\boldsymbol{\theta}]$ contains every unique correlation in $\mathrm{E}\left[(\mathbf{B}\mathbf{x})^{(n)}\right]$.

Now suppose we are given $M \geq 1$ independent and identically distributed (iid) vector observations $\mathbf{x}_1, \ldots, \mathbf{x}_M$. We wish to estimate the n-th order correlations of the underlying process. We can estimate these correlations in the $\mathbf{B}$ domain by computing the product node NST of each observation $F_n : \mathbf{x}_j \mapsto \boldsymbol{\theta}_j$ and then averaging the resulting nonlinear signal coefficients. We estimate $\mathrm{E}[\boldsymbol{\theta}]$ by $\frac{1}{M}\sum_{j=1}^{M} \boldsymbol{\theta}_j$.

We have applied this technique to the problem of acoustic emission signal processing, which is complicated by the complex emission patterns generated by irregularities in the acoustic medium. Such problems arise, for example, in laser optoacoustic tomography for cancer diagnostics. Correlation analyses can aid in illuminating the nature of optoacoustic irregularities in human organs, such as the breast [22]. In the following experiment, $M = 20$ independent acoustic emission trials were performed in the same medium. Emission data for the trials is plotted in Figure 5.

We computed the second-order correlations ($n = 2$) of this data using product node NSTs based in the time, frequency, and wavelet domains. The Daubechies-6 wavelet basis was used in this study [7]. Histograms of the correlation magnitudes were computed for each case and are shown in Figure 6. To quantitatively assess the efficiency of the time, frequency, and wavelet domain representations, the entropy

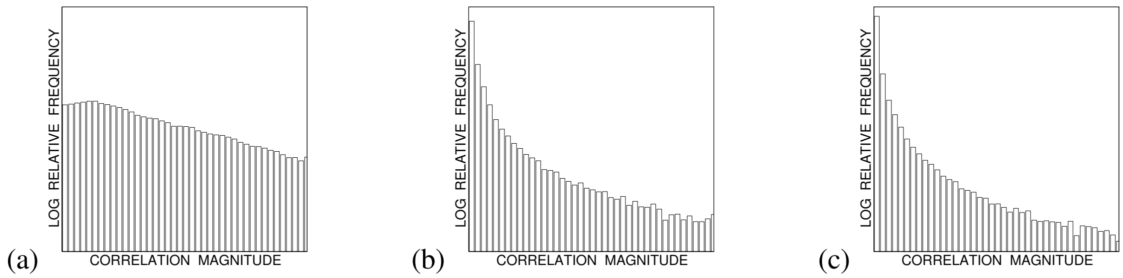


Figure 6: *Histograms of the second-order ($n = 2$) correlations of the acoustic emission signals from Figure 5 in three different basis representations. (a) Histogram of time-domain correlation magnitude, entropy $H = 6.44$. (b) Histogram of frequency-domain correlation magnitude, entropy $H = 2.39$. (c) Histogram of wavelet-domain correlation, entropy $H = 1.82$. To better illustrate both the peakiness and rapid decay of the wavelet-domain correlation, we plot only the first few bins of the histograms on a logarithmic vertical scale.*

of each histogram was computed. The wavelet-domain histogram has a much lower entropy than the time- and frequency-domain histograms, which indicates that the wavelet-domain analysis is more efficient at representing the second-order correlations of the acoustic emission data. Hence, this experiment corroborates theoretical results showing that unconditional bases are optimal for signal compression [10]. Efficient wavelet-based representations can provide more robust and reliable estimates of the higher-order statistics and could provide better insight into the complicated non-stationary correlation structure of the data.

5.2 Volterra filtering

In this Section, we consider Volterra filter realizations based on the NST. We show that a complete n-th order NST is capable of realizing every n-th order Volterra filter. In particular, the summing node transformation leads to an elegant filter bank representation.

The output of a homogeneous n-th order Volterra filter applied to a signal $\mathbf{x} = [x_1, \ldots, x_m]^T$ is given by [15]

$$y = \sum_{1 \le i_1 \le \ldots \le i_n \le m} h_{i_1,\ldots,i_n}\, x_{i_1} \cdots x_{i_n}. \tag{41}$$

The filter output y is simply an n-th order multilinear combination of the samples $x_1, \ldots, x_m$. The set of weights h is called the n-th order *Volterra kernel*. Note that while (41) computes only a single output value given m input values, the extension to online processing of infinite-length signals is straightforward. To treat the input signal x_l, we simply set $\mathbf{x}_l = [x_l, \ldots, x_{l-m+1}]^T$, with m the *memory length* of the filter. The output of (41) is then y_l, a nonlinearly filtered version of x_l.

Since (41) is identical to the multilinear functional (9) appearing in Definition 1, it follows that every n-th order Volterra filter can be computed as a linear combination of the nonlinear signal coefficients $\boldsymbol{\theta} = F_n(\mathbf{x})$. As shown in Section 3, both the product node and summing node structures are capable of computing a complete n-th order signal transformation. The summing node structure is particularly interesting in this application, because it allows us to represent every n-th order Volterra filter using the simple filter bank of Figure 7. Key to this scheme is that *the overcomplete linear transformation, rather than the nonlinearities, manage the signal coupling prescribed by the overall Volterra filter.* Therefore, this new representation

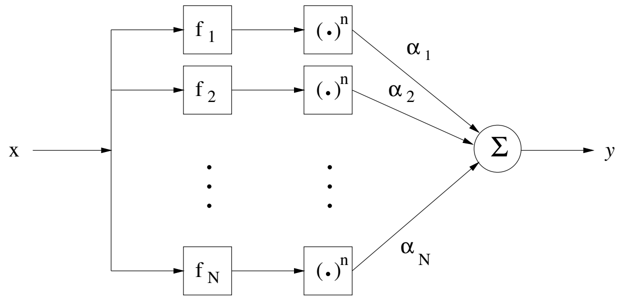


Figure 7: *Volterra filter realization using a summing node NST.*

greatly simplifies the analysis, synthesis, and implementation of Volterra filters.[1]

Volterra filter realizations of this type are often referred to as *parallel-cascade realizations* [23]. Previous work on parallel-cascade designs has relied on complicated numerical optimizations to construct kernel-specific sets of linear filters and, hence, each distinct Volterra filter requires its own unique parallel-cascade realization [5, 23]. In contrast, the summing node NST can represent every n-th order Volterra filter simply by adjusting the output weights $\{\alpha_k\}_{k=1}^N$. The linear filters $\{\mathbf{f}_k\}_{k=1}^N$ of the summing node structure remain the same for every Volterra kernel. Hence, the summing node structure is a *universal structure* for homogeneous Volterra filtering. Nonhomogeneous Volterra filters can also be implemented with the summing node structure by following each linear filter with an n-th degree polynomial nonlinearity instead of the homogeneous n-th order monomial. Moreover, if the Volterra kernel h is low-rank, then it can be represented exactly with a smaller subset ($r < m$) of orthonormal basis vectors [21]. Therefore, low-rank systems can be implemented with a far smaller filter bank.[2]

The weights $\{\alpha_k\}$ corresponding to a specific Volterra filter with kernel h can be computed by solving a system of linear equations. Let $\mathbf{h}$ be a vectorized version of h ordered to correspond to the Kronecker product in (11). According to (23), the Volterra kernel generated by the summing node NST is given by $\sum_{k=1}^N \alpha_k \mathbf{f}_k^{(n)}$. Therefore, to represent the Volterra filter with kernel $\mathbf{h}$ we chose the weights $\{\alpha_k\}$ so that $\sum_{k=1}^N \alpha_k \mathbf{f}_k^{(n)} = \mathbf{h}$. The proper weights are readily obtained by solving this system of linear equations.

As an example, consider the implementation of a homogeneous third-order ($n = 3$) Volterra filter using the summing node NST. Let $\mathbf{B} = \{\mathbf{b}_j\}_{j=1}^m$ be an orthonormal basis for $\mathbb{R}^m$. For example, $\mathbf{B}$ could be the delta, Fourier, or wavelet basis. We design the filters $\mathbf{f}_1, \ldots, \mathbf{f}_N$, $N = \binom{m+2}{3}$, for the summing node transformation using the construction of Theorem 2. Referring to the Theorem, we take $\rho = 2$ and hence $\gamma_0 = 1, \gamma_1 = 2, \gamma_2 = 4, \gamma_3 = 8$. Each filter $\mathbf{f}_k$, $k = 1, \ldots, N$, is a linear combination of the basis vectors: $\mathbf{f}_k = \mathbf{B}\mathbf{a}_k$, with $\mathbf{a}_k$ a vector with elements in the set $\{1, 2, 4, 8\}$. Each $\mathbf{a}_k$ consists of all 1s except for either a single 8, a 2 paired with a 4, or three 2s. Raising the output of each filter to the third power generates third-order interactions between the different distinct components of the input signal represented by the basis

[1]The canonical representation of the Volterra filter (41) is of limited utility, due to the inherent difficulty in interpreting the multidimensional kernel h (particularly when $n > 2$).

[2]Using the *tensor product basis approximation* to the low-rank kernel [21], we can represent the kernel exactly with a filter bank consisting of $\binom{n+r-1}{r} < N$ filters, with $r < m$ the rank of the kernel h. This is particularly useful if the kernel is known to satisfy certain constraints (for example, smoothness, bandlimitedness).

vectors. Taken together, these filters collaborate to generate all possible third-order nonlinear interactions of the signal.

Different types of interactions are produced depending on the choice of basis. The delta basis produces interactions between different time samples of the signal. The Fourier basis yields frequency intermodulations, whereas the wavelet basis produces interactions between wavelet atoms localized in both time and frequency. The fact that wavelet tensor bases are unconditional bases for many tensor spaces suggests that wavelets may provide a more parsimonious representation for Volterra filters than time- or frequency-domain representations.

5.3 Nonlinear System Identification

One common application of Volterra filters is nonlinear system identification [12, 15, 21]. To illustrate the use of the tensor wavelet basis in this context, consider the following problem. Assume that we observe the input and output of a nonlinear system defined by the bilinear operator

$$y(t) \;=\; \iint_{t_1,t_2} h(t_1,t_2)\, x(t-t_1)\, x(t-t_2)\, dt_1\, dt_2. \tag{42}$$

This type of quadratic nonlinearity arises in the analysis of audio loudspeakers, for example [12]. We assume that both the input and output signals are sampled, resulting in the following discrete-time Volterra system

$$y_k \;=\; \sum_{i,j=1}^{m} h_{i,j}\, x_{k-i}\, x_{k-j}. \tag{43}$$

The discretized kernel $h_{i,j}$ can be estimated from the input and output samples using correlation techniques. However, in real applications only a finite number of samples are available and often additive noise is present in the observations. Consequently, the kernel estimates obtained from short data records are "noisy".

Noise can be removed from a kernel estimate by processing the estimate in the Fourier or wavelet domain. Because the Fourier and wavelet bases often provide a concise representation of the kernel, in many cases the separation of the true kernel from the noise can be carried out very easily in these domains. Specifically, the noise in a "raw" kernel estimate can be removed by truncating a Fourier or wavelet expansion of the estimate. In the following example, we will show that wavelet-domain noise removal can outperform Fourier-domain processing.

To illustrate this point, we simulate the identification of the nonlinear system (43) with the quadratic kernel h depicted in Figure 8(a). This kernel was obtained from actual measurements on an audio loudspeaker [12]. In our simulation, we treat this kernel as an "unknown" model we wish to identify. Using an iid zero-mean, unit-variance Gaussian input sequence $\mathbf{x}$ to "probe" the system, we computed the outputs y according to (43) (with no additive observation noise). In total, we generated 5000 input and output measurements and, from these two sequences, identified the kernel using the following correlation estimator. Note that

$$\mathrm{E}\,[Y X_i\, X_j] \;=\; \begin{cases} 2\,h_{i,i} + \sum_k h_{k,k}, & i=j \\ 2\,h_{i,j}, & i \neq j. \end{cases} \tag{44}$$

Letting $s_{i,j}$ denote the sample average estimate of $\mathrm{E}[Y X_i X_j]$, we have the following estimator for the Volterra kernel[3]

$$\widehat{h}_{i,j} = \begin{cases} \frac{1}{2}\left(s_{i,i} - \frac{1}{m+2}\sum_k s_{k,k}\right), & i = j \\ \frac{1}{2}\, s_{i,j}, & i \neq j. \end{cases} \tag{45}$$

While the simple correlation estimator (45) converges to the true kernel, with only a finite number of data the resulting estimate is typically very noisy due to the variability of the sample correlation estimator about the true correlation values. A simple noise removal procedure is based on expanding this "raw" kernel estimate in an orthonormal basis and then discarding the small terms of this expansion (which presumably correspond to noise and not signal). Let $\left\{\widehat{\theta}_l\right\}$ denote the coefficients of the raw estimate $\widehat{h}$ in the basis expansion. Then the coefficients $\left\{\widetilde{\theta}_l\right\}$ of the truncated series kernel estimate $\widetilde{h}$ can be expressed in terms of a hard threshold applied to the coefficients $\left\{\widehat{\theta}_l\right\}$

$$\widetilde{\theta}_l = \begin{cases} \widehat{\theta}_l, & \left|\widehat{\theta}_l\right| \geq \tau \\ 0, & \left|\widehat{\theta}_l\right| < \tau, \end{cases} \tag{46}$$

with τ the threshold level. Many choices for the threshold value are possible; usually τ is chosen based on some estimate of the noise level in the data. The better the basis "matches" the true kernel, the more efficient this procedure will be at noise removal. In our experiment, we expanded the raw estimate in the wavelet and Fourier tensor bases and then discarded the terms in the expansions whose coefficient magnitudes fell below the threshold value $\tau = \sqrt{2\,\log(m^2)}\,\sigma$, with σ the standard deviation of the noise and $m^2 = 1024$ the dimension of the discretized kernel. This threshold choice is suggested in [11] as a probabilistic upper bound on the noise level. In practice, σ must be estimated from the raw kernel estimate $\widehat{h}$. However, since we had access to the true kernel in this simulation, we computed σ directly from the difference between the true kernel and the raw estimate. Figure 8(c) and (d) show the estimates that result from hard thresholding in the wavelet domain and Fourier domain, respectively. Wavelet-based truncation provides a much better kernel estimate than both the original raw estimate (b) and the truncated Fourier expansion estimate (d). In fact, the Fourier-based method oversmooths the estimate and results in a worse mean-squared-error (MSE) than that of the original raw estimate. This simulation demonstrates the utility of wavelet-based representations for the analysis of real-world nonlinear systems.

6 Conclusions

In this paper, we have developed two new structures for computing n-th order NSTs. The product and summing node NSTs, while simple, can represent all n-th order nonlinear signal interactions. Both transformations have an elegant interpretation in terms of tensor spaces. The product node NST yields an orthogonal transformation in the tensor space, and thus it is especially well-suited to estimation problems. The summing

[3]The estimator $\widehat{h}$ is derived as follows. Let $\mu_{i,i} = E[YX_i^2]$. Setting $\mu_{i,i} = 2\,h_{i,i} + \sum_k h_{k,k}$ and re-arranging gives $2h_{i,i} = \mu_{i,i} - \sum_k h_{k,k}$. Summing over i produces $2\sum_i h_{i,i} = \sum_i \mu_{i,i} - m\sum_k h_{k,k}$ and hence $\frac{1}{m+2}\sum_i \mu_{i,i} = \sum_i h_{i,i}$. A simple substitution then yields $h_{i,i} = \left(\mu_{i,i} - \frac{1}{m+2}\sum_k \mu_{k,k}\right)/2$. The estimator $\widehat{h}_{i,i}$ is obtained by substituting the sample cross-moments $s_{i,i}$ for the true moments $\mu_{i,i}$. The estimator $\widehat{h}_{i,j}$, $i \neq j$, is obtained in a similar fashion.

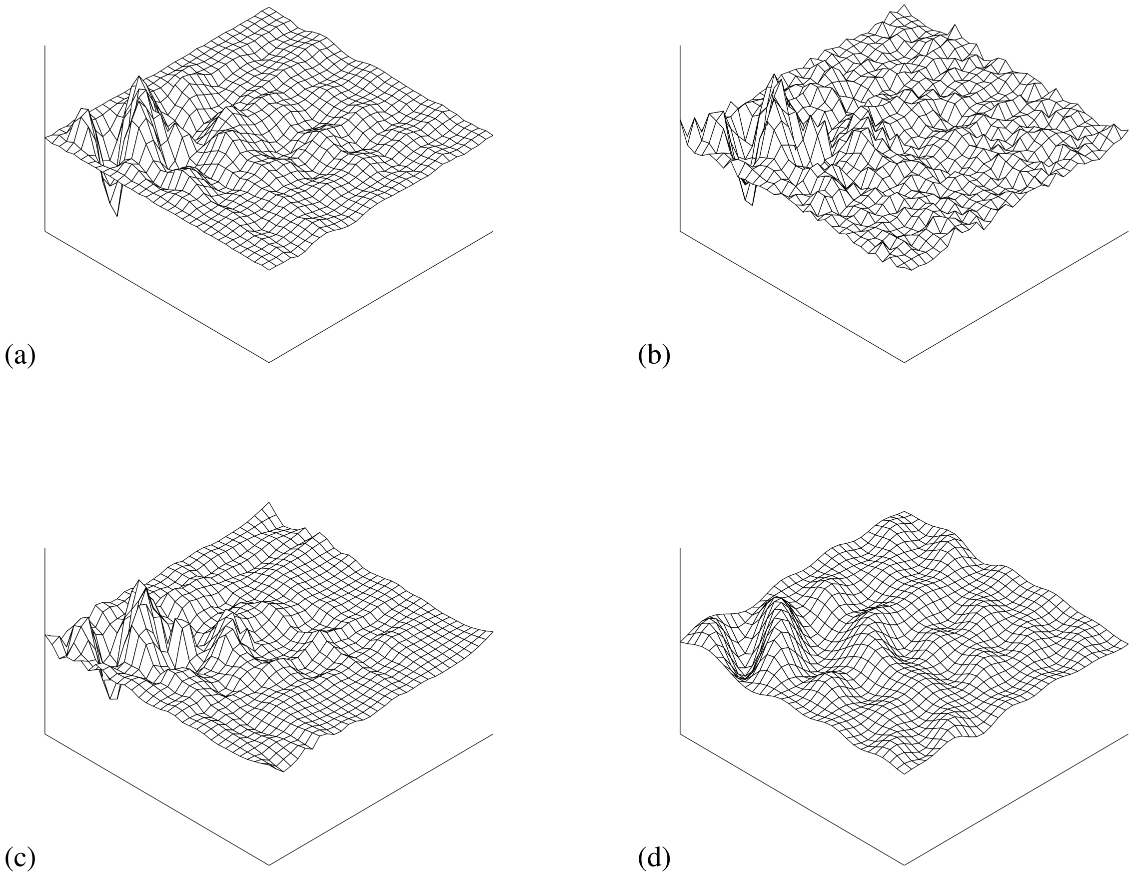


Figure 8: *Nonlinear system identification using tensor bases. Estimates of the quadratic Volterra kernel of an audio loudspeaker obtained using thresholding in a tensor basis expansion. (a) True kernel h [12]. (b) Raw estimate $\widehat{h}$ obtained using* (45), *MSE=0.20. (c) Estimate $\widetilde{h}$ obtained through a truncated Daubechies-8 wavelet expansion of $\widehat{h}$, MSE=0.15. (d) Estimate $\widetilde{h}$ obtained through truncated Fourier expansion of $\widehat{h}$, MSE=0.40. The wavelet estimator provides a more faithful estimate than the Fourier estimator, which oversmooths.*

node NST results in a redundant filter bank structure natural both for analyzing and interpreting nonlinear interactions and for designing efficient implementations. Not only does the summing node architecture suggest new, efficient algorithms for nonlinear processing, it also decouples the processing into linear dynamics and static nonlinearities. Hence, this new framework for nonlinear analysis and processing may provide new insights into the inner workings of nonlinear systems.

NSTs are not constrained to a fixed choice of basis. However, we have shown that wavelet bases provide an optimal framework for NSTs in the sense that wavelet tensor bases are unconditional for many important tensor spaces. Because the wavelet basis provides a more concise representation of many real-world signals, more robust estimates of higher-order moments/cumulants/spectra and Volterra kernels may be obtained via the wavelet representation as compared to time- or frequency-domain approaches.

Finally, we have focused on the classical L_p tensor spaces in our theoretical analysis of wavelet-domain nonlinear processing. However, new results in the statistical literature suggest that more general spaces such as Besov and Triebel spaces are extremely useful for characterizing real-world signals [10]. Therefore, an important avenue for future work will be to extend the results of this paper to these more general settings, possibly using the results of [27]. Another issue currently under investigation is the relationship between polynomial-based processing (higher-order statistics, Volterra filters) and other important types of nonlinear processing that use sigmoidal (neural networks), threshold (wavelet shrinkage) [10], or weighted highpass nonlinearities [20]. The results of this paper could serve as a link between these important areas of nonlinear signal processing.

A Proof of Theorem 2

To prove the Theorem, we must show that the set $\{\mathbf{f}_k\}_{k=1}^N$ generated by (22) and (25) satisfies (24). That is, we must show that the tensors $\left\{\mathbf{f}_k^{(n)}\right\}$ span the symmetric tensor space $S^n(\mathbb{R}^m)$. Recall that each filter $\mathbf{f}_k$ has the form

$$\mathbf{f}_k \;=\; \sum_{j=1}^{m} a_{j,k}\,\mathbf{b}_j, \tag{47}$$

with $\mathbf{a}_k = [a_{1,k}, \ldots, a_{m,k}]^T$. Now consider

$$\begin{aligned}\mathbf{f}_k^{(n)} &= \left(\sum_{j=1}^{m} a_{j,k}\mathbf{b}_j\right)^{(n)} \\ &= \sum_{1\le i_1\le\ldots\le i_n\le m} a_{i_1,k}\cdots a_{i_n,k}\,\mathbf{b}_{i_1}\otimes\cdots\otimes\mathbf{b}_{i_n}.\end{aligned} \tag{48}$$

With Ω denoting the set of permutations of the set $\{1,\ldots,n\}$, we can write

$$\mathbf{f}_k^{(n)} \;=\; \sum_{1\le i_1\le\ldots\le i_n\le m} a_{i_1,k}\cdots a_{i_n,k}\left(\sum_{\omega\in\Omega}\mathbf{b}_{i_{\omega(1)}}\otimes\cdots\otimes\mathbf{b}_{i_{\omega(n)}}\right). \tag{49}$$

According to Theorem 2.6 in [28], the collection of tensors

$$\left\{\sum_{\omega\in\Omega}\mathbf{b}_{i_{\omega(1)}}\otimes\cdots\otimes\mathbf{b}_{i_{\omega(n)}} \;:\; 1\le i_1\le\cdots\le i_n\le m\right\} \tag{50}$$

is a basis for $S^n(\mathbb{R}^m)$. Let $\{\mathbf{s}_k\}_{k=1}^N$, $N = \binom{m+n-1}{n}$, denote these basis vectors, and set $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]^T$. Then, from (49), we can write

$$\mathbf{f}_k^{(n)} = \mathbf{S}\mathbf{u}_k, \tag{51}$$

with $\mathbf{u}_k$ an $N \times 1$ column vector containing all degree-n monomials in the $a_{1,k}, \ldots, a_{m,k}$ coefficients. Next, define

$$\mathbf{F} = \left[\mathbf{f}_1^{(n)}, \ldots, \mathbf{f}_n^{(n)}\right] = \mathbf{S}\mathbf{U}, \tag{52}$$

with $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$. $\mathbf{U}$ is an $n \times n$ matrix and, since $\mathbf{S}$ is a basis for $S^n(\mathbb{R}^m)$, it follows that $\mathrm{Span}\left\{\mathbf{f}_k^{(n)}\right\}_{k=1}^N = S^n(\mathbb{R}^m)$ if and only if $\mathbf{U}$ is invertible.

The remainder of the proof shows that the invertibility of $\mathbf{U}$ is guaranteed if the Vandermonde matrix

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \gamma_0 & \cdots & \gamma_n \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \gamma_0^n & \cdots & \gamma_n^n \end{bmatrix} \tag{53}$$

is invertible.

First, we show that $\mathbf{U} = \mathbf{Q}^T\mathbf{V}^{(m)}\mathbf{Q}$, with $\mathbf{Q}$ a full rank matrix defined as follows. Consider the weights $a_{1,k}, \ldots, \mathbf{a}_{m,k}$. Form a vector of products of these weights using the Kronecker product:

$$\mathbf{q}_k = \begin{bmatrix} 1 \\ a_{1,k} \\ a_{1,k}^2 \\ \vdots \\ a_{1,k}^n \end{bmatrix} \otimes \begin{bmatrix} 1 \\ a_{2,k} \\ a_{2,k}^2 \\ \vdots \\ a_{2,k}^n \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ a_{m,k} \\ a_{m,k}^2 \\ \vdots \\ a_{m,k}^n \end{bmatrix}. \tag{54}$$

Note that every monomial in $\mathbf{u}_k$ is included in the vector $\mathbf{q}_k$. Define the matrix $\mathbf{Q}$ as the map $\mathbf{Q}^T : \mathbf{q}_k \mapsto \mathbf{u}_k$. Note that $\mathbf{Q}$ is full rank and does not depend on k.

Now consider $\mathbf{V}^{(m)}$, the m-fold tensor product of the Vandermonde matrix $\mathbf{V}$. Each row of $\mathbf{V}^{(m)}$ corresponds to a particular monomial form ($a_{1,k}^{l_1} \cdots a_{m,k}^{l_m}$, for example). Each column of $\mathbf{V}^{(m)}$ corresponds to a particular set of weights ($[\gamma_{l_1}, \ldots, \gamma_{l_m}]^T = \mathbf{a}_k$ from (25), for example). Now consider $\mathbf{Q}^T\mathbf{V}^{(m)}\mathbf{Q}$. The action of $\mathbf{Q}^T$ on the left extracts the rows of $\mathbf{V}^{(m)}$ corresponding to degree-n monomials. Applying $\mathbf{Q}$ on the right extracts the columns of $\mathbf{V}^{(m)}$ corresponding to the specific weights in the N vectors defined in (25). Therefore, $\mathbf{U} = \mathbf{Q}^T\mathbf{V}^{(m)}\mathbf{Q}$. The special construction of the vectors $\{\mathbf{a}_k\}_{k=1}^N$ in (25) should now be clear: The values $\mathbf{a}_k = [\gamma_{l_1}, \ldots, \gamma_{l_m}]^T$ correspond to the powers in one of the monomials $a_{1,k}^{l_1} \cdots a_{m,k}^{l_m}$, and hence by applying $\mathbf{Q}$ to both sides of $\mathbf{V}^{(m)}$ we select the n-th order monomials with the values corresponding to the weights in the vectors $\{\mathbf{a}_k\}_{k=1}^N$.

We now claim that taking $\gamma_r = \rho^r$, $r = 0, \ldots, n$, $|\rho| \neq 1$, $\rho \neq 0$, in (53) implies that $\mathbf{U}$ is invertible. Note that with this choice $\mathbf{V}$ is real-symmetric and invertible. It follows that $\mathbf{V}^{(m)}$ is also real-symmetric and invertible. Therefore, since $\mathbf{Q}$ has full rank, $\mathbf{U} = \mathbf{Q}^T\mathbf{V}^{(m)}\mathbf{Q}$ is also real-symmetric and invertible.[4] This completes the proof. □

[4] If $\mathbf{U}$ is not invertible, then there exists a vector $\mathbf{x}$ such that $\mathbf{x}^T\mathbf{U}\mathbf{x} = 0$. This implies the existence of $\mathbf{y} = \mathbf{Q}\mathbf{x}$ such that $\mathbf{y}^T\mathbf{V}^{(m)}\mathbf{y} = 0$ and hence $\mathbf{V}^{(m)}$ has a zero eigenvalue, contradicting the assumption that $\mathbf{V}^{(m)}$ is invertible.

B Proof of Theorem 3

We must show that Span $\left\{\mathbf{f}_k^{(2)}\right\}_{k=1}^N = S^2(\mathbb{R}^m)$. Following the proof of Theorem 2, we have

$$\mathbf{f}_k^{(2)} \;=\; \mathbf{S}\mathbf{u}_k, \tag{55}$$

with $\mathbf{u}_k$ an $N\times 1$ column vector containing all degree 2 monomials in the elements of $\mathbf{a}_k = [a_{1,k}, \ldots, a_{m,k}]^T$ and

$$\left[\mathbf{f}_1^{(2)}, \ldots, \mathbf{f}_N^{(2)}\right] \;=\; \mathbf{S}\mathbf{U}, \tag{56}$$

with $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]$. It follows that Span $\left\{\mathbf{f}_k^{(2)}\right\}_{k=1}^N = S^2(\mathbb{R}^m)$ if and only if $\mathbf{U}$ is invertible.

To show that $\mathbf{U}^{-1}$ exists in this case, let us take a closer look at the columns of $\mathbf{U}$. Recall that each column of $\mathbf{U}$ is denoted by $\mathbf{u}_k$ and is generated by computing all cross-products between the elements of $\mathbf{a}_k$ (Each $\mathbf{a}_k$ is an m-vector with entries of 1 or 0, and each has at most 2 non-zero entries). Consider first the columns $\mathbf{u}_k$ that correspond to $\mathbf{a}_k$ vectors with a single non-zero entry. These columns also contain a single non-zero entry. For example, if $a_{i,k} = 1$ and $a_{j,k} = 0$ $(i \neq j)$, then $\mathbf{u}_k$ has a single non-zero entry in the position corresponding to the monomial $a_{i,k}^2 = 1$. There are a total of m such columns, each with a single 1 in a unique location corresponding to such an product. Clearly, these columns are linearly independent of one another, as each has a single non-zero entry in a different location. Now consider the columns $\mathbf{u}_k$ that correspond to $\mathbf{a}_k$ vectors with a two non-zero entries. If $a_{i_1,k} = 1$, $a_{i_2,k} = 1$, and $a_{j,k} = 0$ $(j \neq i_1, i_2)$, then the column $\mathbf{u}_k$ has non-zero entries in the location corresponding the cross-product $a_{i_1,k}\, a_{i_2,k}$. Note that since no other $\mathbf{a}_l$ $(l \neq k)$ will have non-zero values in both the i_1 and i_2 position, the corresponding $\mathbf{u}_l$ will be zero in the associated cross-product location. Therefore, all $\mathbf{u}_k$ are linearly independent. This completes the proof. $\square$

One might wonder whether this construction using binary weights can be extended to higher orders $n > 2$. Unfortunately, the answer is negative. As we move to higher orders, we require more diversity in the weights used to form the linear filters $\{\mathbf{f}_k\}$. Hence, we require a more complicated construction such as that of Theorem 2.

C Proof of Theorem 4

Recall that the summing node decomposition is complete if and only if for every $\mathbf{h} \in S^n(\mathbb{R}^m)$ there exist $\{\alpha_k\}_{k=1}^N$ such that

$$\mathbf{h} \;=\; \sum_{k=1}^{N} \alpha_k\, \mathbf{f}_k^{(n)}. \tag{57}$$

Here, $\mathbf{h}$ is symmetric and hence contains repeated elements. Also, vectors such as $\mathbf{f}_k^{(n)}$ contain repeated products. To avoid such redundancies, define the vectors $\widetilde{\mathbf{h}}$ and $\left\{\widetilde{\mathbf{f}_k^{(n)}}\right\}$ from which repeated elements in the original vectors $\mathbf{h}$ and $\left\{\mathbf{f}_k^{(n)}\right\}$ have been discarded. For example, if $n = 2$ and $\mathbf{f}_k = [f_1, f_2]^T$, then both $f_1 f_2$ and $f_2 f_1$ occur in $\mathbf{f}_k^{(2)}$. In this case $\widetilde{\mathbf{f}_k^{(2)}} = [f_1^2,\ f_1 f_2,\ f_2^2]^T$. In general, for nonlinear order n and signal length m, the vectors $\widetilde{\mathbf{h}}$ and $\left\{\widetilde{\mathbf{f}_k^{(n)}}\right\}$ each contain exactly $N = \binom{m+n-1}{n}$ elements. With this notation in place, (57) can be rewritten as

$$\widetilde{\mathbf{h}} = \sum_{k=1}^{N} \alpha_k \, \widetilde{\mathbf{f}_k^{(n)}}. \tag{58}$$

To guarantee that the summing node structure can represent every $\mathbf{h} \in S^n(\mathbb{R}^m)$, we must have that $\mathrm{Span}\left\{\widetilde{\mathbf{f}_k^{(n)}}\right\}_{k=1}^{N} = \mathbb{R}^N$.

To determine a set of spanning vectors, consider the following argument. Suppose that we randomly choose the vectors $\{\mathbf{f}_k\}$ as independent realizations from a common probability distribution. Furthermore, assume that this distribution has a density (that is, the distribution is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^m$). Then $\mathrm{Span}\left\{\widetilde{\mathbf{f}_k^{(n)}}\right\}_{k=1}^{N} = \mathbb{R}^N$ with probability one. This follows from the following result regarding the invertibility of the n-th order moment matrix of an $\mathbb{R}^m$–valued random vectors.

Lemma 1 [19] *If* $\mathbf{f}$ *is an* $\mathbb{R}^m$*–valued random vector having a density, then* $\mathrm{E}\left[\widetilde{\mathbf{f}^{(n)}}\,\widetilde{\mathbf{f}^{(n)}}^T\right]$ *is invertible.*

To see how this result relates to the problem at hand, let $\widetilde{\mathbf{F}} = \left[\widetilde{\mathbf{f}_1^{(n)}}, \ldots, \widetilde{\mathbf{f}_N^{(n)}}\right]$, and note that the matrix $\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T$ can be viewed as the sample n-th order moment matrix of the density for $\{\mathbf{f}_k\}_{k=1}^{N}$. Theorem 1 implies that the sample moment matrix $\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T$ is invertible with probability one if the number of samples is greater than N (see Remark 5.2 in [19]). This in turn implies that $\widetilde{\mathbf{F}}$ is full rank and that $\mathrm{Span}\left\{\widetilde{\mathbf{f}_k^{(n)}}\right\}_{k=1}^{N} = \mathbb{R}^N$. $\square$

Acknowledgments: The authors thank Klaus Floret, Michael Fraizer, and Yves Meyer for their helpful comments and suggestions; Alexander Oraevsky for providing the acoustic emission data; and Walter Frank for supplying the loudspeaker kernel measurements.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1989.

[2] T. André, R. Nowak, and B. Van Veen, "Low Rank Estimation of Higher Order Statistics," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '96*, Atlanta, GA, May 1996.

[3] W. Beckner, "Inequalities in Fourier Analysis," *Annals of Math.*, 102, pp. 159-182, 1975.

[4] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, Vol. 25, pp. 772-781, 1978.

[5] H.-H. Chiang, C. L. Nikias, and A. N. Venetsanopoulos, "Efficient implementations of quadratic filters," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 34, pp. 1511-1528, 1986.

[6] K. C. Chou and L. P. Heck, "A multiscale stochastic modelling approach to the monitoring of mechanical systems," *IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, Philadelphia, pp. 25–27, 1994.

[7] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.

[8] A. Defant and K. Floret, *Tensor Norms and Operator Ideals*, Elsevier, 1993.

[9] J. Diestel, H. Jarchow, and A. Tonge, *Absolutely Summing Operators*, Cambridge University Press, 1995.

[10] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *Applied Comp. Harm. Anal.*, pp. 100-115, vol. 1, no. 1, Dec. 1993.

[11] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, pp. 613-627, vol. 41, no. 3, May 1995.

[12] W. Frank, "An efficient approximation to the quadratic Volterra kernel and its application in realtime loudspeaker linearization," *Signal Processing*, pp. 97-113, vol. 45, no. 1, 1995.

[13] B. R. Gelbaum and J. Gil de Lamadrid, "Bases of tensor products of Banach spaces," *Pacific J. Math.*, 11, pp. 1281-1286, 1961.

[14] B. Jawerth and W. Sweldens, "An Overview of Wavelet Based Multiresolution Analysis," *SIAM Rev.*, vol.36, no. 3, pp.377-412, 1994.

[15] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Proc. Mag.*, vol. 8, no. 3, pp. 10-26, 1991.

[16] Y. Meyer, *Wavelets and Operators*, Cambridge University Press, 1992.

[17] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework*, Prentice-Hall, 1993.

[18] M. H. Neumann and R. von Sachs, "Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra," preprint.

[19] R. Nowak and B. Van Veen, "Invertibility of Higher Order Moment Matrices," *IEEE Trans. Signal Proc.*, vol. 43, no. 3, pp. 705-708, 1995.

[20] R. D. Nowak and R. G. Baraniuk, "Optimal weighted highpass filters using multiscale analysis," *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, pp. 224-229, San Antonio, 1996.

[21] R. Nowak and B. Van Veen, "Tensor Product Basis Approximations for Volterra Filters," *IEEE Trans. Signal Proc.*, vol. 44, no. 1, pp. 36-50, January 1996.

[22] A. Oravesky, personal communication, 1996.

[23] T. M. Panicker and V.J. Mathews, "Parallel-cascade realizations and approximations of truncated Volterra systems," *Proceedings of the 1996 IEEE Intl. Conf. Acoustics, Speech, and Signal Proc.*, pp. 1589-1592, Atlanta, 1996.

[24] G. Pisier, *Factorization of Linear Operators and Geometry of Banach Spaces*, AMS, 1986.

[25] G. L. Sicuranza, "Quadratic filters for signal processing," *Proc. IEEE*, vol. 90, p p. 1263-1285, Aug. 1992.

[26] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[27] W. Wang, "A discrete transform and Triebel-Lizorkin spaces on the bidisc," *Trans. Amer. Math. Soc.*, vol. 347, no. 4, Apr. 1996.

[28] T. Yokonuma, *Tensor Spaces and Exterior Algebra*, AMS, Providence, 1991.

FIGURE CAPTIONS:

Figure 1: Nonlinear signal transformation (NST) $F_n : \mathbf{x} \mapsto \boldsymbol{\theta}$. The front end processing (projection onto the basis $\{\mathbf{b}_j\}$) is linear; the back end processing (by η from (1) or (2)) is nonlinear.

Figure 2: Filter bank realization of the summing node NST. By combining the basis vectors as in (21), we can decompose an arbitrary summing node NST into a parallel cascade of a redundant set of linear filters $\{\mathbf{f}_k\}_{k=1}^N$, each followed by a simple monomial nonlinearity $(\cdot)^n$.

Figure 3: Comparison of different bases $\{\mathbf{b}_j\}$ for nonlinear signal processing. The choice of basis employed in the linear front end of the NST of Figure 1 determines in which domain we represent signal interactions. Consider a second-order NST, which generates squares β_j^2 and cross-products $\beta_i\beta_j$ of the signal coefficients. We illustrate two basis elements $\mathbf{b}_i$ and $\mathbf{b}_j$ from three different bases, in both the time domain and the frequency (squared magnitude) domain. In the delta basis, each $\mathbf{b}_j$ is a unit pulse, so β_j is simply a sample of the signal. The corresponding NST represents coupling between different time lags of the signal. In the Fourier basis, each $\mathbf{b}_j$ is a sinusoid, so β_j is a Fourier coefficient of the signal. The corresponding NST represents intermodulations between different frequencies. In the wavelet basis, each $\mathbf{b}_j$ is localized in both time and frequency simultaneously, so β_j measures the time-frequency content of the signal. The corresponding NST represents coupling between different localized wavelet atoms.

Figure 4: Graphical depiction of the elements of the two-dimensional multiresolution and tensor wavelet bases. The aspect ratios of the tiles correspond roughly to the size of the "regions of support" of the basis elements. (a) The multiresolution wavelet basis consists only of products of pairs of wavelets and scaling functions from the same scale; hence all tiles have the same aspect ratio. (b) The tensor wavelet basis consists of products of pairs of wavelets and scaling functions from all possible scales; hence many different aspect ratios result. (Strictly speaking, these mosaics illustrate the organization of the *coefficients* obtained upon expanding onto these bases. Nevertheless, there is a direct correspondence between the size of a coefficient tile and the size of the region of support of the associated basis elements: Basis functions broad in one direction result in fewer expansion coefficients in that direction and hence a narrower tile.)

Figure 5: Data from an acoustic emission experiment. (a) Emission from a typical trial. (b) Overlay of data from twenty trials.

Figure 6: Histograms of the second-order ($n = 2$) correlations of the acoustic emission signals from Figure 5 in three different basis representations. (a) Histogram of time-domain correlation magnitude, entropy $H = 6.44$. (b) Histogram of frequency-domain correlation magnitude, entropy $H = 2.39$. (c) Histogram of wavelet-domain correlation, entropy $H = 1.82$. To better illustrate both the peakiness and rapid

decay of the wavelet-domain correlation, we plot only the first few bins of the histograms on a logarithmic vertical scale.

Figure 7: Volterra filter realization using a summing node NST.

Figure 8: Nonlinear system identification using tensor bases. Estimates of the quadratic Volterra kernel of an audio loudspeaker obtained using thresholding in a tensor basis expansion. (a) True kernel h [12]. (b) Raw estimate $\widehat{h}$ obtained using (45), MSE=0.20. (c) Estimate $\widetilde{h}$ obtained through a truncated Daubechies-8 wavelet expansion of $\widehat{h}$, MSE=0.15. (d) Estimate $\widetilde{h}$ obtained through truncated Fourier expansion of $\widehat{h}$, MSE=0.40. The wavelet estimator provides a more faithful estimate than the Fourier estimator, which oversmooths.

BIOGRAPHIES

Richard G. Baraniuk (M–93, SM–98) received the BSc degree in 1987 from the University of Manitoba, the MSc degree in 1988 from the University of Wisconsin-Madison, and the PhD degree in 1992 from the University of Illinois at Urbana-Champaign, all in Electrical Engineering. In 1986, he was a research engineer with Omron Tateisi Electronics in Kyoto, Japan. After spending 1992–1993 with the Signal Processing Laboratory of Ecole Normale Supérieure, in Lyon, France, he joined Rice University, where he is currently an Associate Professor of Electrical and Computer Engineering. He spent Autumn 1998 at the Isaac Newton Institute of Cambridge University, England as the Rosenbaum Fellow. His research interests lie in the area of signal and image processing and include wavelets and time-frequency analysis.

Dr. Baraniuk received a NATO postdoctoral fellowship from NSERC in 1992, a National Young Investigator award from the National Science Foundation in 1994, and a Young Investigator award from the Office of Naval Research in 1995.