# Enhanced Pseudo Power Signatures For Nonstationary Signal Classification: The Projector Approach

Vidya Venkatachalam[1]
Computational Mathematics Laboratory
Rice University

Jorge L. Aravena
Dept. of Electrical & Computer Engr.
Louisiana State University

EDICS : 2-TIFR

## Abstract

*The classification of nonstationary signals of unknown duration is of great importance in areas like oil exploration, moving target detection, and pattern recognition. In an earlier work, we provided a solution to this problem, based on the wavelet transform, by defining representations called* **pseudo power signatures** *for signal classes which were independent of signal length, location and magnitude, and proposed a simple approach using the Singular Value Decomposition to generate these signatures. This paper offers a new approach resulting in more discriminating signatures. The enhanced signatures are obtained by solving a nonlinear minimization problem involving an inverse projection. The problem formulation, solution procedure, and computational algorithm are presented in this work. An analysis of the projection signatures, and their efficacy in separating highly correlated signal classes are demonstrated through simulation examples.*

[1]POC author : Dr. Vidya Venkatachalam (supported by DARPA/AFOSR and Northrop-Grumman)
Computational Mathematics Laboratory
Dept. of Mathematics - MS 136
6100 Main St.
Rice University
Houston, TX 77005
Tel : (713) 527 8101-2372
Fax : (713) 285 5231
Email : vidyav@rice.edu

# 1 Introduction and previous work

The classification of nonstationary signals of unknown duration is of great importance in areas like oil exploration, moving target detection, and pattern recognition. Consider the following representative classification problem :

*Signals are obtained by propagating electromagnetic waves through several layers of different classes of materials. The goal is the determination of the various classes present and the thickness of each layer. We denote the presence of a particular class as the occurrence of an event.*

The problem was described as a generic classification problem in a previous work ([1]), and was solved by introducing the concept of *pseudo power signatures.* For signals in each class, these signatures capture information at different scales, independent of the signal duration. Essentially, the signatures characterize the scale power distribution ([2]) in a manner independent of time. The signatures are then invariant to time shifts, and *pieces of signals* are all characterized by the *same signature.* This is an extremely important characteristic, and one that is not generally present in current classification schemes ([3]). For $x \in L^2(\Re)$ with $CWT$, $c_\psi^x \in H = L^2(\Re^2, C_\psi^{-1} \frac{da db}{a^2})$, where $\psi$ is an admissible wavelet, we approximated $c_\psi^x(a,b)$ by a separable element of the form[2]

$$c_\psi^x(a,b) \approx s_\psi^x(a) r_\psi^x(b)$$

[2]We have shown ([4]) that there do not exist any admissible wavelets that admit a separable $CWT$ function. Thus, we can only approximate a $CWT$ function by a separable form.

where $s_\psi^x \in S = L^2(\Re, C_\psi^{-1}\frac{da}{a^2})$, and $r_\psi^x \in R = L^2(\Re, db)$. The normalized function $s_\psi^x$ corresponds to the pseudo power signature of $x$. In [4], we enumerated a simple approach to determine this signature by performing a Singular Value Decomposition ($SVD$) of the Continuous Wavelet Transform ($CWT$), and extracting the principal component ([5]). For discrete signals, we reduced the problem of determination of the pseudo power signatures to a standard matrix $SVD$ problem, and readily obtained the discrete approximation to the signatures from the principal component of the discretized $CWT$ matrix. The results shown in [4] indicate that the signatures obtained from the principal component of the discretized $CWT$ matrix are easy to compute, and are insensitive to noise, but are limited by a lack of fine discriminating capability. Consequently, they are not well suited for classification problems. The $SVD$ matrix technique had two obvious shortcomings: (1) It implicitly assumed that the $CWT$ was piecewise constant. (2) It made the more seriously erroneous assumption that *the orthogonality of the vectors* $u_i, u_j \in l^2(\mathcal{Z})$, *and* $v_i, v_j \in l^2(\mathcal{Z})$, $i \neq j$, *implied that the elements* $s_i r_i, s_j r_j \in H$ *defined as* $s_i r_i(a,b) = \sum_{l,n} 2^{2l} u_i(l)\overline{v_i(n)} q_{l,n}(a,b)$ *and* $s_j r_j(a,b) = \sum_{l,n} 2^{2l} u_j(l)\overline{v_j(n)} q_{l,n} n(a,b)$, *are orthogonal in* $H$ [3]. We can readily see that such an assumption is not valid in the space $H$ with its weighted inner product ([6]). Hence, given $c_\psi^x \in M$, though the principal component of $c_\psi^x$, denoted by $PC[c_\psi^x]$, is such that $\left\langle c_\psi^x, c_\psi^x - PC[c_\psi^x] \right\rangle_H = 0$, this property no longer holds true when we consider the discretization used in [4], to determine $PC[c_\psi^x]$,

[3]

$$q_{l,n}(a,b) = \begin{cases} 1, & 2^l \leq a < 2^{l+1}, \quad n \leq b < n+1 \\ 0, & elsewhere \end{cases}$$

resulting in reduced signature quality.

A much more significant and subtle limitation lies in the fact that assuring a good approximation to the $CWT$ does not necessarily assure a good approximation to the signal that is being classified. Since the separable approximation cannot be a $CWT$, associating a time signal to it is not obvious. The $CWT$ permits the definition of an isometry between $L^2(\Re)$ and the subspace of wavelet transforms $M \subset H$. Hence, what one needs to consider is the *projection of the separable terms onto the subspace* $M$. The association of a time signal to the separable term in $H$ is then logically defined as the element in $L^2(\Re)$ associated with the projection in $M$ of the separable term. The situation is best described in Figure 1. In the figure, $c_\psi^x \in M$ denotes a $CWT$. The vectors $s_1 \otimes r_1, s_2 \otimes r_2$ represent the best separable approximations obtained using different notions of orthogonality (one obtained from the $SVD$ of the discretized $CWT$ matrix, and the other directly from $PC[c_\psi^x]$). The vector $s_3 \otimes r_3$ denotes the separable term that provides the best orthogonal projection onto the subspace of wavelet transforms. Our first result will show that this last term provides the best separable representation of the signal $x(t)$. The normalized scale component, $s \in S$, will then be used to define a pseudo power signature for $x(t)$. Since this signature is obtained as a result of a projection, it is referred to as a projection signature. The following sections formulate and solve the problem of directly determining the pseudo power signatures using a suitably defined projection operator.

## 2 Orthogonal projections

The first step in the determination of the projection signature is the definition of a suitable orthogonal projection operator $\mathcal{K} : H \to M$.

**Theorem 2.1** *There exists an orthogonal projection operator $\mathcal{K} : H \to M$ defined as follows :*

$$\mathcal{K}\left[c\right](a,b) = C_\psi^{-1} \int_\alpha \int_\beta \overline{c_\psi^{\psi_{ab}}(\alpha,\beta)}c(\alpha,\beta)\frac{d\beta d\alpha}{\alpha^2}, \ \ \forall\, c \in H$$

The proof is given in [7].

The next result establishes a fundamental connection between this projector and the continuous wavelet transform.

**Theorem 2.2** *Consider the Continuous Wavelet Transform, $\Gamma : L^2(\Re) \to H$, defined by $[\Gamma x](a,b) = \langle x, \psi_{a,b}\rangle$ ; $x \in L^2(\Re)$. This operator is known to have the closed subspace $M \subset H$ as its range space. The adjoint operator $\Gamma^* : H \to L^2(\Re)$ satisfies the following conditions :*

*1) $\Gamma^*\Gamma : L^2(\Re) \to L^2(\Re)$ is the identity operator*

*2) $\Gamma\Gamma^* : H \to H$ is the orthogonal projector, $\mathcal{K}$ introduced in Theorem 2.1.*

The proof of this theorem is also included in [7]. A result which readily follows from Theorems 2.1 and 2.2 is given by :

**Corollary 2.1** *To every $c \in H$, there corresponds one and only one $\hat{x} \in L^2(\Re)$ such that $\Gamma[\hat{x}] = \mathcal{K}\left[c\right] \in M$, i.e.*

$$c_\psi^{\hat{x}}(a,b) = \mathcal{K}\left[c\right](a,b)$$

For any given $x \in L^2(\Re)$, let $c_\psi^x \in M$ denote its $CWT$ with respect to an admissible $\psi \in L^2(\Re)$. Consider the element $sr \in H$ ($sr$ is unitarily equivalent to $s \otimes r$). Let $\hat{c} = \mathcal{K}[sr] \in M$, and $\hat{x} \in L^2(\Re)$ the element associated with $sr$ by Corollary 2.1. It intuitively follows that if we determine $sr \in H$ such that $\| c_\psi^x - \hat{c} \|_M$ is minimized, then we effectively minimize $\| x - \hat{x} \|_2$. Hence, we can expect that $\hat{c}$, and consequently, $sr$, will better characterize the intrinsic properties of $x$. However, we have no formal proof that the orthogonal projection operator $\mathcal{K}$, when restricted to the set of separable elements in $H$, is one-one. Thus, in order to ensure the determination of a *unique* projection signature, a regularizing term $\alpha \|sr\|$ is added to the minimization problem. For analysis purposes, $\alpha = 1$. The minimization problem can then be represented as follows :

**For a given $c_\psi^x \in M$, find the decomposition $s_\psi^x r_\psi^x \in H$ that minimizes the index**

$$J(s_\psi^x, r_\psi^x) = \left\{ \| c_\psi^x - \mathcal{K}[s_\psi^x r_\psi^x] \|_M^2 + \| s_\psi^x r_\psi^x \|_H^2 \right\}$$

**where $\mathcal{K}$ is the orthogonal projection operator defined earlier.**

This is an infinite dimensional nonlinear minimization problem, and requires the solution of the inverse projection problem. The problem formulation and solution procedure for the infinite dimensional case is discussed in [7]. However, for a practical application, the problem needs to be reduced to a finite dimensional one. The problem formulation and the corresponding solution procedure for the finite dimensional case are discussed in the following sections.

# 3 Problem formulation

The first step towards developing a finite dimensional representation for the infinite dimensional minimization problem is to determine a suitable discretization for the elements $c^x_\psi, \mathcal{K}[s^x_\psi r^x_\psi] \in M$, $s^x_\psi \in S$, and $r^x_\psi \in R$. As discussed in [4], given $c^x_\psi \in M$, one can obtain a discretized equivalent using the concept of frames and frame operators ([8]), and a wavelet $\psi \in L^2(\Re)$ of compact support that arises from a multiresolution analysis ($MRA$). The set of discretized coefficients $\{c^x_\psi(2^l, n)\}_{l,n}$ can then be determined using the Shensa algorithm ([9]), and we can approximate $\{c^x_\psi(2^l, n)\}_{l,n}$ using finitely many coefficients, and thus obtain a finite dimensional discretized $CWT$ coefficient matrix $C^x_\psi \in \mathcal{C}^{L\times N}$. However, the problem of finding a discrete approximation to the orthogonal projection operator $\mathcal{K} : H \to M$ is more involved. The approach followed here approximates $\mathcal{K}$ by using a successive application of the inverse and forward Shensa algorithms.

## 3.1 Discrete approximation to the projection operator

In this analysis, it is assumed that the wavelet $\psi \in L^2(\Re)$ arises from a multiresolution. Let $\phi \in V_0 \subset L^2(\Re)$ denote the scaling function associated with the multiresolution ($V_0$ is one of the spaces of the multiresolution ladder). Note that the collection $\{\phi_{0,n}\}_n$, where $\phi_{0,n} = \phi(t-n)$, constitutes an $ONB$ for $V_0$. There is an associated implicit assumption that we constrain the computation to the subspace $V_0$. With these assumptions, we can define a frame operator $F_2 : V_0 \to l^2(\mathcal{Z})$ as

$$F_2[x](n) = \langle x, \phi_{0,n}\rangle_2 \,, \quad x \in V_0$$

The adjoint operator $F_2^* : l^2(\mathcal{Z}) \to V_0$ is then given by

$$F_2^*[z](t) = \sum_n z(n)\phi_{0,n}, \quad z \in l^2(\mathcal{Z})$$

We also define a second operator, $\mathcal{T} : H \to l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$, which is the approximation to the sampling operator, as

$$\mathcal{T}[c](l,n) = c(2^l, n), \quad c \in H$$

The adjoint operator $\mathcal{T}^* : l^2(\mathcal{Z}^2, \frac{1}{2^{4l}}) \to H$ is then defined as

$$\mathcal{T}^*[h](a,b) = \sum_l \sum_n h(l,n) q_{l,n}(a,b), \quad h \in l^2(\mathcal{Z}^2, \frac{1}{2^{4l}})$$

where $q_{l,n}$ is the piecewise constant function defined earlier. Thus, $\mathcal{T}^*\mathcal{T}[c]$ is a piecewise constant approximation to $c \in H$ of the form

$$c(a,b) \approx \mathcal{T}^*\mathcal{T}[c](a,b) = \sum_{l=1}^{L} \sum_{n=1}^{N} c(2^l, n) q_{l,n}(a,b)$$

Let $\mathcal{S} : l^2(\mathcal{Z}) \to l^2(\mathcal{Z}^2)$ denote the forward Shensa operator. Clearly, an element in the range of $\mathcal{S}$ is also in $l^2(Z^2, \frac{1}{2^{4l}})$ (since $l$ assumes only non-negative integer values). Thus, $\mathcal{S} : l^2(\mathcal{Z}) \to l^2(Z^2, \frac{1}{2^{4l}})$. Then, for $x \in V_0$, the map $\mathcal{S}F_2$ defines a matrix $C_\psi^x$ with samples, $c_\psi^x(2^l, n)$, of the wavelet transform of $x$, i.e. $\mathcal{S}F_2 : V_0 \to l^2(Z^2, \frac{1}{2^{4l}})$ is defined as

$$\mathcal{S}F_2[x](l,n) = \langle x, \psi_{l,n} \rangle_2, \quad x \in V_0$$

The adjoint operator $F_2^*\mathcal{S}^* : l^2(Z^2, \frac{1}{2^{4l}}) \to V_0$ is then obtained as

$$F_2^*\mathcal{S}^*[C](t) = \sum_{l,n} C(l,n)\psi_{l,n}(t)$$

Note that, if $C$ was a matrix of discretized $CWT$ coefficients, $\mathcal{S}^*$ is exactly the inverse Shensa operator. Thus, the adjoint operator $\mathcal{S}^* : l^2(Z^2, \frac{1}{2^{4l}}) \to l^2(\mathcal{Z})$ is effectively the extended inverse Shensa operator. Clearly, $\mathcal{S}^*\mathcal{S} = I$. It can be shown that, in this framework, $\mathcal{S}\mathcal{S}^*$ is the discrete approximation to $\mathcal{K}$.

With the above terminology, the approximation to the operators $\Gamma$ and $\Gamma^*$ is given by

$$\begin{aligned} \Gamma &\approx \mathcal{T}^*\mathcal{S}F_2 : V_0 \to H \\ \Gamma^* &\approx F_2^*\mathcal{S}^*\mathcal{T} : H \to V_0 \end{aligned}$$

As shown in [7], the orthogonal projection operator $\mathcal{K} : H \to M$ can be represented as $\mathcal{K} = \Gamma\Gamma^*$. This result can be used to obtain a discrete approximation to $\mathcal{K}$ as $\mathcal{K} = \mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}$. It is clear that $\mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}$ is not an orthogonal projection, since for any $c \in M$, $\mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}c$ only provides a piecewise constant approximation to $c \in M$. However, if we assume that the $CWT$ is indeed piecewise constant, then $\mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}$ can be used as an approximation to the orthogonal projection operator. To see this more clearly, observe that

$$\mathcal{T}^*\mathcal{S}\mathcal{S}^*\mathcal{T}[c](a,b) = \sum_{l',n'} \left\langle \sum_{l,n} c(2^l, n)\psi_{l,n}(t), \psi_{l',n'} \right\rangle_2 q_{l',n'}(a,b), \quad c \in H$$

$$
\begin{aligned}
&= \sum_{l,n} c(2^l, n) \sum_{l',n'} c_{\psi}^{\psi_{l,n}}(2^{l'}, n') q_{l',n'}(a,b) \\
&= \sum_{l,n} c(2^l, n) c_{\psi}^{\psi_{l,n}}(a,b), \quad if\ c_{\psi}^{\psi_{l,n}}\ is\ piecewise\ constant \\
&= c(a,b), \quad if\ c \in M
\end{aligned}
$$

With $s_d(l) = s_{\psi}^{x}(2^l)$, $s_d \in \mathcal{C}^L$, and $r_d(n) = r_{\psi}^{x}(n)$, $r_d \in \mathcal{C}^N$, the element $\mathcal{K}[s_{\psi}^{x} r_{\psi}^{x}]$ can be approximated by the finite dimensional matrix $\tilde{C} \in \mathcal{C}^{L \times N}$ resulting from the operation $\mathcal{S}\mathcal{S}^*[s_d r_d^T]$.

The infinite dimensional minimization can now be formulated as the following finite dimensional problem :

**Given a matrix $C_{\psi}^{x} \in \mathcal{C}^{L \times N}$ of samples on the Shensa grid of the $CWT$ of $x \in L^2(\Re)$, determine the rank one matrix $s_d r_d^T \in \mathcal{C}^{L \times N}$ such that the following functional is minimized**

$$
J(s_d, r_d) = \parallel C_{\psi}^{x} - \mathcal{S}\mathcal{S}^* \left[ s_d r_d^T \right] \parallel_2^2 + \left\| s_d r_d^T \right\|_2^2
$$

# 4 Solution to the minimization problem

This section presents the solution procedure for the finite dimensional minimization problem. The technique uses an iterative approach, where, in each iteration, we successively minimize with respect to the vectors $s_d \in \mathcal{C}^L$, and $r_d \in \mathcal{C}^N$. The basic framework leading to the solution procedure is established below. The entire development is in $l^2$.

## 4.1 Existence of the minimizer

The first step in developing a solution procedure is to establish the existence of a solution to the finite dimensional minimization problem. The finite dimensional minimization problem has a solution based on the following result.

**Theorem 4.1** *There exists* $\bar{s}_d \in B_L = \{s_d \in \mathcal{C}^L;\ \ \|s_d\|_2 \leq 1\}$, *and* $\bar{r}_d \in \mathcal{C}^N$ *such that*

$$J(\bar{s}_d, \bar{r}_d) = \inf_{s_d \in B_L, r_d \in \mathcal{C}^N} J(s_d, r_d)$$

The proof is given in Appendix $A$.

## 4.2 Convergence to the optimal

Once the existence of the minimizer has been established, we can develop a procedure for its determination. The procedure followed here is an iterative one, and requires successively solving a set of necessary conditions in each iteration. From Lemma A.2, one can immediately see that the iterative procedure outlined produces a monotonically decreasing sequence of costs $\{J^i\}_i$, since one has the result $J(s_d^0, r_d^0) \geq J(\tilde{s}_d^1, r_d^0) \geq \cdots \geq J(s_d^i, r_d^i) \geq \cdots$, from which one can extract the monotonically decreasing sequence, $\{J^i\}_i;\ \ J^i = J(s_d^i)$, defined on the unit ball $B_L$ which is compact. The convergence of the iterative procedure is established based on the following result.

**Theorem 4.2** *There exists* $\bar{s}_d \in B_L$, *and* $\bar{J} \geq 0$ *such that the sequence of costs* $\{J^i\}_i$ *converges to* $\bar{J} = J(\bar{s}_d)$.

**Proof**. By Lemma A.3, the real valued functional $J(s_d)$ is continuous on the compact set $B_L$. Hence, it is guaranteed to attain its maximum and minimum on $B_L$. The iterative procedure produces a monotonically decreasing sequence of costs $\{J^i\}_i$, whose limit $\bar{J}$ exists by Lemma A.3. Hence, $\lim J^i = \bar{J}$. The set $B_L$ is also sequentially compact. This implies that from the sequence $\{s_d^i\}_i$, one can extract a subsequence $\{s_d^{i_j}\}_{i_j}$ that converges to some $\bar{s}_d \in B_L$ such that by the continuity of $J$, $J(\bar{s}_d) = \lim J(s_d^{i_j}) = \lim J^{i_j} = \bar{J}$, establishing the convergence of the algorithm to the optimal solution. □

An iterative procedure using successive minimization with respect to $s_d$, and $r_d$, which converges to the limiting solution whose existence is guaranteed, has thus been developed. It is important to note that the procedure offers no guarantee that the minimum attained is global. This is a general problem in nonlinear minimization techniques. The sufficient condition to ensure that the minimum attained is indeed global is that the functional $J$ be *jointly* convex in the variables $s_d$ and $r_d$, which is not the case in this problem.

# 5 Computational algorithm

The last section outlined a solution procedure for the determination of the discrete projection signatures for signal classes. In this section, the computational algorithm developed based on the procedure is presented. The computational algorithm used to generate the projection signatures is given below.

1. Select a wavelet $\psi \in L^2(\Re)$ which arises from a $MRA$, and the number of levels $L$ to be used in the filter bank corresponding to the $MRA$. Denote the analysis low

pass filters as $f \in \mathcal{C}^{N_\psi}$, and the analysis high pass filters as $g \in \mathcal{C}^{N_\psi}$.

2. For the given finite discrete input signal $x \in \mathcal{C}^{N_x}$, determine the discretized $CWT$ coefficient matrix $C_\psi^x \in \mathcal{C}^{L \times N}$ using the forward Shensa algorithm. Note that $N = N_x + (2^L - 1)(N_\psi - 1)$.

3. Based on the scalogram $SC_\psi^x(l,n) = \mid C_\psi^x(l,n) \mid^2$ obtained, modify the value of $L$ such that $C_\psi^x(l,n) \approx 0$, for all $l \geq L$, $n \geq N$. Recompute $C_\psi^x$ using the modified value of $L$.

4. Pick random vectors $s_d^0 \in \mathcal{C}^L$, $r_d^0 \in \mathcal{C}^N$, and set a value for $tol$.

   - At the $i$ th stage, set $r_d = r_d^{i-1}$. Using the conjugate gradient technique ([10]), with gradient given by $\Delta_{s_d} = \Re\{(s_d {r_d^{i-1}}^T + \mathcal{S}\mathcal{S}^* \left[s_d {r_d^{i-1}}^T\right] - C_\psi^x) {r_d^{i-1}}^c\}$, solve the minimization problem for $s_d$. Let $\tilde{s_d}^i$ denote the solution.

   - Set $s_d^i = \frac{\tilde{s_d}^i}{\|\tilde{s_d}^i\|}$.

   - Next, with $s_d = s_d^i$, using the conjugate gradient technique with gradient $\Delta_{r_d} = \Re\{(s_d^i {r_d}^T + \mathcal{S}\mathcal{S}^* \left[s_d^i {r_d}^T\right] - C_\psi^x)^T {s_d^i}^c\}$, solve the minimization problem for $r_d$ . Let $r_d^i$ denote the solution.

   - Compute the cost function $J(s_d^i, r_d^i)$. If ( $J(s_d^{i-1}, r_d^{i-1}) - J(s_d^i, r_d^i)$ ) $<$ $tol$, terminate.

5. end.

## 5.1 Computational complexity of the algorithm

The minimization problem under consideration is separately quadratic in $s_d$, and $r_d$. Consequently, the use of the conjugate gradient technique for each minimization guarantees convergence for each minimization in $\mathcal{O}(V)$ steps, where $V$ is the size of the vector over which one is minimizing. It is clear that the computational cost associated with the technique depends on the number of iterations $I$ required to reach the optimal solution, which in turn, is a function of *tol*, and the initial condition. The complexity of each iteration is largely a factor of the complexity of the conjugate gradient technique, and the complexity of the implementation of $\mathcal{S}$ and $\mathcal{S}^*$. The complexity of the conjugate gradient technique is $(6V+2)$ multiplications and $(6V-3)$ additions per iteration. For a practical implementation, especially when $V$ is large (like, for example, when we are minimizing with respect to $r_d$), we usually prescribe a termination criteria so that convergence occurs in far fewer than $V$ steps ([10]). The complexity of $\mathcal{S}$ is $2LN_\psi N_x$ multiplications and $2L(N_\psi - 1)N_x$ additions, and that of $\mathcal{S}^*$ is exactly the same ([11]). Thus, the complexity of the Shensa algorithm is a linear function of $N_x$, $L$, and $N_\psi$. If we assume an average of $T_s$ steps for convergence of the conjugate gradient technique for the minimization with respect to $s_d$, and $T_r$ steps for the minimization with respect to $r_d$, the overall cost of the projection algorithm is given by $I[(6L + 2 + 4LN_\psi N_x)T_s + (6N + 2 + 4LN_\psi N_x)T_r]$ multiplications and $I[(6L-3+4L(N_\psi-1)N_x)T_s+(6N-3+4L(N_\psi-1)N_x)T_r]$ additions.

# 6 Simulation results

This section presents the results of applying the iterative solution technique to different nonstationary signals. These results serve to illustrate the potential capabilities of the projection signatures, and also the limitations of the computational procedure used to determine these signatures. A summary of the simulation experiments and the corresponding results is given in Table 1.

## 6.1 Signature quality and application to classification

The first experimental result presented is the classification problem discussed in [4]. The signals considered (Figure 2) are the simple modulated sinc ($sinc(x) = \frac{sin(\pi x)}{\pi x}$) functions $\{x1, x2, x3\}$ given by

1) $x_1(t) \;=\; e^{j.5\pi t} sinc(\frac{t}{3})$

2) $x_2(t) \;=\; e^{j.6\pi t} sinc(\frac{t}{3})$

3) $x_3(t) \;=\; e^{j1.55\pi t} sinc(\frac{t}{3})$

The projection signatures obtained using the $Db4$ wavelet, for $L = 6$, and $r_d^0 = \chi_x$ [4], are also shown in Figure 2. Observe how the projection signatures clearly separate the highly correlated signals $x_1$ and $x_2$. These signatures were then applied to the classification problem described in Figure 3. The figure displays the composite signal obtained with $x1$ in the interval [-125:-50], $x2$ in the interval [-50:50] and $x3$ in the interval [50:115].

[4] $\chi_x \in \mathcal{C}^N$ is the characteristic function of $x \in \mathcal{C}^{N_x}$ defined as

$$\chi_x(n) \;=\; \begin{cases} 1, & 1 \leq n \leq N_x \\ 0, & N_x \leq n \leq N \end{cases}$$

The results of the classification using a correlation approach are shown here. Two assumptions are made in performing this classification.

1) All the signal classes are present.

2) Only one signal class is present at any given time.

It was established that the pseudo power signature of a signal represents the normalized scale power distribution, and is independent of time. Thus, one can get an accurate picture of the signal composition, with particular reference to the location of the transition points, if one determines the correlation of each signature with the discretized $CWT$ of the composite signal for each $b$. Figure 3 also shows the correlation graphs of the projection signatures with the discretized $CWT$ of the composite signal. From the correlation graphs, we can conclude with a high degree of confidence that the signal $x_1$ is present in $[-125 : -50]$, $x_2$ in the segment $[-50 : 50]$, and $x_3$ in $[50 : 115]$. Notice the high correlation values obtained using the projection signatures, and the clear demarkation of the transition points. The high correlation values are of great importance when we need to classify a signal where *it is not known apriori if all the events are present.*

## 6.2 Reliability analysis

For use in classification applications, it is extremely important that the projection signatures be reliable measures of the classes they represent. For samples of different signals in the same signal class, the projection signatures should not show significant variations. Consider a common example of a speech signal, say for example, the letter "$A$" as spoken by a person on several different occasions (this constitutes a class). It is reasonable to expect that slight variations will exist between every utterance of the letter by the same person. However, a reliable signature for the signal class should be fairly insensitive to these variations. Figure 4 presents the results of the reliability experiment using this example. In the figure, Sample #1 represents the projection signatures for 5 different samples of the letter "$A$" as spoken by a person on a given day, and Sample #2, the projection signatures for samples obtained on a different day. Observe the very high correlations ($.9705 - .9950$) between all the different signatures. The wavelet used was $Db4$ with $L = 8$. As before, in every case, $r_d^0$ was taken to be the characteristic function of the sample signal.

The experiment described above is significant for several reasons. The reliability results obtained from the experiment reflect those obtained from several other similar experiments. First, the experiment shows that the projection signatures are true measures of the signal class they characterize, which is an essential requirement of any representation. Next, it shows that the pseudo power signature technique can be applied to real data signals which may have noise and random variations, even though every signal essentially belongs to the same class. Finally, it reiterates the claim that the signatures are independent of signal duration, since every signal sample considered in the example was of a duration different from the others.

## 6.3 Robustness analysis

Robustness of the projection signatures in the presence of noise adds to their effectiveness. While the projection signatures were fairly robust in the presence of low noise (as seen in the speech example), they are not very robust in the presence of a lot of noise (Figure 5). This is because these signatures closely characterize the true signal, and hence, are more sensitive to signal variations. Also, since the projection algorithm was not guaranteed to attain the global minimum, it is possible that different local minima are attained when the signal is highly corrupted by noise. Thus, when dealing with very noisy signals, it might be necessary to prefilter the signal to remove some of the noise before applying the projection technique.

Another study of importance in the robustness analysis of the projection signatures is the sensitivity to the initial condition. From experimental analysis, for $\alpha = 1$, the projection signatures did not depend on the initial choice of $s_d^0$. The robustness analysis then is effectively based on the sensitivity to the initial vector $r_d^0$. In many cases, the projection signatures were completely unaffected by the choice of the initial condition. However, for the three signal classes shown in Figure 2, the projection signatures (especially $S1$, and $S2$) show quite some variation when a different initial point is used, as is evident from Figure 6. In the figure, the results are shown for $r_d^0$ chosen randomly, and for $r_d^0$ obtained from the principal component of the $SVD$ of the discretized $CWT$ matrix for each signal. This result suggests that the solution technique does not converge to the global minimum in all cases. This is a common limitation of nonlinear minimization algorithms, and we usually get around it by analyzing different initial conditions, and selecting the one that gives best results. For classification problems, one of the main requirements is that the projection signatures obtained should clearly separate even closely spaced signal classes. Intuitively, we can see that for a signal $x$, if $r_d^0 = \chi_x$, then it has little information pertaining to the intrinsic signal properties, which would 'force' the associated projection signature to capture most of the information about the intrinsic properties of the signal. On this basis, one could conclude that the projection signature would better represent the signal class. This conclusion was borne out by several experimental studies on different signal classes and a variety of initial conditions. Thus, the initial vector $r_d^0$ was always set at $r_d^0 = \chi_x$ for the determination of the projection signatures.

## 6.4 Effect of the analyzing wavelet

The solution procedure used to generate the projection signatures assumes a fixed admissible wavelet $\psi \in L^2(\Re)$ of compact support that arises from a $MRA$. It is worthwhile to study the effect of using different wavelet functions on the projection signatures of signals. An example of the effect of using different wavelet functions of the Daubechies family ([8]) is illustrated in Figure 7. It can be observed that while there are some variations in the projection signatures obtained, they are essentially of the same nature. Nevertheless, the problem of optimal wavelet selection for better classification is an area which needs to be studied in depth with a wider variety of wavelets. One possible advantage of this study is that if one can indeed select an optimal wavelet, one might be able to obtain sufficiently discriminating signatures using the simple, and more robust, matrix $SVD$ approach ([4]).

## 6.5 Convergence issues

In the discussion on the computational complexity of the projection algorithm, an average number of iterations for the convergence of the conjugate gradient algorithm for each minimization was assumed. From several experiments, it was observed that convergence was usually achieved in $L$ steps for the minimization with respect to $s_d$, but far fewer than $N$ steps (less than 15 steps for $N = 30000$) for the minimization with respect to $r_d$ without the need for a termination criteria. The number of iterations $I$ required for convergence to the minimum showed a wide variation, with values as low as $I = 3$ for a signal of length $N_x = 201$, to values as high as $I = 19$ for a signal of length $N_x = 25000$. In both cases, $N_\psi = 8$, and $L = 8$. For illustration purposes, some sample results on eight different signals of length $N_x = 201$, with $N_\psi = 8$, and $L = 8$, are shown in Figure 8. It is seen that, in most of the cases, the maximum reduction in cost is achieved in the first iteration, with only marginal improvements in the subsequent steps. This figure also serves to validate the monotonically decreasing nature of the cost function. From the figure, and several other simulation results, one may then conclude that the algorithm used to generate the projection signatures converges to the minimum in relatively few iterations; i.e., it has a reasonably fast convergence rate.

## 6.6 Quality of the classification

The last issue addressed in this section deals with the actual quality of the classification results. For an unambiguous classification, we would ideally like to have correlation graphs which are relatively smooth over different segments, and have sharp transition points. While the second criteria is reasonably met by the correlation graphs shown in Figure 3, the first is not. There exist some oscillations in the correlation graphs over the different segments which cannot be easily explained.

A second issue regarding the quality of the classification is the use of the simple correlation technique as the basis for the classification. While it is true that the correlation approach gives fairly good results for artificially generated signals, it is reasonable to expect that we might need a more sophisticated technique when dealing with real data signals. At the very least, we might expect to need some additional techniques, along with the straightforward correlation approach. These techniques could take the form of preprocessing the signal to remove noise, and removing cross - correlation effects occurring due to interaction between adjacent signal classes. The latter technique would be extremely useful when dealing with signals where two or more events are present at the same time. This is an issue for further study, and will be addressed in greater depth in a future work.

# 7 Conclusions and future work

In this paper, we proposed a technique using the $CWT$ to compute pseudo power signatures for signal classes based on projections. The technique involved a nonlinear minimization, and we provided the complete solution to the minimization problem. We also developed an efficient algorithm for the computation of the signatures using an iterative procedure with fast convergence, and illustrated the quality of the projection signatures through representative examples. It is important to note that the actual classification can be done very quickly, since the signatures are vectors of very small dimension. This approach has wide applicability, in areas as diverse as oil exploration, hidden mine detection, moving target detection, system identification, and pattern recognition.

Though the methodology has some distinct advantages, it also suffers from certain limitations. This leaves scope for further work in the area. Two problems we are currently exploring are techniques to obtain more accurate discrete representations for the orthogonal projection operator $\mathcal{K}$, and the classification problem where two or more events are present at any given time. The latter is a very practical situation, since it is unrealistic to expect that adjacent signal components will not interact with each other, and that a composite signal consisting of several different components, will not exhibit some characteristics which result from these interactions. An additional area we are exploring is the applicability of these signatures to speech processing. Our initial results indicate several exciting possibilities for speaker, and speech recognition schemes.

## References

[1] V. Venkatachalam, and J.L. Aravena, "Nonstationary signal classification using pseudo power signatures," *1998 IEEE International Conference on Circuits and Systems*, Monterey, CA, June 1 - 3, 1998.

[2] O. Rioul and M. Vetterli, "Wavelets and Signal Processing," *Signal Processing Magazine*, vol.8, pp. 14-38, Apr. 1991.

[3] Jack McLaughlin and Les E. Atlas, "Applications of Operator Theory to Time-

Frequency Analysis and Classification," *accepted by IEEE Trans. on Signal Processing.*

[4] Vidya Venkatachalam and Jorge L. Aravena, "Nonstationary Signal Classification Using Pseudo Power Signatures : The Matrix $SVD$ Approach," *Revised version submitted to IEEE Trans. Circuits and Systems - II.*

[5] L. Cohen, "*Time-Frequency Analysis*," Prentice Hall Signal Processing Series, 1995.

[6] M. Reed and B. Simon, "*Methods of Modern Mathematical Physics 1 : Functional Analysis*," Academic Press, Inc., San Diego, 1980.

[7] Vidya Venkatachalam, "Pseudo Power Signatures For Nonstationary Signal Analysis And Classification," *Ph.D Dissertation*, Louisiana State University, Dec. 1998.

[8] I. Daubechies, "*Ten Lectures on Wavelets*," Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[9] Mark J. Shensa, "The Discrete Wavelet Transform: Wedding the À Trous and Mallat Algorithms," *IEEE Trans. on Signal Processing*, vol. 40, pp. 2464-2482, Oct. 1992.

[10] G. H. Golub and C. F. Van Loan, "*Matrix Computations - II Edition*," The Johns Hopkins University Press, 1989.

[11] Olivier Rioul and Pierre Duhamel, "Fast Algorithms For Discrete And Continuous Wavelet Transforms," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp 569 - 586, March 1992.

# 8 Appendix $A$

**Theorem A.1** *There exists* $\bar{s}_d \in B_L = \{s_d \in \mathcal{C}^L;\ \|s_d\|_2 \leq 1\}$, *and* $\bar{r}_d \in \mathcal{C}^N$ *such that*

$$J(\bar{s}_d, \bar{r}_d) = \inf_{s_d \in B_L, r_d \in \mathcal{C}^N} J(s_d, r_d)$$

**Proof**. The discrete orthogonal projection approximation operator $\mathcal{S}\mathcal{S}^*$ on the separable finite dimensional Hilbert space $l^2(\mathcal{C}^{L\times N})$, is isometrically equivalent to a square matrix $\tilde{K} : \mathcal{C}^{LN\times LN}$. This isometry, $T : \mathcal{C}^{L\times N} \to \mathcal{C}^{LN}$, essentially rearranges the elements of an $L \times N$ matrix into a column vector by stacking its rows. Let $T^* : \mathcal{C}^{LN} \to \mathcal{C}^{L\times N}$ denote the adjoint operator, with the property that $T^*T = I$. Then, $\tilde{K} = T\mathcal{S}\mathcal{S}^*T^*$. Observe that $\tilde{K}$ has the properties that it is Hermitian ($\tilde{K} = \tilde{K}^*$), idempotent ($\tilde{K}^2 = \tilde{K}$), and positive semidefinite ($\tilde{K} \geq 0$). With $I_L$ denoting the identity matrix of size $L$, and $c = TC_\psi^x$, the functional $J(s_d, r_d)$ can be redefined as

$$\begin{aligned} J(s_d, r_d) &= \| TC_\psi^x - T\mathcal{S}\mathcal{S}^*T^*Ts_dr_d^T \|_2^2 + \left\|Ts_dr_d^T\right\|_2^2 \\ &= \| c - \tilde{K}(r_d \otimes I_L)s_d \|_2^2 + \|(r_d \otimes I_L)s_d\|_2^2 \end{aligned}$$

where $\otimes$ denotes the standard Kronecker product. For a fixed $r_d$, one can thus define the following subproblem for minimization with respect to $s_d$ :

$$\min_{s_d} J(s_d) = \| c - \tilde{K}(r_d \otimes I_L)s_d \|_2^2 + \|(r_d \otimes I_L)s_d\|_2^2 \tag{A.1}$$

Similarly, denoting $c_T = TC_\psi^{xT}$, and $I_N$ as the identity matrix of size $N$, one can define the minimization problem with respect to $r_d$ for a fixed $s_d$ as

$$\min_{r_d} J(r_d) = \| c_T - \tilde{K}(s_d \otimes I_N) r_d \|_2^2 + \|(s_d \otimes I_N) r_d\|_2^2 \tag{A.2}$$

**Lemma A.2** *The solutions to the minimization problems defined in Equations A.1 and A.2 exist, and are unique.*

**Proof**. The positive real valued functional $J(\cdot)$ defined in Equations A.1 and A.2 is quadratic in the variables $s_d$ and $r_d$ respectively. Moreover, the quadratic form is positive definite, and so $J$ is (separately) convex in $s_d$, and $r_d$. Hence, each minimization problem has a unique solution. □

One can now determine the conditions on the unique minimizers to each of the subproblems. Since the functional $J$ is separately convex in $s_d$, and $r_d$, the first order necessary conditions as determined using Calculus of Variations, become sufficient to determine the minimizers. For minimization with respect to $s_d$ ($r_d$ fixed), equation A.1 simplifies to $J(s_d) = \langle c, c \rangle - 2\Re \langle (r_d \otimes I_L)^* c, s_d \rangle + \left\langle (\tilde{K} + I)(r_d \otimes I_L) s_d, (\tilde{K} + I)(r_d \otimes I_L) s_d \right\rangle$. Taking variations with respect to $s_d$, $\delta J_s = 2\Re \left\langle (r_d \otimes I_L)^* (\tilde{K} + I)(r_d \otimes I_L) s_d - (r_d \otimes I_L)^* c, \delta s_d \right\rangle$, where $\delta s_d$ is completely arbitrary. Setting $\delta J_s = 0$, one obtains the necessary condition for minimization with respect to $s_d$. Following a similar approach, the necessary condition for minimization with respect to $r_d$ is also obtained. These conditions are :

1) $(r_d \otimes I_L)^* (\tilde{K} + I)(r_d \otimes I_L) s_d - (r_d \otimes I_L)^* c = 0, \quad r_d \;\; \mathit{fixed}$

2) $(s_d \otimes I_N)^* (\tilde{K} + I)(s_d \otimes I_N) r_d - (s_d \otimes I_N)^* c_T = 0, \quad s_d \;\; \mathit{fixed.}$

Let $B_L = \{s_d \in \mathcal{C}^L;\ \ \|s_d\|_2 \leq 1\}$ denote the closed unit ball in $l^2(\mathcal{Z})$ of dimension $L$, which is compact. For a fixed $s_d$ of unit norm, $s_d \in B_L$. Let $\bar{r}_d = P_{s_d}^{-1}(s_d \otimes I_N)^* c_T$ be the solution to Equation A.2, with $P_{s_d} = (s_d \otimes I_N)^*(\tilde{K} + I)(s_d \otimes I_N)$. Then, the functional $J(s_d, \bar{r}_d) = \left\langle c, c - (s_d \otimes I_N) P_{s_d}^{-1}(s_d \otimes I_N)^* c_T \right\rangle$ is effectively a function of $s_d \in B_L$.

**Lemma A.3** *The real valued functional $J(s_d)$ is continuous on the compact set $B_L = \{s_d \in \mathcal{C}^L;\ \ \|s_d\|_2 \leq 1\}$.*

**Proof**. The real valued functional $J(s_d)$ defined on the compact set $B_L$ is given by $J(s_d) = \left\langle c, c - (s_d \otimes I_N) P_{s_d}^{-1}(s_d \otimes I_N)^* c_T \right\rangle, \ \ s_d \in B_L$. Let $\tilde{K} + I = \Lambda^2 > 0$, where $\Lambda$ is positive definite and self-adjoint. Let $\lambda_{min} > 0$ and $\lambda_{max} > 0$ denote the minimum and maximum eigen values of $\Lambda$. Defining $X(s_d) = \Lambda s_d \otimes I_N$, $P_{s_d} = X^*(s_d) X(s_d)$. Then,

$$\lambda_{min}^2 \parallel c \parallel^2 \quad \leq \langle \Lambda c, \Lambda c \rangle \quad \leq \lambda_{max}^2 \parallel c \parallel^2;\ \forall c \in \mathcal{C}^{LN} \tag{A.3}$$

$$\lambda_{min} \parallel s_d \parallel \quad \leq \parallel X(s_d) \parallel \quad \leq \lambda_{max} \parallel s_d \parallel \tag{A.4}$$

$$\lambda_{max}^{-2} \parallel r_d \parallel^2 \quad \leq \left\langle r_d, P_{s_d}^{-1} r_d \right\rangle \quad \leq \lambda_{min}^{-2} \parallel r_d \parallel^2;\ \forall r_d \in \mathcal{C}^N; s_d \in B_L \tag{A.5}$$

$$\textit{and} \quad \|\delta_s\| \quad \leq \|s_{d_1}\| + \|s_{d_2}\| \quad = 2;\ \ \delta_s = s_{d_1} - s_{d_2},\ s_{d_1}, s_{d_2} \in B_L \tag{A.6}$$

Now, $P_{s_{d_1}} - P_{s_{d_2}} = X^*(s_{d_2})X(\delta_s) + X^*(\delta_s)X(s_{d_2}) + X^*(\delta_s)X(\delta_s)$, and hence, from Equations A.4 and A.6,

$$\left\| P_{s_{d_1}} - P_{s_{d_2}} \right\| \leq 4\lambda_{max}^2 \parallel \delta_s \parallel \tag{A.7}$$

Also, $P_{s_{d_2}}^{-1} - P_{s_{d_1}}^{-1} = P_{s_{d_2}}^{-1} \left( P_{s_{d_1}} - P_{s_{d_2}} \right) P_{s_{d_1}}^{-1}$ which implies from Equations A.5 and A.7, $\parallel P_{s_{d_2}}^{-1} - P_{s_{d_1}}^{-1} \parallel \leq 4\lambda_{min}^{-4}\lambda_{max}^2 \parallel \delta_s \parallel;\ \forall s_{d_1}, s_{d_2} \in B_L$. Observe that one can express $J(s_d) =$

$\langle c, c\rangle - \frac{1}{4}\left\langle X^*(s_d)\Lambda c, P_{s_d}^{-1} X^*(s_d)\Lambda c_T\right\rangle$ from which one can obtain $\mid J(s_{d_2}) - J(s_{d_1}) \mid \leq A \|\delta_s\|$ where $A$ is a constant that depends only on $\| c \|$, $\lambda_{min}$, and $\lambda_{max}$. Then, given any $\epsilon > 0$, there exists $\delta = \frac{\epsilon}{2A} > 0$, such that for all $\| s_{d_1} - s_{d_2} \|_2 < \delta$, $\mid J(s_{d_2}) - J(s_{d_1}) \mid < \epsilon$. Thus, $J$ is continuous on $B_L$. □

The existence of the minimizing solution to the finite dimensional problem is established as follows. Fix $s_d \in B_L$. Then, by Lemma A.2, there exists $r_d(s_d) \in \mathcal{C}^N$ such that $J(s_d, r_d(s_d)) = J(s_d) = \inf_{r_d \in \mathcal{C}^N} J(s_d, r_d)$. The real valued functional $J$ is thus defined on the compact set $B_L$, and is continuous (Lemma A.3) on $B_L$. Hence, it is guaranteed to attain its maximum and minimum on the set, i.e. there exists $\bar{s}_d \in B_L$, such that

$$J(\bar{s}_d, r_d(\bar{s}_d)) = \inf_{s_d \in B_L} J(s_d, r_d(s_d)) = \inf_{s_d \in B_L, r_d \in \mathcal{C}^N} J(s_d, r_d)$$

Denoting $\bar{r}_d = r_d(\bar{s}_d)$, the existence of the minimum is established. □

Table 1: SIMULATION RESULTS ON THE PROJECTION SIGNATURES

| EXPERIMENT | ANALYSIS |
|---|---|
| Signature quality | Highly distinct |
| Classification | Unambiguous, with sharp transition points |
| Reliability | Good |
| Robustness | Not very robust in the presence of much noise |
| Sensitivity to initial condition | Sensitive to choice of initial condition |
| Analyzing wavelet | Some variation in quality |
| Computational requirements | Very reasonable with fast convergence |

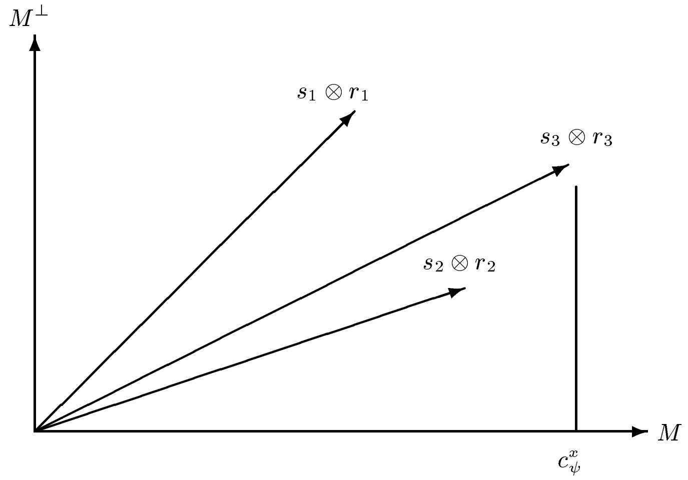


Figure 1: LIMITATIONS IN THE PRINCIPAL COMPONENT APPROACH TO DETERMINE SIGNATURES

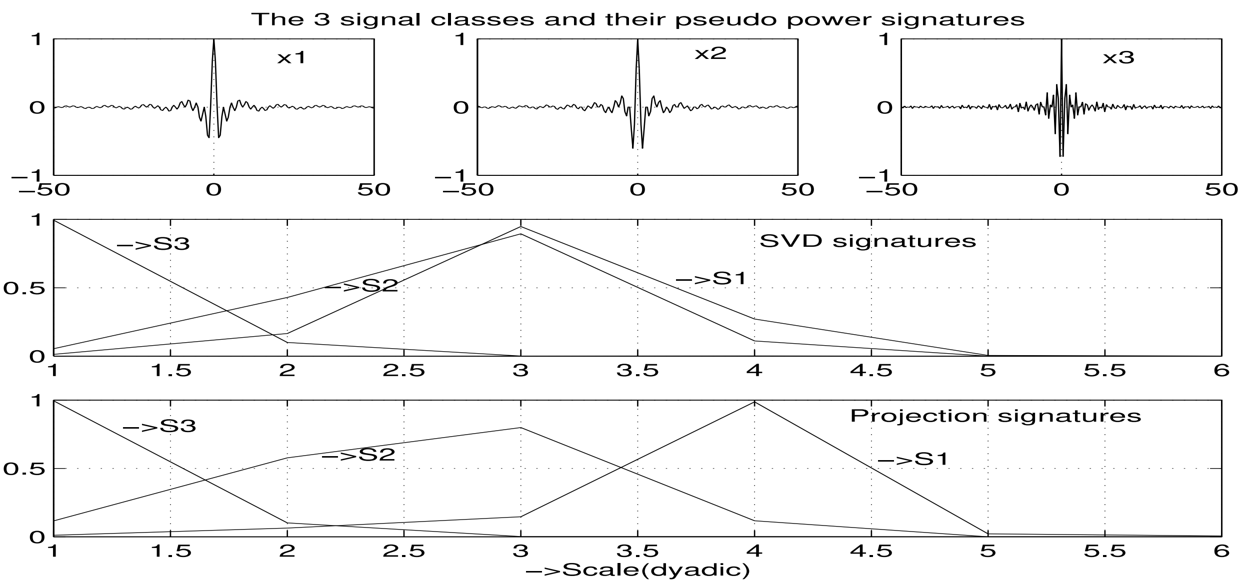


Figure 2: THE 3 SIGNAL CLASSES, AND THEIR MATRIX $SVD$ AND PROJECTION SIGNATURES

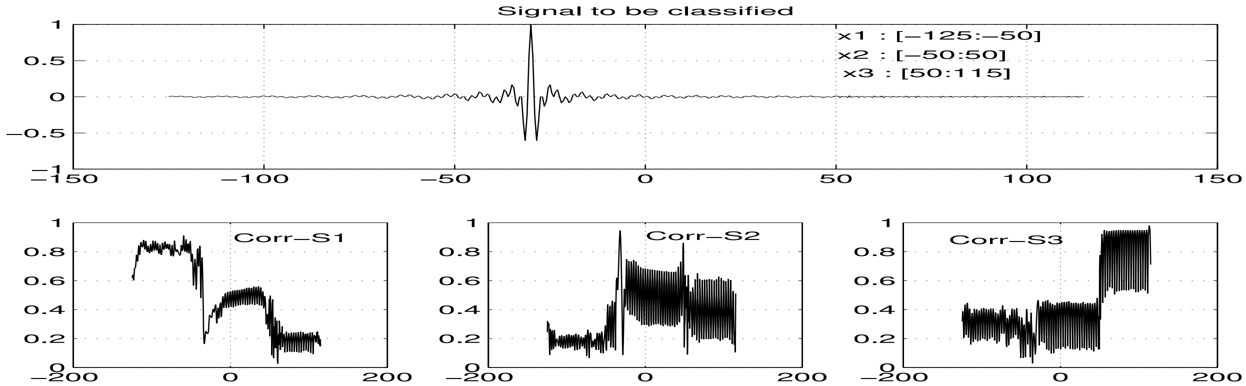


Figure 3: THE PROJECTION SIGNATURES APPLIED TO THE CLASSIFICATION PROBLEM

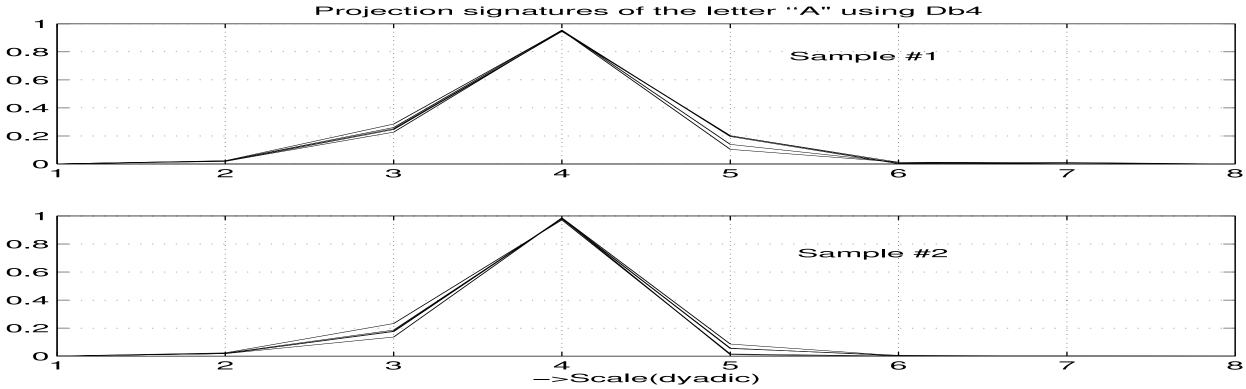


Figure 4: RELIABILITY TEST RESULTS FOR THE PROJECTION SIGNATURES

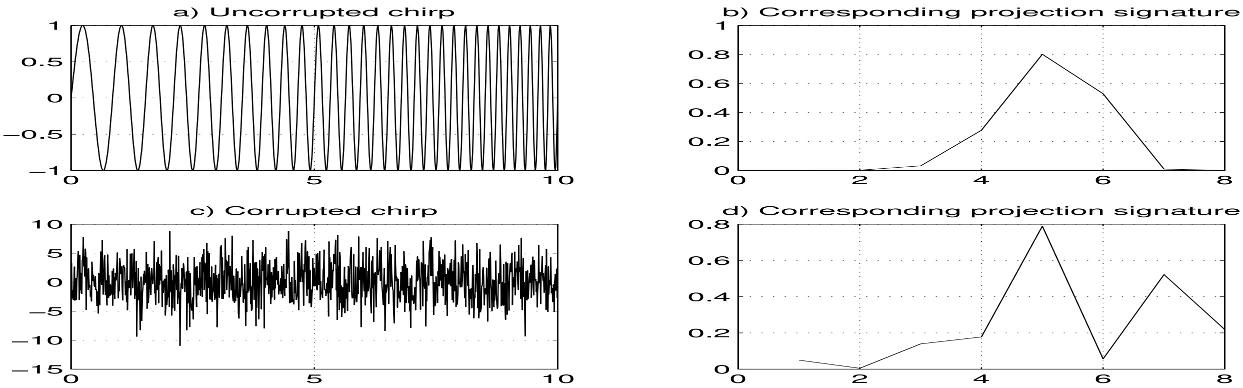


Figure 5: ROBUSTNES OF THE PROJECTION SIGNATURES

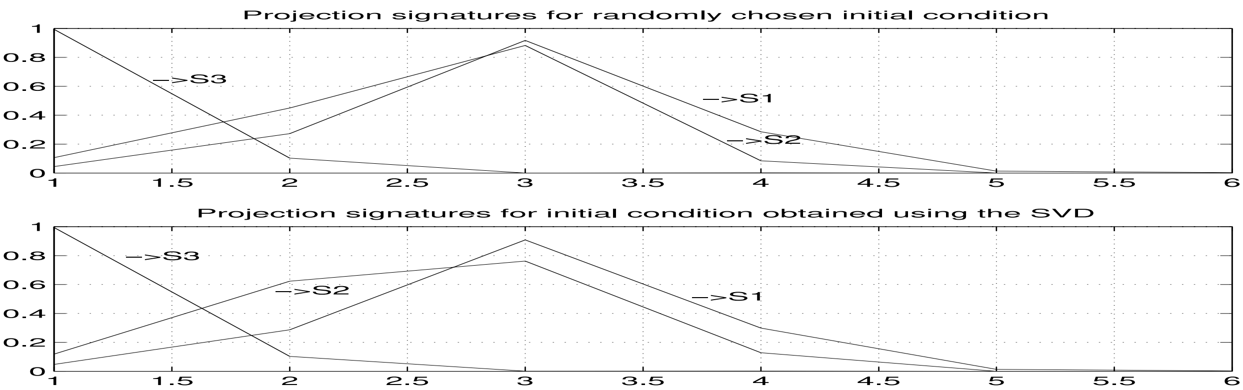


Figure 6: EFFECT OF DIFFERENT INITIAL CONDITIONS $r_d^0$ ON THE PROJECTION SIGNATURES

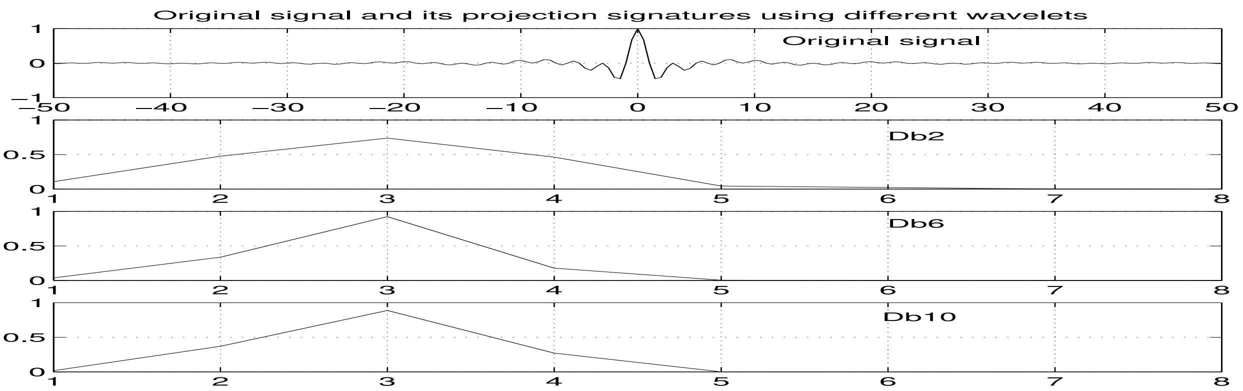


Figure 7: EFFECT OF DIFFERENT WAVELETS ON THE PROJECTION SIGNATURES

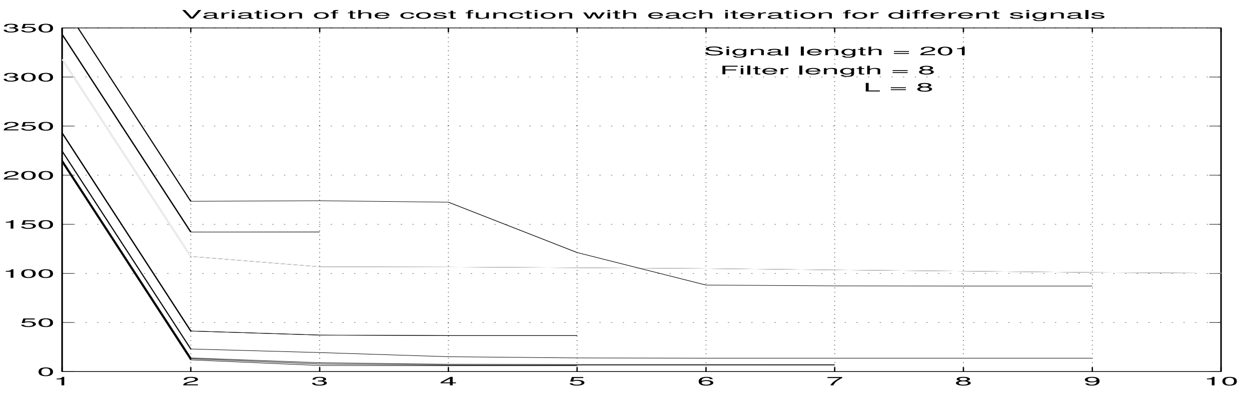


Figure 8: VARIATION OF THE COST FUNCTION WITH EACH ITERATION

## Table and illustration captions

- SIMULATION RESULTS ON THE PROJECTION SIGNATURES: Table 1.
- LIMITATIONS IN THE PRINCIPAL COMPONENT APPROACH TO DETERMINE SIGNATURES: Figure 1.
- THE 3 SIGNAL CLASSES, AND THEIR MATRIX $SVD$ AND PROJECTION SIGNATURES: Figure 2.
- THE PROJECTION SIGNATURES APPLIED TO THE CLASSIFICATION PROBLEM: Figure 3.
- RELIABILITY TEST RESULTS FOR THE PROJECTION SIGNATURES: Figure 4.
- ROBUSTNES OF THE PROJECTION SIGNATURES: Figure 5.
- EFFECT OF DIFFERENT INITIAL CONDITIONS $r_d^0$ ON THE PROJECTION SIGNATURES: Figure 6.
- EFFECT OF DIFFERENT WAVELETS ON THE PROJECTION SIGNATURES: Figure 7.
- VARIATION OF THE COST FUNCTION WITH EACH ITERATION: Figure 8.