BASIS SELECTION FOR WAVELET PROCESSING OF SPARSE SOURCE SIGNALS

Ian Atkinson and Farzad Kamalabadi

University of Illinois at Urbana-Champaign
Department of Electrical and Computer Engineering and the Coordinated Science Laboratory

ABSTRACT

An attractive property of wavelet bases is their ability to sparsely represent piecewise polynomial signals. The sparsity of a wavelet-domain representation depends on several factors such as the mother wavelet, the number of decomposition levels, and the structure of the original signal. We consider the problem of selecting an overcomplete or dyadic wavelet basis that can sparsely represent a *sparse* piecewise polynomial signal. Most existing applications that apply wavelet-domain processing techniques to signals that are inherently sparse have not considered the sparsity of underlying signal when selecting a wavelet basis. By accounting for the initial sparseness of a signal, the maximum wavelet filter length and number of decomposition levels can be computed. Selecting a wavelet basis that satisfies these maximum values guarantees that the resulting wavelet-domain representation will be at least as sparse as the original signal. This criteria for wavelet basis selection is of use in applications having sparse source signals.

***Index Terms*—** Wavelet transforms, Signal representations

1. INTRODUCTION

During the past decade, wavelet-based signal processing methods have received great popularity and success [1–4]. The ability of a wavelet basis to sparsely represent piecewise polynomail signals makes wavelets naturally suited to problems that benefit from sparse representations (i.e., compression, denoising, and approximation). In general, the sparsity of a wavelet representation of a signal depends on both the wavelet basis (determined by the mother wavelet and the number of decomposition levels) and the exact form of the source signal.

Most applications of wavelets focus on piecewise polynomial signals that have large spatial support or are composed of features that are much larger than the length of the wavelet (e.g., [1–4]). For such signals, almost any wavelet with a sufficiently large number of vanishing moments will provide a sparse representation (although some more sparse than others). The number of decomposition levels, $\mathcal{J}$, is typically selected to satisfy $1 < N/2^{\mathcal{J}} \ll N$ where N is the length of the signal.

In some important applications, however, the signal of interest does not have large spatial support and is not composed of large features. Such a source signal can be described as having some degree of sparsity in the signal domain. Examples of signals with this inherent sparsity include activation patterns in functional magnetic resonance imaging (fMRI) and the blood oxygenation level dependent signal in event-related fMRI. When wavelets have been applied to such signals in the past, the sparseness of the true signal has been ignored when choosing a wavelet basis (e.g., [5, 6]).

In this paper, we address the question of how to select a wavelet basis in which to express a signal that is known to be sparse. Although we consider only single dimensional signals, the concepts and results can be extended to higher dimensions. What differentiates this work from traditional wavelet methods (such as wavelet-based denoising and compression) is that we assume that the true signal has some degree of initial sparsity *before* being expressed in a wavelet-domain and consider this sparsity when selecting the mother wavelet and number of decomposition levels that define the wavelet basis.

2. WAVELET BASIS SELECTION

If $\mathbf{a}$ denotes a finite-length signal, then the sparsity of $\mathbf{a}$ can be expressed as

$$\text{Sparse}\,(\mathbf{a}) = 1 - ||\mathbf{a}||_0 / \text{ len}\,(\mathbf{a}) \tag{1}$$

where $||\cdot||_0$ is the number of non-zero elements of the argument and len $(\cdot)$ is the length of the argument. This measure of sparsity is the percentage of total elements in $\mathbf{a}$ that have a value of zero. Other sparsity measures can be imagined, such as the number of coefficients containing 95% of the total signal energy, but (1) represents the most strict sparsity measure.

Let us define a *feature* of a signal as one or more consecutive samples that can be represented as a finite-order polynomial. Let $\text{FS}_m\,(\cdot)$ denote the size (number of samples) of the m^{th} feature of its argument. Denote the number of non-zero elements of feature m as $||\cdot||_{0,m}$ and the polynomial order of feature m as $\text{FO}_m\,(\cdot)$.

Let $\mathbf{a}_J$ be a length-N signal with M features where N is an integer power of two and $J = \log_2(N)$. Since we are interested in signals that are initially sparse we will assume that

$$\sum_{m=1}^{M} \text{FS}_m\,(\mathbf{a}_J) \ll N \tag{2}$$

Let $R = \max_m \text{FO}_m\,(\mathbf{a}_J)$ be the maximum polynomial order of all features and let ϕ and ψ be length-L scaling and wavelet functions with $R+1$ vanishing moments, respectively. Finally, denote by $\mathbf{a}_j$ and $\mathbf{d}_j$ the approximation and detail coefficients at scale j of the wavelet expansion of $\mathbf{a}_J$ achieved using ϕ and ψ.

In order for the wavelet-domain representation of $\mathbf{a}_J$ to be at least as sparse as the original signal, the inequality

$$1 - \frac{||\mathbf{a}_{J-\mathcal{J}}||_0 + \sum_{j=J-\mathcal{J}}^{J-1} ||\mathbf{d}_j||_0}{\text{len}\,(\mathbf{a}_{J-\mathcal{J}}) + \sum_{j=J-\mathcal{J}}^{J-1} \text{len}\,(\mathbf{d}_j)} \geq \text{Sparse}\,(\mathbf{a}_J) \tag{3}$$

must be true for $0 < \mathcal{J} \leq J$. Our goal is therefore to determine suitable values for the wavelet filter length (L) and the number of decomposition levels ($\mathcal{J}$) such that (3) is true.

To be as conservative as possible, we will assume that during the wavelet expansion no features merge. This assumption requires the features to be sufficiently spaced, which is possible due to (2). If this assumption is violated and two or more features do merge during the wavelet expansion, the resulting wavelet-domain signal will be *more* sparse than if no merging had occurred. Therefore, if L and $\mathcal{J}$ are selected such that (3) is true under the "no feature merging" assumption, then (3) will also be true if there is feature merging.

2.1. Basis Selection For Dyadic Wavelets

Since the standard wavelet transform includes downsampling after filtering at each decomposition level, the scales of a standard wavelet expansion have dyadic sampling [7]. It will therefore be helpful to consider the wavelet-domain signal both before and after downsampling at each scale. Let $\bar{\mathbf{a}}_j$ and $\bar{\mathbf{d}}_j$ denote the non-downsampled approximation and detail coefficients at scale j. As described above, the standard (downsampled) wavelet coefficients are denoted by $\mathbf{a}_j$ and $\mathbf{d}_j$.

The wavelet decomposition of $\mathbf{a}_J$ is given by

$$\bar{a}_j[k_j] = \sum_{l=0}^{L-1} \phi[l] a_{j+1}[k_j - l] \qquad a_j[k_j] = \bar{a}_j[2k_j]$$

$$\bar{d}_j[k_j] = \sum_{l=0}^{L-1} \psi[l] a_{j+1}[k_j - l] \qquad d_j[k_j] = \bar{d}_j[2k_j]$$

for $k_j = 0, 1, \ldots, 2^{j-J}N-1$ and $j = J-\mathcal{J}, J-\mathcal{J}+1, \ldots, J-1$. For the conservative assumption that no features merge when moving from scale $j+1$ to scale j, the feature size of $\bar{\mathbf{a}}_j$ and $\mathbf{a}_{j+1}$ are related as

$$\mathrm{FS}_m(\bar{\mathbf{a}}_j) = \mathrm{FS}_m(\mathbf{a}_{j+1}) + L - 1 \tag{4}$$

From (4), we see that $\mathrm{FS}_m(\mathbf{a}_j)$ will be an integer satisfying the inequality

$$\lfloor \mathrm{FS}_m(\bar{\mathbf{a}}_j)/2 \rfloor \leq \mathrm{FS}_m(\mathbf{a}_j) \leq \lceil \mathrm{FS}_m(\bar{\mathbf{a}}_j)/2 \rceil \tag{5}$$

If $\mathrm{PC}_m^R(\cdot)$ is the number of coefficients in feature m that can be described by an R^{th}-order polynomial, then

$$\mathrm{PC}_m^R(\bar{\mathbf{a}}_j) = \max\Big(\mathrm{PC}_m^R(\mathbf{a}_{j+1}) - (L-1),\ 0\Big) \tag{6}$$

$$\mathrm{PC}_m^R(\mathbf{a}_j) \geq \lfloor \mathrm{PC}_m^R(\bar{\mathbf{a}}_j)/2 \rfloor \tag{7}$$

This means that the number of non-zero elements in the m^{th} feature of the detail signal at scale j is

$$||\bar{\mathbf{d}}_j||_{0,m} \leq \mathrm{FS}_m(\mathbf{a}_{j+1}) + (L-1) - \mathrm{PC}_m^R(\bar{\mathbf{a}}_j) \tag{8}$$

$$||\mathbf{d}_j||_{0,m} \leq \lceil ||\bar{\mathbf{d}}_j||_{0,m}/2 \rceil \tag{9}$$

To satisfy the sparsity constraint given in (3), L and $\mathcal{J}$ must be selected such that the inequality

$$\sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_{J-\mathcal{J}}) + \sum_{m=1}^{M} \sum_{j=J-\mathcal{J}}^{J-1} ||\mathbf{d}_j||_{0,m} \leq \sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_J) \tag{10}$$

is true. Unfortunately, due to the *max* function in (6) and the *floor* and *ceiling* functions in (5), (7), and (9), no simplified expressions can be calculated for either the filter length or the number of decomposition levels. However, maximum values for both L and $\mathcal{J}$ can be computed by recursively applying (4)-(9) and using (10) as a stopping condition.

Although we cannot derive simple expressions from (10) that exactly describe the maximum values of L and $\mathcal{J}$, it is possible compute simple expressions that approximate the maximum filter length and decomposition levels. Since all features have integer size and for any integer A, the inequalities $2\lfloor A/2 \rfloor \leq A+1$ and $2\lceil A/2 \rceil \geq A-1$ are true, it can be shown that

$$\mathrm{FS}_m(\mathbf{a}_j) \leq \frac{\mathrm{FS}_m(\mathbf{a}_J)}{2^{J-j}} + (1 - 2^{-J+j})L \tag{11}$$

Furthermore, since $A \leq \max(A, 0)$ for any value A, it can be shown that

$$\mathrm{PC}_m^R(\bar{\mathbf{a}}_j) \geq \frac{\mathrm{FS}_m(\mathbf{a}_J)}{2^{J-(j+1)}} - 2(1 - 2^{-J+j})L + 1 \tag{12}$$

Using (8), (9), (11), and (12), the number of non-zero detail coefficients at scale j can be upper bounded as

$$||\mathbf{d}_j||_{0,m} \leq 2(1 - 2^{-J+j})L \tag{13}$$

Substituting (11) and (13) into (9) and simplifying gives inequalities describing the maximum filter length and number of decomposition levels.

$$L \leq \frac{\frac{1}{M}\sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_J)\left(1 - 2^{-\mathcal{J}}\right)}{2\mathcal{J} - (1 - 2^{-\mathcal{J}})} \tag{14}$$

$$\frac{2\mathcal{J}}{1 - 2^{-\mathcal{J}}} \leq \frac{1}{ML} \sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_J) + 1 \tag{15}$$

Interestingly, (14) and (15) reveal that the maximum values for both L and $\mathcal{J}$ depend on the average feature size and not on the total or minimum feature size as one may expect. To select a dyadic wavelet basis for a sparse source signal one needs simply to determine the average feature size (generally dictated by the specifics of the problem), choose either L or $\mathcal{J}$, and use (14) or (15) to select the other value.

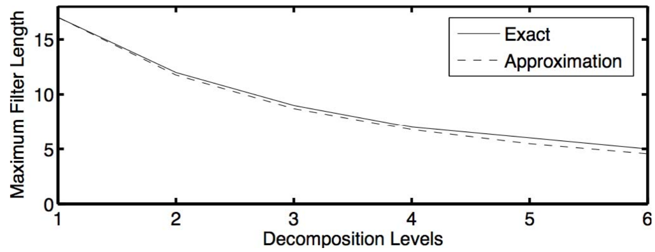

Figure 1. Exact and approximate maximum filter length for a dyadic wavelet expansion of signal with an average feature size of 51. The exact maximum filter length (solid) was found by numerically solving (10) while the approximate maximum filter length (dashed) was found using (14).

Figure 1 compares finding the maximum filter length by numerically solving (10) to reach an exact solution and using (14) to get an approximate value. From this graph, we see that (14) is a very good approximation to the exact maximum filter length, especially for a small number of decomposition levels. When the maximum L is restricted to integer multiples of 2, the exact and approximate solutions match exactly for all but two decomposition levels with the approximate solution being more conservative in cases where there is a difference.

In addition to satisfying (14), L must be *large* enough to allow ψ to have R vanishing moments. Therefore, when a Daubechies wavelet, which provides the maximum number of vanishing moments for a given filter length [7], is used, $L \geq 2(R+1)$. This lower bound on L allows us to consider the maximum polynomial order for which we can satisfy (3) with *some* value of L and $\mathcal{J}$. From (4) and (9) we can reason that in order for a $J-j$ level expansion to be more sparse than a $J-j+1$ level expansion, the number of coefficients added due to the edge effects of the feature must be less than the number of polynomail coefficients annihilated by the wavelet filter. This means that if an expansion with $\mathcal{J} = 1$ cannot be made more sparse than the original signal, we cannot hope to increase sparsity by increasing $\mathcal{J}$. Making the substitutions $L = 2(R+1)$ and $\mathcal{J} = 1$ in (14) and solving for R gives

$$R \leq \frac{1}{6M} \sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_J) - 1 \tag{16}$$

which bounds the maximum polynomial order for which we are guaranteed to be able to find a wavelet filter that can provide a representation that is at least as sparse as the original signal.

2.2. Basis Selection for Overcomplete Wavelets

The classic wavelet transform includes downsampling operations that cause dyadic wavelet expansions to be shift-variant. Overcomplete wavelets are often used to overcome the shift-varying nature of dyadic wavelet expansion [2,7] and it has been demonstrated that use of overcomplete, rather than dyadic, wavelets often leads to higher quality results in applications such as denoising [2].

An overcomplete wavelet expansion is computed exactly like a standard wavelet expansion except that there are no downsampling operations and modified wavelet and scaling functions are used at each decomposition level [7]. Let ϕ_j and ψ_j denote the wavelet and scaling functions applied to move from scale $j+1$ to scale j. These functions are obtained by upsampling and scaling ϕ and ψ by appropriate amounts [7]. An overcomplete wavelet expansion is given by

$$a_j[k] = \sum_{l=0}^{L_j-1} \phi_j[l] a_{j+1}[k-l] \qquad d_j[k] = \sum_{l=0}^{L_j-1} \psi_j[l] a_{j+1}[k-l]$$

for $k = 0, 1, \ldots, N-1$, $j = J - \mathcal{J}, J - \mathcal{J} + 1, \ldots, J-1$, and where the length of ϕ_j and ψ_j is

$$L_j = 2^{J-j-1}(L-1) + 1 \tag{17}$$

From (17) we may reason that for an overcomplete wavelet expansion and no feature merging, the size of feature m of $\mathbf{a}_j$ is

$$\begin{aligned}\mathrm{FS}_m(\mathbf{a}_j) &= \mathrm{FS}_m(\mathbf{a}_{j+1}) + 2^{J-j-1}(L-1) \\ &= \mathrm{FS}_m(\mathbf{a}_J) + (L-1)\left(2^{J-j} - 1\right)\end{aligned} \tag{18}$$

The number of non-zero coefficients in $\mathbf{a}_j$ arising from feature m that are not entirely or partially due to edge effects, and are therefore still guaranteed to be described by an R^{th}-order polynomial, is

$$\mathrm{PC}_m^R(\mathbf{a}_j) = \max\left(\mathrm{PC}_m^R(\mathbf{a}_{j+1}) - (L_j - 1),\ 0\right) \tag{19}$$

This means that the number of non-zero elements in feature m of $\mathbf{d}_j$ can be expressed as

$$||\mathbf{d}_j||_{0,m} = \mathrm{FS}_m(\mathbf{a}_{j+1}) + (L_j - 1) - \mathrm{PC}_m^R(\mathbf{a}_j) \tag{20}$$

To satisfy (3) and ensure that sparseness is maintained or improved we require that the ratio of number of non-zero coefficients to total coefficients used to represent $\mathbf{a}_J$ does not increase at scale $J - \mathcal{J}$. For an overcomplete wavelet expansion this implies that the following must be true:

$$\sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_{J-\mathcal{J}}) + \sum_{m=1}^{M} \sum_{j=J-\mathcal{J}}^{J-1} ||\mathbf{d}_j||_{0,m} \leq (\mathcal{J}+1) \sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_J) \tag{21}$$

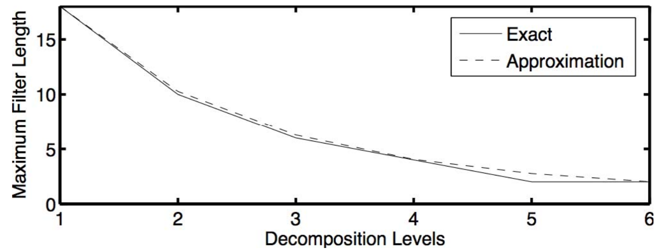

Figure 2. Exact and approximate maximum filter length for an overcomplete wavelet expansion of signal with an average feature size of 51. The exact maximum filter length (solid) was found by numerically solving (21) while the approximate maximum filter length (dashed) was found using (24).

Similar to the dyadic wavelet case, the presence of the *max* function in (19) forces (21) to be solved numerically when an exact solution is required. However, by noting again that $A \leq \max(A, 0)$ for any value A, the number of polynomial coefficients in feature m at scale j can be bound as:

$$\mathrm{PC}_m^R(\mathbf{a}_j) \geq \mathrm{PC}_m^R(\mathbf{a}_{j+1}) - (L_j - 1) \tag{22}$$

Note that (19) and (22) will be equivalent for large features (relative to L) or small values of $J - j$. Using (18), (20), and (22) the total number of non-zero detail coefficients in feature m at all decomposition levels can be shown to be:

$$\sum_{k=j}^{J-1} ||\mathbf{d}_j||_{0,m} \leq 2(L-1)(2^{J-j+1} - 2 - J + j) \tag{23}$$

Substituting (18) and (23) into (21), and solving for L, we find that maximum filter length for an overcomplete wavelet expansion is

$$L \leq \frac{\mathcal{J} \frac{1}{M} \sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_J)}{5(2^{\mathcal{J}} - 1) - 2\mathcal{J}} + 1 \tag{24}$$

Performing the same substitution, but simplifying to isolate the $\mathcal{J}$ terms, gives an expression describing the maximum number of decomposition levels for an overcomplete wavelet expansion:

$$\frac{2^{\mathcal{J}} - 1}{\mathcal{J}} \leq \frac{1}{5}\left(\frac{\sum_{m=1}^{M} \mathrm{FS}_m(\mathbf{a}_J)}{M(L-1)} + 2\right) \tag{25}$$

Just as in the case of a dyadic wavelet expansion, the maximum values for L and $\mathcal{J}$ of an overcomplete wavelet expansion depend on the average feature size of the original sparse signal. Selection of a suitable overcomplete wavelet basis for representing a sparse source signal simply requires choosing either L or $\mathcal{J}$ and using (24) or (25) to compute the other.

Figure 2 compares finding the maximum filter length by numerically solving (21) to reach an exact solution and using (24) to get an approximate value. We can see from Figure 2 that (24) is a very good approximation to the exact maximum filter length for all decomposition levels. When the maximum L is restricted to integer multiples of 2, the exact and approximate solutions match exactly for all decomposition levels. The absence of the floor and ceiling functions in the overcomplete wavelet expressions make it possible to approximate the maximum filter length very accurately with a relatively simple expression.

Similar to the dyadic wavelet case, L must be large enough to allow ϕ to have $R+1$ vanishing moments. This means that $L \geq 2(R+1)$. Substituting $L = 2(R+1)$ and $\mathcal{J} = 1$ into (24), and solving for R, we see that the maximum polynomial order must satisfy

$$R \leq \frac{1}{6M} \sum_{m=1}^{M} \mathrm{FS}_m \left(\mathbf{a}_J\right) - \frac{1}{2} \tag{26}$$

in order for an overcomplete wavelet basis to be constructed that is guaranteed to satisfy (3).

3. DYADIC OR OVERCOMPLETE WAVELET BASIS?

Having seen the criteria for selecting a wavelet basis so as to maintain or increase the sparsity of an initially sparse signal, we can now consider whether there are any benefits to using a dyadic wavelet basis rather than an overcomplete wavelet basis or vise-versa. Since the maximum values of L and $\mathcal{J}$ depend on the average feature size, it is sufficient to consider signals having only a single feature.

Let $\mathbf{s}_1$ and $\mathbf{s}_2$ denote length-1024 signals each having a single feature where $\mathrm{FS}_1(\mathbf{s}_1) = 51$, $\mathrm{FO}_1(\mathbf{s}_1) = 2$, $\mathrm{FS}_1(\mathbf{s}_2) = 10$, and $\mathrm{FO}_1(\mathbf{s}_2) = 0$. If the initial sparseness of these signals is not considered one would be tempted to select a wavelet basis with many levels of decomposition (i.e., 8-10) and a filter length long enough to allow for a sufficient number of vanishing moments. However, from Figures 1 and 2 we can see that using more than 5 decomposition levels and a dyadic wavelet basis, or 3 decomposition levels and an overcomplete wavelet basis, for a signal with an average feature size of 51 restricts the maximum filter length to less than 6. Since $\mathrm{FO}_1(\mathbf{s}_1) = 2$, we know that $L \geq 6$. It is therefore obvious that considering the sparsity of the original signal is important.

A complete view of how L and $\mathcal{J}$ influence the sparsity of a wavelet expansion is most easily obtained by considering multiple candidate bases with varying filter lengths and decomposition levels. Figure 3 shows the sparsity (coded as a grayscale) achieved when $\mathbf{s}_1$ and $\mathbf{s}_2$ are expressed in a number of wavelet bases. The dashed line in each of these *sparsity images* indicates the boundary of parameters that satisfy (3). This dashed line corresponds to the values for L and $\mathcal{J}$ that can be computed as described in Section 2.

From Figure 3 (and Figures 1 and 2), we see that when $\mathcal{J} = 1$, using an overcomplete wavelet basis allows for a slightly longer filter to be used. This is expected due to the fact that an overcomplete expansion does not include any downsampling and therefore will always be at least as sparse as the *worst* case dyadic expansion when a single decomposition level is used. Conversely, for a fixed filter length supported by both dyadic and overcomplete bases, use of a dyadic wavelet basis allows for more decomposition levels to be employed. Again, this is expected since the total number of wavelet-domain coefficients in an overcomplete wavelet expansion is $(\mathcal{J}+1)N$ rather than $2^{\mathcal{J}}N$.

Finally, comparing (16) and (26) we see that for the same average feature size an overcomplete wavelet basis is able to satisfy (3) (with $\mathcal{J} = 1$) for a slightly larger polynomial order than a dyadic wavelet basis (although the two will often be equal since R must be an integer). Depending on the application, the restricted number of decomposition levels allowed by an overcomplete wavelet may be a worthwhile trade-off to gain shift-invariance and the ability to handle slightly higher order polynomials.

4. CONCLUSIONS

Selecting a wavelet basis that can sparsely represent a piecewise polynomial signal that is composed primarily of zeros requires accounting for the form of the original signal. By conservatively considering the sparsity of the original signal, expressions describing the maximum wavelet filter length and number of decomposition levels can be determined for both dyadic and wavelet basis. When the maximum values dictated by these expressions are observed, the resulting wavelet-domain representation is guaranteed to be at least as sparse as the original signal.

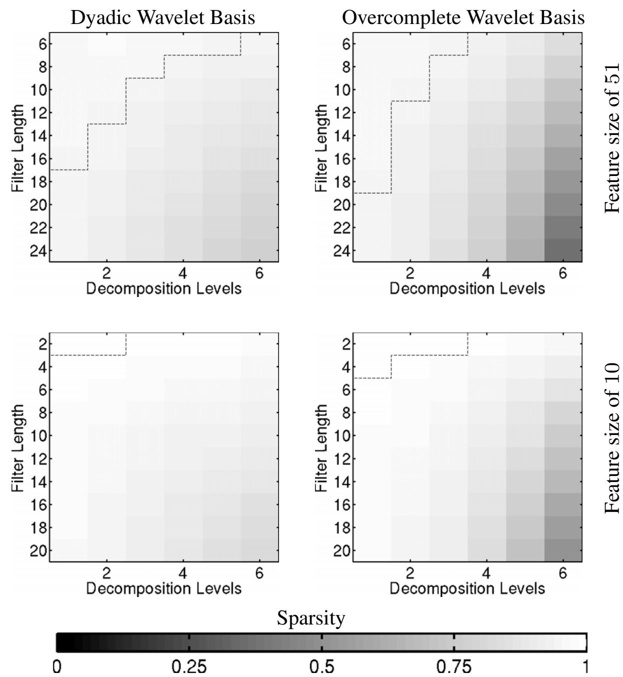

Figure 3. Sparsity achieved using various dyadic (left) and overcomplete (right) wavelet bases for a length-1024 source signal with a single feature of size of 51 (top) and 10 (bottom). The dashed line indicates the boundary of parameters that guarantee the wavelet-domain signal is at least as sparse as the original signal.

5. REFERENCES

[1] David L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data," in *Proceedings of Symposia in Applied Mathematics*, 1993, vol. 47, pp. 173–205.

[2] M. Lang, H. Guo, J. Odegard, C. Burrus, and R. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, Jan. 1996.

[3] Sandeep P. Ghael, Akbar M. Sayeed, and Richard G. Barniuk, "Improved wavelet denoising via empirical Wiener filtering," in *Proceedings of SPIE*, July 1997, pp. 389–399.

[4] Dong Wei and Alan C. Bovik, *Wavelet Denoising for Image Enhancement*, pp. 117–123, Academic Press, San Diego, 2000.

[5] Alle Meije Wink and Jos B. T .M. Roerdink, "Denoising functional MR images: A comparison of wavelet denoising and gaussian smoothing," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 374–387, Mar. 2004.

[6] D. Van De Ville, T. Blu, and M. Unser, "Integrated wavelet processing and spatial statistical testing of fMRI data," *NeuroImage*, vol. 23, no. 4, pp. 1472–1485, Dec. 2004.

[7] Martin Vetterli and Jelena Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, Upper Saddle River, 1995.