Wavelets and Random Processes: Optimal Matching in the Bhattacharyya Sense

Nirmal Keshava and José M. F. Moura *
Carnegie Mellon University
Department of Electrical and Computer Engineering
5000 Forbes Avenue, Pittsburgh, PA 15213-3890
keshava@ece.cmu.edu

Abstract

The use of wavelet packet bases to represent deterministic signals has led to optimal representations that minimize a desired cost function. In this paper, we present an algorithm that uses wavelet packet bases to construct a random process that is matched to an arbitrary correlation matrix. An analytical expression for the Bhattacharyya coefficient gauges the similarity between the two processes and leads to two iterative algorithms that are used jointly to find the desired quantities. Eigendecompositions for the two processes reduce the problem to finding the best wavelet-based unitary matrix and a set of eigenvalues. Examples illustrates the technique.

1. Introduction

In target detection and classification, wavelet theory has proven to be a capable tool for capturing localized activity in time/space and frequency. As an alternative to the Fourier Transform for deterministic signals, it provides a multiresolution decomposition that can methodically reveal multiple levels of detail in a signal.

The idea of matching wavelet-based, multiresolution representations to deterministic signals is extended in [2, 7] to wavelet packet bases. This generalization of the traditional wavelet filter bank structure permits the representation of a signal by one of many possible bases, each of which is constructed by a unique ensemble of scalings and translations of a wavelet/scaling filter pair. By employing a suitable cost criteria, the best basis from the entire family of bases is chosen to represent a single deterministic signal.

For the analysis of stochastic processes, however, no comprehensive framework exists for optimally representing random signals using wavelets. Several efforts [8, 1, 3, 9] have been made to incorporate the statistics of stochastic processes into the mathematical framework of wavelets.

In this paper, an algorithm is presented which constructs a wavelet-based correlation matrix that approximates an arbitrary correlation matrix by using the Bhattacharyya coefficient as a cost criteria. Because the Bhattacharyya coefficient is actually a bound on the probability of error when detecting the presence of either of the two processes, it is a useful metric for measuring the similarity of the wavelet-based process and the original process.

The wavelet-based process is constructed by inserting the eigendecomposition of its correlation matrix, as well as that of the original process, into the analytic expression for the Bhattacharyya coefficient. As a result, the optimization reduces to the selection of the two defining quantities of the wavelet-based process: its unitary matrix of eigenvectors and the associated eigenvalues. The unitary matrix may be any of the possible wavelet packet bases in the tree spawned by the wavelet/scaling filter pair. Each basis in the tree has an optimal set of eigenvalues that maximizes the similarity between the original process and the wavelet-based process.

2. Preliminaries

Let Σ be the $N \times N$ correlation matrix for a real-valued, zero-mean, Gaussian stochastic process. The power of the process, or the trace of Σ, is P. Let $\hat{\Sigma}$ be the correlation matrix for the wavelet-based process whose entries are to be determined. Let Σ and $\hat{\Sigma}$ have the following eigendecompositions:

$$\Sigma = U \cdot S \cdot U^H \tag{1}$$

$$\hat{\Sigma} = \hat{U} \cdot \hat{S} \cdot \hat{U}^H. \tag{2}$$

U is a unitary matrix containing the eigenvectors of Σ. S is a diagonal matrix containing the corresponding eigenvalues of Σ. $\hat{U}$ and $\hat{S}$ are defined similarly.

The Bhattacharyya coefficient, ρ, for Σ and $\hat{\Sigma}$ when both processes are real, Gaussian, and equally likely, is [4]:

$$\rho(\Sigma, \hat{\Sigma}) = e^{-\mu} \tag{3}$$

*This work was partially supported by DARPA through ONR grant N00014-91-J183 and AFOSR grant F49620-96-O436.

where

$$\mu = \frac{1}{2} \ln \frac{|\ \frac{\Sigma+\hat{\Sigma}}{2}\ |}{\sqrt{|\ \Sigma\ ||\ \hat{\Sigma}\ |}}. \quad (4)$$

From (3) and (4) it follows that: $0 \leq \rho \leq 1$. The larger the value of ρ, the greater the similarity between Σ and $\hat{\Sigma}$.

Substituting the eigendecompositions in (1) and (2), ρ can be shown to become:

$$\rho(\Sigma, \hat{\Sigma}) = 2^{-N/2}\ |\ S\ |^{\frac{1}{4}} |\ \hat{S}\ |^{\frac{1}{4}} |\ \hat{S} + \hat{U}^H \Sigma \hat{U}\ |^{-\frac{1}{2}}\ . \quad (5)$$

Since the first two terms in (5) contain neither $\hat{U}$ nor $\hat{S}$, maximizing ρ is equivalent to maximizing the product of the last two determinants:

$$W(\Sigma, \hat{\Sigma}) = |\ \hat{S}\ |^{\frac{1}{4}} |\ \hat{S} + \hat{U}^H \Sigma \hat{U}\ |^{-\frac{1}{2}}\ . \quad (6)$$

Let $\{c_k, d_k\}$ be any orthonormal, compactly supported, dyadic, scaling/wavelet filter pair that spawns a wavelet packet tree structure. Any wavelet packet basis that arises from $\{c_k, d_k\}$ may be expressed as an orthonormal basis for $\Re^N$ where N is constrained to be a power of 2. Let $\hat{U}$ contain the N basis vectors, where each vector in $\hat{U}$ resides on a node in the wavelet packet tree.

3. Approach

The full algorithm for matching a wavelet-based correlation matrix to an arbitrary correlation matrix starts with a scaling/wavelet filter pair, $\{c_k, d_k\}$, for any orthonormal, compactly supported, dyadic, wavelet. Successive upsampling and filtering of the filter coefficients leads to vector sequences that are orthonormal on the same scale. Vector sequences that exceed N, the matrix dimension, are wrapped circularly. In [5] the theory is presented for adapting the discrete wavelet transform (DWT) from the filtering domain to the matrix domain. These results have been extended here to unitary matrices and unitary wavelet packet bases and supply an alternative perspective to traditional forms of wavelet-based analysis.

The maximization of $\rho(\Sigma, \hat{\Sigma})$ is facilitated by considering the squared reciprocal of the expression in (6):

$$\begin{aligned} \Omega(\Sigma, \hat{\Sigma}) &= \left(\frac{1}{W(\Sigma, \hat{\Sigma})}\right)^2 \\ &= |\ \hat{S}\ |^{-\frac{1}{2}} |\ \hat{S} + \hat{U}^H \Sigma \hat{U}\ |\ . \end{aligned} \quad (7)$$

The maximization of $W(\Sigma, \hat{\Sigma})$ is now rephrased as the equivalent task of minimizing $\Omega(\Sigma, \hat{\Sigma})$ with respect to the same quantities as before, $\hat{U}$ and $\hat{S}$. The following subsections separately address the minimization with respect to these two entities before the concepts are combined into one complete algorithm.

3.1. Optimal Eigenvalues: Fixed-Point Algorithm

Consider the constrained problem of minimizing $\Omega(\Sigma, \hat{\Sigma})$ when a unitary basis matrix, $\hat{U}$, has been fixed. What remains is to find the matrix, $\hat{S}$ that contains the eigenvalues, $g_1, \ldots, g_N$, for the wavelet-based process, $\hat{\Sigma}$ that maximize $\Omega(\Sigma, \hat{\Sigma})$, subject to the constraint

$$\sum_{i=1}^{N} g_i = P \quad (8)$$

i.e., the power in the wavelet-based process equals the power in the original process.

$\Omega(\Sigma, \hat{\Sigma})$ can be expressed as:

$$\Omega(\Sigma, \hat{\Sigma}) = \left| \begin{bmatrix} g_1^{-\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & g_N^{-\frac{1}{2}} \end{bmatrix} \right| \cdot \left| \begin{bmatrix} g_1 + a_{11} & \ddots & \ddots & a_{1N} \\ \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots \\ a_{N1} & \ddots & \ddots & g_N + a_{NN} \end{bmatrix} \right| \quad (9)$$

where $a_{ij}, i, j = 1, \ldots, N$, are the elements of the symmetric, positive definite matrix, A,

$$A = \hat{U}^H \Sigma \hat{U}. \quad (10)$$

For notational convenience, allow the first determinant in (9) to be known as G, where

$$G = \prod_{i=1}^{N} g_i^{-\frac{1}{2}}, \quad (11)$$

and the second determinant in (9) to be called V. Then,

$$\Omega(\Sigma, \hat{\Sigma}) = GV. \quad (12)$$

The i-th principal minor of V, which is the determinant of the matrix formed by deleting the i-th row and i-th column, is then denoted as V_{ii}.

To minimize $\Omega(\Sigma, \hat{\Sigma})$ with respect to the eigenvalues in $\hat{S}$, derivatives are first taken with respect to g_i:

$$\frac{\partial \Omega}{\partial g_i} = \frac{\partial G}{\partial g_i} V + G \frac{\partial V}{\partial g_i} \quad (13)$$

$$= -\frac{1}{2} \frac{1}{g_i} \frac{1}{\sqrt{G}} V + \frac{1}{\sqrt{G}} V_{11}. \quad (14)$$

Setting $\frac{\partial \Omega}{\partial g_i}$ equal to 0 in (14) and solving for g_i leads to the following expression:

$$g_i = \frac{1}{2}\frac{V}{V_{ii}}. \tag{15}$$

Since V, and hence V_{ii}, are a function of $g_1, \ldots, g_N$, this expression results in a fixed-point algorithm for g_i. Using a superscript to denote iterations, the iterative expression that finds the optimal values for $g_1, \ldots, g_N$ is expressed as

$$\begin{bmatrix} g_1^j \\ \vdots \\ \vdots \\ g_N^j \end{bmatrix} = k^{j-1} V^{j-1} \begin{bmatrix} \frac{1}{V_{11}^{j-1}} \\ \vdots \\ \vdots \\ \frac{1}{V_{NN}^{j-1}} \end{bmatrix} \tag{16}$$

where a constant, k^{j-1}, has been inserted as a normalizing factor to preserve the total power, P, in the original process. The requirement that $g_1, \ldots, g_N$ be strictly positive is met, since both V and $V_{ii}, i = 1, \ldots, N$, are guaranteed to be positive. It has been found that an appropriate choice for the starting values, $g_1^0, \ldots, g_N^0$, is the main diagonal of A.

3.2. Optimal Eigenvectors: Basis Migration Algorithm

Alternately, minimizing $\Omega(\Sigma, \hat{\Sigma})$ when the eigenvalue matrix of the wavelet-based process, $\hat{S}$, is fixed, requires a technique to find the only remaining variable quantity, the unitary basis matrix, $\hat{U}$. The challenge of finding the best basis from a wavelet packet tree structure is a significant one. In [2, 7, 6] the problem is tackled when the cost function is *additive;* the result is an unconstrained optimization in which branches of the wavelet packet tree can be pruned independently of other branches until the overall cost is minimized.

The Bhattacharyya coefficient, unfortunately, is not an additive cost function, i.e., the branches of the wavelet packet tree cannot be pruned independently and still lead to an optimum solution. A naive approach to optimizing a non-additive cost function would be to evaluate it for every possible wavelet packet basis. A more efficient alternative is to pick an initial basis and allow its vectors to "migrate" up and down the branches of the tree until it arrives at a new basis which maximizes the Bhattacharyya coefficient.

Consider the equality in (7). With $\hat{S}$ fixed, define

$$\Omega_b(\Sigma, \hat{\Sigma}) = |\, \hat{S} + \hat{U}^H \Sigma \hat{U} \,| \tag{17}$$

$$= |\, U^H \hat{U} \hat{S} \hat{U}^H U + S \,|\,. \tag{18}$$

Let a superscript, j, indicate successive values for $\hat{U}$. Without sacrificing generality, assume $\hat{U}^0$, the initial choice for the unitary basis, is populated by the vectors at the bottom

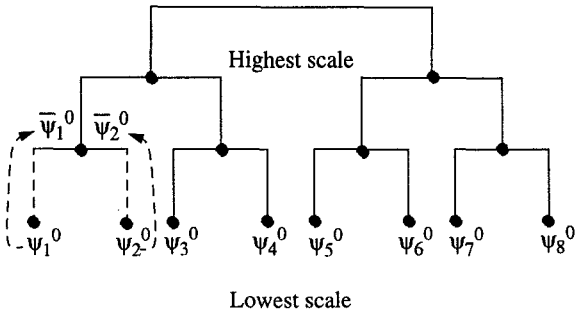

Figure 1. The wavelet packet tree

scale of the wavelet packet tree. Figure 1 illustrates this concept. These vectors provide an STFT-like decomposition of a signal. $\hat{U}^0$ is:

$$\hat{U}^0 = [\ \psi_1^0 \quad \psi_2^0 \,|\, \psi_3^0 \quad \psi_4^0 \,|\, \cdots \,|\, \psi_{N-1}^0 \quad \psi_N^0\ ]. \tag{19}$$

The vertical bars demarcate leaves on the tree that have been spawned from the same branch. Vectors from the same branch may either migrate together up to the node at the next highest scale, or be filtered further by $\{c_k, d_k\}$ to yield basis vectors on the next lower scale. In the case of the basis in (19), since the vectors are already on the lowest scale, they may only migrate upwards.

Each pair of vectors in $\hat{U}^0$ (e.g., $[\psi_1^0 \psi_2^0]$) may be replaced by an alternative pair of vectors (e.g., $[\overline{\psi}_1^0 \overline{\psi}_2^0]$) residing on the next higher scale, while maintaining the unitarity of the matrix. Since there are $N/2$ groups in (19) there are $2^{N/2} - 1$ new bases to which $\hat{U}^0$ may migrate.

To illustrate the algorithm, consider the impact on $\Omega_b^0(\Sigma, \hat{\Sigma})$ when $[\psi_1^0 \ \psi_2^0]$ is replaced by $[\overline{\psi}_1^0 \ \overline{\psi}_2^0]$. Rewriting (18),

$$\Omega_b^0(\Sigma, \hat{\Sigma}) = |\, U^H \hat{U}^0 \hat{S} \hat{U}^{0H} U + S \,| \tag{20}$$

$$= |\, U^H \sum_{i=1}^{i=2} g_i \psi_i^0 \psi_i^{0H} U \tag{21}$$

$$+ U^H \sum_{i=3}^{i=N} g_i \psi_i^0 \psi_i^{0H} U + S \,|\,. \tag{22}$$

If $[\psi_1^0 \ \psi_2^0]$ is replaced by $[\overline{\psi}_1^0 \ \overline{\psi}_2^0]$, the net change in contribution by the first summation in (22) is

$$\delta_{12}^0 = U^H \Big(\sum_{i=1}^{i=2} \overline{g}_i \overline{\psi}_i^0 \overline{\psi}_i^{0H} - \sum_{i=1}^{i=2} g_i \psi_i^0 \psi_i^{0H} \Big) U, \tag{23}$$

where $\overline{g}_1$ and $\overline{g}_2$ is a redistribution of the power in g_1 and g_2 when moving from one scale to another. This reallocation of power among vectors on a common branch is not necessary but is performed to maximize the similarity between

Σ and $\hat{\Sigma}$ and, ultimately, enhance the speed of the complete algorithm. A significant point about the reallocation is that the total power *within* a branch is reorganized, yet completely conserved. The algorithm described in Section 3.1 is responsible for optimally redistributing power *between all vectors* while maintaining the overall power constraint. The values for $\overline{g}_i, i = 1, 2$, are:

$$\overline{g}_i = \frac{\overline{\psi}_i^{0H} K \overline{\psi}_i^0}{trace([\overline{\psi}_1^0 \overline{\psi}_2^0]^H K [\overline{\psi}_1^0 \overline{\psi}_2^0])}(g_1 + g_2). \quad (24)$$

To decide whether the difference in migrating from $[\psi_1^0 \psi_2^0]$ to $[\overline{\psi}_1^0 \overline{\psi}_2^0]$ lowers $\Omega_b^0(\Sigma, \hat{\Sigma})$, and hence maximizes the Bhattacharyya coefficient, the $N \times N$ matrix, δ_{12}^0, is multiplied, element-wise, with the gradient of the original system in (21), and summed. The gradient of the original system, ∇^0, is the partial derivative of $\Omega_b^0(\Sigma, \hat{\Sigma})$ with respect to each matrix element in the original system:

$$\begin{aligned} \nabla^0 &= \text{Adj}(U^H \hat{U}^0 \hat{S} \hat{U}^{0H} U + S) && (25) \\ &= |\hat{\Sigma} + \Sigma| (U^H \hat{U}^0 \hat{S} \hat{U}^{0H} U + S)^{-1} && (26) \\ &= c^0 (U^H \hat{U}^0 \hat{S} \hat{U}^{0H} U + S)^{-1} && (27) \end{aligned}$$

where c^0 is a constant that may be ignored. The impact caused by migrating from $[\psi_1^0 \psi_2^0]$ to $[\overline{\psi}_1^0 \overline{\psi}_2^0]$ on $\Omega_b^0(\Sigma, \hat{\Sigma})$ is therefore given by the innovation value, ϵ_{12}^0:

$$\epsilon_{12}^0 = \sum_{i,j=1}^{N} \delta_{12}^0(i,j) \nabla^0(i,j). \quad (28)$$

If $\epsilon_{12}^0 > 0$, then the migration increases $\Omega_b^0(\Sigma, \hat{\Sigma})$ and subsequently decreases the similarity between the two proceses. If $\epsilon_{12}^0 < 0$ then the new basis will result in a wavelet-based process that is closer to the original process, Σ, and the migration is justified.

This procedure is repeated for each group of vectors in the basis. Migration of one group of vectors to another scale only occurs if its impact is to decrease $\Omega_b(\Sigma, \hat{\Sigma})$. When all the basis vector groups have been evaluated for migration, and a new basis, $\hat{U}^1$, has been found, the procedure is repeated, initiated by calculating ∇^1, using $\hat{U}^1$, and $\overline{S}$, the diagonal matrix of reallocated eigenvalues, if the eigenvalues have been reallocated, or S, if they have not.

4. The Complete Matching Algorithm

Starting with a scaling/wavelet filter pair, an initial wavelet packet basis, $\hat{U}^0$, is chosen. Initial results indicate that selecting $\hat{U}^0$ to be the highest scale of wavelet packet basis vectors provides an adequate starting point for a number of classes of matrices. $\hat{S}^0$, the initial eigenvalue matrix, utilizes the diagonal of A^0:

$$A^0 = \hat{U}^{0H} \Sigma \hat{U}^0. \quad (29)$$

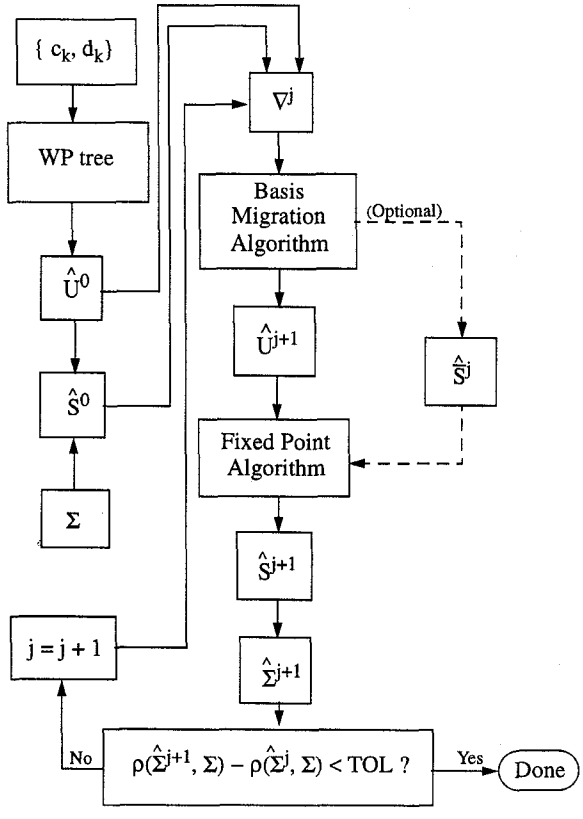

Figure 2. The complete matching algorithm

$\hat{\Sigma}^0$, the first approximation to Σ, may be constructed:

$$\hat{\Sigma}^0 = \hat{U}^0 \hat{S}^0 \hat{U}^{0H}. \quad (30)$$

$\rho^0(\Sigma, \hat{\Sigma}^0)$ is calculated and recorded.

$\hat{U}^1$ is obtained by migrating $\hat{U}^0$ using the basis migration algorithm discussed in Section 3.2 when $\hat{S}^0$ is employed as the set of eigenvalues. Successive values of $\hat{U}$ will reflect incremental changes in vectors when compared to the preceding unitary basis. The eigenvalues in $\hat{S}^1$ are determined by iterating the fixed point algorithm described in Section 3.1 when $\hat{U}^1$ is the fixed unitary matrix. If the eigenvalues in $\hat{S}^0$ were reallocated to obtain $\hat{U}^1$, then $\overline{\hat{S}^0}$ is the starting point for determining $\hat{S}^1$. $\hat{\Sigma}^1$ is created in the same manner as $\hat{\Sigma}^0$. $\rho^1(\Sigma, \hat{\Sigma}^1)$ is then calculated, and the entire procedure iterates until successive values of $\rho^j(\Sigma, \hat{\Sigma}^j)$ are equal within a prescribed tolerance. Figure 2 illustrates the complete matching algorithm.

5. Examples

Two examples demonstrate how the complete algorithm arrives at an optimal basis and set of eigenvalues without

Iteration	Scales of basis vectors	$\rho(\Sigma, \hat{\Sigma}^i)$
1	{3,3,3,3,3,3,3,3}	0.2791
2	{2,2,2,2,2,2,2,2}	0.4562
3	{1,1,1,1,3,3,3,3}	0.7643
4	{1,1,1,1,2,2,2,2}	0.7932
5	{1,1,1,1,3,3,3,3}	0.8158

Table 1. Results for Markov Process

exhaustively testing all possibilities. The first correlation matrix to be matched is an 8×8 first-order Markov process whose eigenvectors are known to be non-harmonic sinusoids. The second correlation matrix is an 8×8 bandlimited process whose eigenvectors are from the family of angular prolate spheroidal functions. The wavelet/scaling filter pair used corresponds to the Daubechies-2 wavelet. The wavelet packet tree has three levels, level three corresponding to the highest scale (best temporal resolution and worst frequency resolution), and level one corresponding to the lowest scale (worst temporal resolution and best frequency resolution). In both cases, the initial basis, $\hat{U}^0$, are the vectors residing on the highest scale. The results were confirmed by an exhaustive search over all possible bases.

5.1. First-Order Markov Process

Σ is a first-order Markov process with correlation coefficient, $\rho = 0.95$. The entries of $\Sigma(i,j)$ are given by:

$$\Sigma(i,j) = \rho^{|i-j|}. \tag{31}$$

Table 1 summarizes the results of the iteration. The algorithm arrived at a basis whose vectors come from both the highest scale, and the lowest scale. It is worth noting that although the algorithm started with basis vectors on the highest scale, and then immediately migrated to the next lower scale, half of the vectors in the final basis returned to the highest scale. This is attributed to the effect of the fixed-point eigenvalue algorithm in Section 3.1 which dynamically "steers" ∇, and returns half of the basis vectors in $\hat{U}$ to the highest scale to arrive at a higher value of $\rho(\Sigma, \hat{\Sigma})$.

5.2. Bandlimited Process

Σ is a bandlimited process whose bandwidth in the normalized frequency domain is $\omega = 0.41$. The entries of $\Sigma(i,j)$ are given by:

$$\Sigma(i,j) = \frac{\sin(2\pi\omega(i-j))}{\pi(i-j)}. \tag{32}$$

Table 2 summarizes the results of the iteration. The phenomenon of upwards and downwards vector migration discussed in Section 5.1 was also observed here.

Iteration	Scales of basis vectors	$\rho(\Sigma, \hat{\Sigma}^i)$
1	{3,3,3,3,3,3,3,3}	0.7658
2	{2,2,2,2,2,2,2,2}	0.8328
3	{1,1,1,1,1,1,1,1}	0.9450
4	{1,1,1,1,1,1,2,2}	0.9467

Table 2. Results for Bandlimited Process

6. Conclusion

We have demonstrated an algorithm for constructing a wavelet-based correlation matrix that approximates the correlation matrix for an arbitrary random process in the Bhattacharyya sense. The algorithm searches a wavelet packet tree for the unitary basis of eigenvectors that maximizes the similarity between the two processes and also finds the associated eigenvalues.

Future work on this algorithm will concentrate on demonstrating convergence and investigating its dependence on starting values. These issues will become pivotal when the algorithm is tested on several varieties of processes, in particular, high-pass and low-pass processes.

References

[1] M. Bassevile, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky. Modeling and estimation of multiresolution stochastic processes. *IEEE Trans. Info. Theory*, 38(2):766–784, Mar. 1992.

[2] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Info. Theory*, 38(2):713–718, Mar. 1992.

[3] R. W. Dijkerman and R. R. Mazumdar. Wavelet representations of stochastic processes and multiresolution stochastic models. *IEEE Trans. in ASSP*, 42(7):1640–1652, July 1994.

[4] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, Inc., Philadelphia, Pennsylvania, 1990.

[5] N. Getz. A fast discrete periodic wavelet transform. Technical Report UCB/ERL M92/138, University of California, Berkeley, 1992.

[6] C. Herley, J. Kovaeević, K. Ramchandran, and M. Vetterli. Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms. *IEEE Trans. Sig. Proc.*, 41(12):3341–3359, Dec. 1993.

[7] K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distortion sense. *IEEE Trans. on Image Proc.*, 2(2):160–175, Apr. 1993.

[8] G. W. Wornell. A Karhunen-Loéve-like expansion for 1/f processes via wavelets. *IEEE Trans. Info. Theory*, 36(4):859–861, July 1990.

[9] J. Zhang and G. Walter. A wavelet-based KL-like expansion for wide-sense stationary random processes. *IEEE Trans. in ASSP*, 42(7):1737–1745, July 1994.