

Adaptive Wavelet Thresholding for Multichannel Signal Estimation

Ian Atkinson, Farzad Kamalabadi, Douglas L. Jones, and Minh N. Do

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

ABSTRACT

In this paper, we illustrate how a recently proposed wavelet-based estimation scheme for 2-D multichannel signals can utilize an overcomplete wavelet expansion or the BayesShrink adaptive wavelet-domain threshold to improve estimation results. The existing technique approximates the optimal estimator using a DFT and an orthonormal 2-D DWT to efficiently decorrelate the signal in both channel and space, and a wavelet-domain threshold to suppress the noise. Although this technique typically yields signal-to-noise ratio (SNR) gains of over 12 dB, results can be improved 1 to 1.5 dB by replacing the critically-sampled wavelet expansion with an overcomplete wavelet expansion. In addition, provided that the detail subbands of the original signal channels each obey a generalized Gaussian distribution, average channel SNR gains can be improved 3 dB or more using the BayesShrink adaptive wavelet-domain threshold.

Keywords: Adaptive, Estimation, Multichannel, Overcomplete, Wavelet

1. INTRODUCTION

Estimation of signals from noisy observations is an important topic with applications in areas such as communication, medical imaging, remote imaging, and astronomy. A subset of this broad problem is the estimation of multichannel signals. In this paper, we consider the estimation of 2-D multichannel signals, examples of which include hyperspectral and multispectral imagery, sequences of video frames, and series of functional magnetic resonance imaging scans. Wavelets have proven useful in many signal estimation problems.¹⁻³ In particular, wavelet-based estimation of images has received considerable attention and shown great promise in recent years.⁴⁻⁷

Unlike single-channel signal estimation, multichannel signal estimation requires estimating *each* channel of the signal. A basic approach to this problem would be to process each channel separately. One major shortcoming of this approach is that treating each channel as a separate signal fails to utilize any information that can be derived from the *interchannel* relationships present in the signal. A more sophisticated technique would attempt to remove noise by accounting not only for the individual channels, but also for the correlation among them. Exploiting this knowledge can improve estimation quality compared to treating each channel individually.

The minimum mean-square-error (MSE) estimate of a 2-D multichannel signal* from its noisy observation is achieved by the Wiener filter. The Wiener filter uses second-order statistics to optimally decorrelate the data by transforming it into the Karhunen-Loève (KL) domain. In the KL domain, signal and noise are easily discerned, allowing the noise to be attenuated while distorting the signal minimally. Despite its mathematically optimal results, the Wiener filter suffers from several drawbacks that often make it impractical for use. First, the requirement for second-order statistical knowledge of both the signal and corrupting noise makes the Wiener filter signal-dependent. Second, for large datasets calculation of the Wiener filter requires the inversion of a large matrix, which is generally a non-trivial task. Finally, for non-wide-sense-stationary (non-WSS) signals, the Wiener filter takes a form for which no efficient implementation has been found.

Recent work,^{8,9} has proposed a DFT and wavelet-based technique for estimating signals from noisy 2-D multichannel observations. This technique exploits *both* the intrachannel and interchannel signal correlations using a discrete Fourier transform (DFT) and 2-D discrete wavelet transform (DWT) to approximate the optimal

*From here forward, we will refer to a 2-D multichannel signal simply as a multichannel signal.

data decorrelation used by the Wiener filter without requiring statistical knowledge of the signal. Wavelet-domain thresholding techniques are employed to suppress noise after decorrelating the data via the DFT and 2-D DWT.

The goal of the work presented here is to illustrate the potential gains that can be realized by application of an overcomplete wavelet expansion or an adaptive wavelet-domain thresholding to the estimation of 2-D multichannel signals. Traditionally, these techniques have been applied only to estimation of 1-D or 2-D signals.^{5,7,10} Here, using the BayesShrink adaptive wavelet threshold,⁷ we demonstrate their value for the estimation of 2-D multichannel signals.

We begin in Section 2, which provides a brief review of wavelet-based estimation of single-channel 2-D signals and discusses benefits of using an overcomplete wavelet expansion and adaptive wavelet-domain threshold. Section 3 reviews the wavelet-based 2-D multichannel estimation method^{8,9} and illustrates how an overcomplete wavelet expansion or adaptive wavelet threshold may be incorporated into the technique. A simulation is provided in Section 4 that demonstrates the additional gains provided using an overcomplete wavelet expansion or adaptive wavelet threshold.

2. WAVELET-BASED ESTIMATION OF 2-D SIGNALS

Wavelets have been used extensively in a variety of signal processing applications including compression, denoising, analysis and estimation.^{2-4,11} Wavelet-based estimation often provides results comparable to Wiener filtering while remaining extremely simple as it relies only on fixed transforms and thresholding operations. In this section we review the basic process for wavelet-based estimation of 2-D signals.

Let $x[n_1, n_2]$ denote the $N \times N$ piece-wise smooth signal $s[n_1, n_2]$ that has been degraded by the zero-mean Gaussian noise $w[n_1, n_2]$. Given these noisy observations, our goal is to estimate the original signal $s[n_1, n_2]$ as closely (in terms of MSE) as possible. Let \mathbf{x} , \mathbf{s} and \mathbf{n} be the $N \times N$ matrices containing the samples of the observation, signal, and noise, respectively. Additionally, let \mathcal{W}_J and \mathcal{W}_J^{-1} denote the 2-D orthonormal forward and inverse wavelet operators with J decomposition levels, respectively.^{11,12}

In the spatial domain, \mathbf{x} , \mathbf{s} , and \mathbf{n} are related by

$$\mathbf{x} = \mathbf{s} + \mathbf{n} \quad (1)$$

Applying the 2-D DWT to (1) results in the following wavelet-domain relation of the signals.

$$\mathbf{X} = \mathcal{W}_J(\mathbf{s}) + \mathcal{W}_J(\mathbf{n}) = \mathbf{S} + \mathbf{N} \quad (2)$$

Since noise is zero-mean Gaussian with variance σ_n^2 , \mathbf{N} will likewise be zero-mean Gaussian with variance σ_n^2 , since $\mathcal{W}_J(\cdot)$ is a unitary transform. Furthermore, we know that \mathbf{S} will be compacted well since \mathbf{s} is piece-wise smooth. Together, these facts imply that \mathbf{X} will be composed of a few large-magnitude coefficients (signal + noise) and many small-magnitude coefficients (noise). We can state equivalently that the 2-D DWT approximates the Karhunen-Loève (KL) transform, an observation that has been used to replace KL transforms with the DWT in several applications.

In wavelet-based estimation, noise coefficients are zeroed and signal coefficients are retained, which is accomplished by thresholding the wavelet-domain coefficients. Using this approach, one can estimate a 2-D signal using wavelet techniques via

$$\hat{\mathbf{s}} = \mathcal{W}_J^{-1}(\eta_\tau(\mathbf{X})) \quad (3)$$

where $\eta_\tau(\cdot)$ represents the thresholding operator. Generally $\eta_\tau(\cdot)$ comes in two flavors, a *hard* threshold or a *soft* threshold. In this paper, we will use the soft threshold, which takes the following form

$$\eta_\tau(k) = \begin{cases} k \frac{|k| - \tau}{|k|} & |k| > \tau \\ 0 & \text{else} \end{cases} \quad (4)$$

where τ is the threshold level and k is a (possibly complex) wavelet-domain coefficient. Obviously, the threshold level is critical to estimation quality as it determines which coefficients will be retained and which will be discarded. Good results, in terms of perceived quality and SNR, are typically achieved using a threshold of $\tau = 3\sigma_n$. It can be shown⁸ that this threshold will, with high probability, remove the largest noise coefficient. In practical situations where the noise variance, σ_n^2 , is unknown, it can be estimated reliably from the finest-scale wavelet coefficients^{2,3} using the robust median estimator.

2.1. Wavelet-based signal estimation using an overcomplete wavelet expansion

Despite its widespread use and celebrated results, the wavelet-based estimation scheme outlined in Section 2 is not without limitations. Most notable is that the 2-D DWT is *not* shift-invariant due to the downsampling operations employed by the DWT. The implication of this shift-variance is that an estimate from an observation and its shifted version using the standard wavelet-based estimation technique will generally produce different results. As there is no way to know which shift will yield the optimal estimate, we can use the *Algorithme á Trous*¹¹ to compute an overcomplete, undecimated, shift-invariant DWT. Estimation and denoising using an overcomplete wavelet expansion have been found to provide slightly improved results[†] that correspond to averaging the estimates of all possible shifts of the signal using a critically-sampled DWT.¹⁰

In order to preserve the signal power when transforming the signal into the wavelet-domain using an overcomplete wavelet expansion, the analysis and synthesis filters must each be scaled by a factor of $1/\sqrt{2}$. This filter scaling causes the magnitude of the coefficients at the j^{th} decomposition level to be scaled by $1/2^j$. Since the wavelet coefficients are scaled in magnitude, we must adjust the threshold accordingly as well. Therefore, if we are using a threshold of τ for critically-sampled wavelet coefficients, then the threshold applied to the coefficients of the j^{th} decomposition level of an overcomplete wavelet expansion should be

$$\tau_j = \frac{\tau}{2^j} \tag{5}$$

Neglecting this threshold scaling will result in a larger number of coefficients being zeroed by the thresholding operation than desired. It is important to realize that although we are scaling the threshold for each specific decomposition level, the effect is that of a constant threshold being applied to all subbands.

Removing the downsampling operations from the DWT means that the subbands no longer decrease in size with each decomposition level. Therefore, if the original signal is $N \times N$, each subband of an overcomplete expansion will likewise be $N \times N$. When several levels of decomposition are desired for even a moderately-sized dataset, an overcomplete expansion may be impractical due to computational and/or storage issues.

2.2. BayesShrink adaptive wavelet threshold

A drawback of wavelet-based estimation, and even its overcomplete version, is that the same threshold is applied to all subbands. While this fixed threshold may be ideal for one or more subbands, it will likely be non-ideal for at least one subband. An obvious improvement to using a single threshold is to adapt the threshold for a given subband based on the data and/or some *a priori* information. Recently, a subband adaptive threshold termed *BayesShrink* was proposed⁷ for 2-D signals with detail subband coefficients having a *generalized Gaussian* (GG) distribution. While this method only slightly increases the computational complexity[‡] of the estimation process, it has been found to provide significant improvements over both standard and overcomplete wavelet-based estimation.

It has been experimentally shown and used in many applications^{7,12,13} that for a large class of images, the wavelet coefficients of the *detail* subbands (HH, HL, LH) obey a GG distribution for all decomposition levels. Each subband can be thought of as a random vector with elements that are independent identically distributed GG random variables. The probability density function of a GG random variable is defined as

$$f_{\alpha,\beta}(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|x|}{\alpha}\right)^\beta} \tag{6}$$

[†]Additional 1 to 2 dB SNR gain over wavelet-based estimation using a standard critically-sampled DWT.

[‡]The additional computations required per input sample grow in a log-like manner with the number of decomposition levels to a maximum value of one.

where $\Gamma(x) = \int_0^\infty e^{-z} z^{x-1} dz, x > 0$ is the gamma function. The two parameters of the GG density, α and β , control the overall form of the GG and may vary from subband to subband. Two special cases of the GG distribution worth noting are $\beta = 1$ and $\beta = 2$, which result in the Laplacian and Gaussian distributions, respectively.

It has been shown⁷ that for a range of β values between 0.5 and 4, the optimal soft threshold τ^* for a given detail subband is well approximated by a soft threshold proportional to the standard deviation of the wavelet-domain coefficients for that subband of the signal. Specifically,

$$\tau_B = \frac{\sigma_N^2}{\sigma_S} \quad (7)$$

is within 5% of τ^* where $\sigma_N^2 = \sigma_n^2$, and σ_S is the standard deviation of the wavelet coefficients in the subband of interest for the signal. Since each detail subband can have a different GG distribution and therefore a different variance, a separate threshold is calculated for *each* detail subband using only the data from that subband and the noise variance, which is constant across all subbands. This leads to a data-driven threshold that is adaptive on a *per-subband* level.

3. WAVELET-BASED ESTIMATION OF 2-D MULTICHANNEL SIGNALS

In this section, we revisit the wavelet-based multichannel estimation scheme proposed in^{8,9} and illustrate how an overcomplete wavelet expansion or adaptive wavelet threshold may be incorporated. The i^{th} channel of an observed $N \times N$ zero-mean 2-D multichannel signal with M channels can be expressed as

$$x_i[n_1, n_2] = s_i[n_1, n_2] + w_i[n_1, n_2] \quad (8)$$

for $n_1 = 0, 1, \dots, N-1, n_2 = 0, 1, \dots, N-1$. Here, $s_i[n_1, n_2]$ and $w_i[n_1, n_2]$ are the signal and noise components of the i^{th} channel, respectively. For simplicity, we will assume that N is an integer power of 2 and model the noise as zero-mean Gaussian with variance σ_n^2 and uncorrelated in both channel and space. Lexicographical ordering of the elements of $x_i[n_1, n_2], s_i[n_1, n_2],$ and $w_i[n_1, n_2]$ allows for their representation as the length- N^2 vectors $\mathbf{y}_i, \mathbf{v}_i$ and \mathbf{n}_i , respectively.

$$\mathbf{y}_i = \begin{bmatrix} x_i[0, 0] \\ x_i[1, 0] \\ \vdots \\ x_i[N-1, N-1] \end{bmatrix} = \mathbf{v}_i + \mathbf{n}_i \quad (9)$$

A complete vector representation for a multichannel signal is obtained by stacking the M channel vectors into the length N^2M vector \mathbf{y} that contains the entire multichannel signal.

$$\mathbf{y} = [\mathbf{y}_1^T \quad \mathbf{y}_2^T \quad \cdots \quad \mathbf{y}_M^T]^T = \mathbf{v} + \mathbf{n} \quad (10)$$

3.1. 2-D multichannel Wiener filter

The Wiener filter is the linear estimator which provides the minimum MSE estimate of a signal given its noisy observation. Using the vector representation of a multichannel signal, the *multichannel Wiener filter* has the familiar form

$$\mathbf{G}_w = \mathbf{R}_{\mathbf{v}\mathbf{v}} (\mathbf{R}_{\mathbf{v}\mathbf{v}} + \mathbf{R}_{\mathbf{n}\mathbf{n}})^{-1}, \quad (11)$$

where $\mathbf{R}_{\mathbf{v}\mathbf{v}}$ is the *signal correlation matrix*, which by definition¹⁴ is the $N^2M \times N^2M$ matrix

$$\mathbf{R}_{\mathbf{v}\mathbf{v}} = E [\mathbf{v}\mathbf{v}^H]. \quad (12)$$

Similarly, the *noise correlation matrix* $\mathbf{R}_{\mathbf{n}\mathbf{n}}$, is

$$\mathbf{R}_{\mathbf{n}\mathbf{n}} = E [\mathbf{n}\mathbf{n}^H] = \sigma_n^2 \mathbf{I} \quad (13)$$

where \mathbf{I} is the $N^2M \times N^2M$ identity matrix and we have used the fact that the noise is zero-mean, has variance σ_n^2 , and is uncorrelated in both channel and space.

The eigenexpansion of \mathbf{G}_w can be easily obtained as

$$\mathbf{G}_w = \mathbf{U}_v \mathbf{\Lambda}_G \mathbf{U}_v^H = \sum_{k=1}^{N_v} \mathbf{u}_{v_k} \frac{\lambda_{v_k}}{\lambda_{v_k} + \sigma_n^2} \mathbf{u}_{v_k}^H \quad (14)$$

where N_v , \mathbf{v}_k and λ_{v_k} are the rank, eigenvectors, and eigenvalues of \mathbf{G}_w , respectively, and

$$\mathbf{\Lambda}_G = \text{diag} \left[\frac{\lambda_{v_1}}{\lambda_{v_1} + \sigma_n^2}, \dots, \frac{\lambda_{v_{N_v}}}{\lambda_{v_{N_v}} + \sigma_n^2} \right]. \quad (15)$$

The matrix \mathbf{U}_v^H optimally decorrelates the signal in both channel and space and is therefore termed the *channel-spatial KL transform*. It can be seen from (14) that the Wiener filter creates an estimate by first decorrelating the signal using \mathbf{U}_v^H , next, weighting the decorrelated coefficients with the optimal weighting values $\mathbf{\Lambda}_G$, and finally recorrelating the signal using \mathbf{U}_v . Despite its optimality in terms of performance, the Wiener filter is often impractical or undesirable for the reasons stated earlier.

3.2. Near-optimal wavelet-based multichannel estimator

In general, the correlation between any two samples of a multichannel signal depends on both the spatial positions of the samples and the channels involved. That is

$$E [\mathbf{v}_i(l) \mathbf{v}_j(k)^H] = f(i, j, k, l) \quad (16)$$

where $\mathbf{v}_i(k)$ denotes the k^{th} sample of the i^{th} channel. Accurate modeling of this four-parameter function is difficult. Therefore, we will ignore the intrachannel correlation changes and assume that the correlation between the i^{th} and j^{th} channels can be adequately represented by a single scalar *coherence coefficient* c_{ij} that depends only on channel separation.

$$f(i, j, k, l) = c_{ij} \quad (17)$$

where $c_{ij} = \alpha^{|i-j|}$ for $0 \leq i, j \leq M-1$, and $\alpha \in [0, 1]$.

Recalling the definition of \mathbf{v} , we can reason that the signal correlation matrix (12) can be equivalently represented with an $M \times M$ *block* matrix of the $N^2 \times N^2$ channel cross-correlation matrices

$$\mathbf{R}_{\mathbf{v}_i \mathbf{v}_j} = E [\mathbf{v}_i \mathbf{v}_j^H] = c_{ij} \mathbf{R}_{\mathbf{v}}. \quad (18)$$

This allows the signal correlation matrix to be expressed as

$$\mathbf{R}_{\mathbf{v}\mathbf{v}} = \begin{bmatrix} c_{11} \mathbf{R}_{\mathbf{v}} & c_{12} \mathbf{R}_{\mathbf{v}} & \cdots & c_{1M} \mathbf{R}_{\mathbf{v}} \\ c_{21} \mathbf{R}_{\mathbf{v}} & c_{22} \mathbf{R}_{\mathbf{v}} & & \\ \vdots & & \ddots & \\ c_{M1} \mathbf{R}_{\mathbf{v}} & & & c_{MM} \mathbf{R}_{\mathbf{v}} \end{bmatrix} = \mathbf{C} \otimes \mathbf{R}_{\mathbf{v}} \quad (19)$$

where $\mathbf{R}_{\mathbf{v}}$ is the $N^2 \times N^2$ source signal correlation matrix, \otimes denotes a Kronecker product, and \mathbf{C} is the $M \times M$ *channel coherence matrix* with its $(i^{\text{th}}, j^{\text{th}})$ element being c_{ij} .

Eigenexpansion of \mathbf{C} and $\mathbf{R}_{\mathbf{v}}$ gives

$$\mathbf{C} = \mathbf{U}_C \mathbf{\Lambda}_C \mathbf{U}_C^H \quad (20)$$

$$\mathbf{R}_{\mathbf{v}} = \mathbf{U}_{\mathbf{R}_{\mathbf{v}}} \mathbf{\Lambda}_{\mathbf{R}_{\mathbf{v}}} \mathbf{U}_{\mathbf{R}_{\mathbf{v}}}^H \quad (21)$$

\mathbf{U}_C^H and $\mathbf{U}_{\mathbf{R}_{\mathbf{v}}}^H$ are termed the *channel KL transform* and the *spatial KL transform* as they decorrelate the data optimally in channel and space, respectively. Applying a basic property of Kronecker products,¹⁵ it can be shown that

$$\mathbf{U}_v^H = \mathbf{U}_C^H \otimes \mathbf{U}_{\mathbf{R}_{\mathbf{v}}}^H \quad (22)$$

Algorithm 1 Wavelet-Based Near-Optimal 2-D Multichannel Signal Estimation

- 1: Decorrelate the data across channels using the normalized DFT $\frac{\mathbf{F}_M}{\sqrt{M}}$.
 - 2: Decorrelate each resulting channel spatially using an orthonormal 2-D DWT $\mathcal{W}_J(\cdot)$.
 - 3: Threshold each decorrelated channel using a threshold operator $\eta_\tau(\cdot)$.
 - 4: Recorrelate each thresholded channel spatially using an orthonormal 2-D IDWT $\mathcal{W}_J^{-1}(\cdot)$.
 - 5: Recorrelate the data across channels using the inverse normalized DFT $\frac{\mathbf{F}_M^H}{\sqrt{M}}$.
-

The significance of (22) is that the channel-spatial KL transform, which decorrelates the signal optimally in both channel and space, can be decomposed into a channel KL transform, which decorrelates the data in channel, and a spatial KL transform, which decorrelates the data in space.

Since \mathbf{C} is an $M \times M$ Toeplitz symmetric matrix, it can be viewed alternatively as the correlation matrix of a length- M WSS sequence. As M increases, the basis functions of the KL transform for a WSS sequence tend toward complex exponentials.¹⁴ Therefore, the DFT will asymptotically diagonalize \mathbf{C} . With this observation, we may replace the channel KL transform that decorrelates the data optimally in channel with the $M \times M$ normalized[§] DFT matrix $\frac{\mathbf{F}_M}{\sqrt{M}}$ that asymptotically decorrelates the data optimally in channel.

The role of the spatial KL transform is to decorrelate the signal optimally in space using second-order signal and noise statistics. The approximation to the spatial KL transform should provide near-optimal spatial decorrelation of the signal without requiring knowledge of signal statistics. Section 2 discussed wavelets and their utilization in single channel 2-D signal estimation by using the 2-D DWT to approximately decorrelate the noisy observations. An important point brought forth was that for a wide class of signals a wavelet basis forms an approximate KL basis; that is, the DWT decorrelates the data effectively. Since the DWT approximates a KL transform for a wide class of signals and is a predefined transform well-suited to efficient implementation, we will approximate the spatial KL transform with a 2-D DWT.

A final issue that must be addressed in order to create the near-optimal estimator is how to handle the weighting values. From our model for channel coherence (16), we can assume that the DFT will asymptotically decorrelate the signal in channel leaving no residual channel correlation. Therefore, after decorrelating the signal in channel, the *resulting* channels[¶] can be processed independently without degrading the estimation quality. These resulting channels can be estimated using the wavelet-based techniques described in Section 2 by approximating the ideal weighting values with a wavelet-domain threshold.

Using these approximations, the entire estimation process can be expressed as the five-step algorithm outlined in Algorithm 1. Although we have assumed a zero-mean signal for the optimal estimator (11), it can be shown[§] that this estimation scheme is equally well-suited to non-zero-mean signals as well, with no modification.

3.3. Multichannel signal estimation using an overcomplete wavelet expansion

As discussed in Section 2.1, wavelet-based image estimation using an overcomplete wavelet expansion can improve estimation quality over using an orthonormal wavelet expansion. Since the overcomplete wavelet expansion is a straightforward extension of the critically-sampled wavelet expansion, we can replace the 2-D DWT in Algorithm 1 with an overcomplete wavelet expansion and enjoy similar benefits in multichannel signal estimation.

An overcomplete 2-D wavelet expansion can be calculated using a 2-D version of the *Algorithme á Trous*, an efficient implementation of which requires $8 \cdot L \cdot J$ operations per input sample.¹¹ ^{||} Here, L is the length of the filters used to compute the 2-D DWT; the length of all filters will be the same since we only consider orthogonal wavelets. In contrast to the overcomplete 2-D DWT, the standard 2-D DWT requires only $\frac{8}{3} \cdot L$ operations per input sample.¹¹ Furthermore, thresholding an overcomplete expansion requires thresholding $3 \cdot J \cdot N^2$ samples,

[§] \mathbf{F}_M is the standard $M \times M$ DFT matrix. The normalization by \sqrt{M} makes $\frac{\mathbf{F}_M}{\sqrt{M}}$ a unitary transform.

[¶]The resulting channels refers to the channels *after* channel decorrelation but *before* spatial decorrelation.

^{||}An efficient implementation of the *Algorithme á Trous* accounts for the zeros in the upsampled filters to reduce the total required computations.

as opposed to only $N^2 \left(1 - \frac{1}{2^{2J}}\right)$ samples for a critically sampled 2-D DWT.** Therefore, neglecting the channel decorrelation/redecorrelation, the total operational cost of wavelet-based multichannel signal estimation using an overcomplete wavelet expansion is

$$N^2 M (16 \cdot L \cdot J + 3 \cdot J)$$

compared to only

$$N^2 M \left(\frac{16}{3} L + 1 - \frac{1}{2^{2J}} \right)$$

for a critically sampled wavelet expansion. This drastic increase in computations may limit the practicality of using an overcomplete wavelet expansion for all but small multichannel datasets.

3.4. Multichannel signal estimation using an adaptive wavelet threshold

In this section, we consider the use of the BayesShrink adaptive thresholding scheme introduced in Section 2.2 in multichannel signal estimation. The BayesShrink method of threshold selection assumes that the coefficients of each detail subband obey a GG distribution. Therefore, before using the BayesShrink adaptive threshold, we must verify that the subbands of the decorrelated channels of a multichannel signal will indeed obey a GG distribution.

Since the channel and spatial decorrelation are both linear operations and are applied to different dimensions of the signal, they commute. That is, decorrelating in channel and then in space is equivalent to decorrelating in space and then in channel. This means that the subbands of the i^{th} fully decorrelated channel are simply a linear combination of the subbands of the original channels. Therefore, since a linear combination of GG random variables is also a GG random variable, we can conclude that if all the detail subbands of all the original channels obey a GG distribution, then so will the detail subbands of all of the fully decorrelated channels, allowing the BayesShrink adaptive threshold to be used as the thresholding operation in Algorithm 1.

It can be shown that the *added* computational cost of using the BayesShrink adaptive threshold rather than a standard soft threshold is

$$1 - \frac{1}{2^{2J}}$$

where J is the number of decomposition levels used.⁸ We can see that the additional computational cost of using the BayesShrink adaptive threshold is minimal, having a worst-case scenario of only one additional operation per input sample.

4. SIMULATION EXAMPLES

In this section, we present simulation examples that demonstrate the additional benefits which may be realized using an overcomplete wavelet expansion or adaptive wavelet-domain threshold in multichannel signal estimation. The dataset, *Desert*, is a multispectral dataset having 48 channels all within the visible portion of the spectrum. Each channel has spatial dimensions of 281×251 and was therefore zero-padded to 512×512 prior to processing. Noisy observations were generated by degrading the signal with additive zero-mean Gaussian noise with variance σ_n^2 . Estimation was performed using the following techniques.

1. *Soft*: Estimation scheme outlined in Algorithm 1 using a critically sampled 2-D DWT and a soft wavelet-domain threshold of $3\sigma_n$.
2. *Soft-Overcomplete*: Estimation scheme outlined in Algorithm 1 using an overcomplete 2-D DWT and a soft wavelet-domain threshold of $3\sigma_n$.
3. *BayesShrink*: Estimation scheme outlined in Algorithm 1 using a critically sampled 2-D DWT and the BayesShrink adaptive wavelet-domain threshold.

**Here we assume that the threshold is not applied to the coarse approximation subband LL_J .

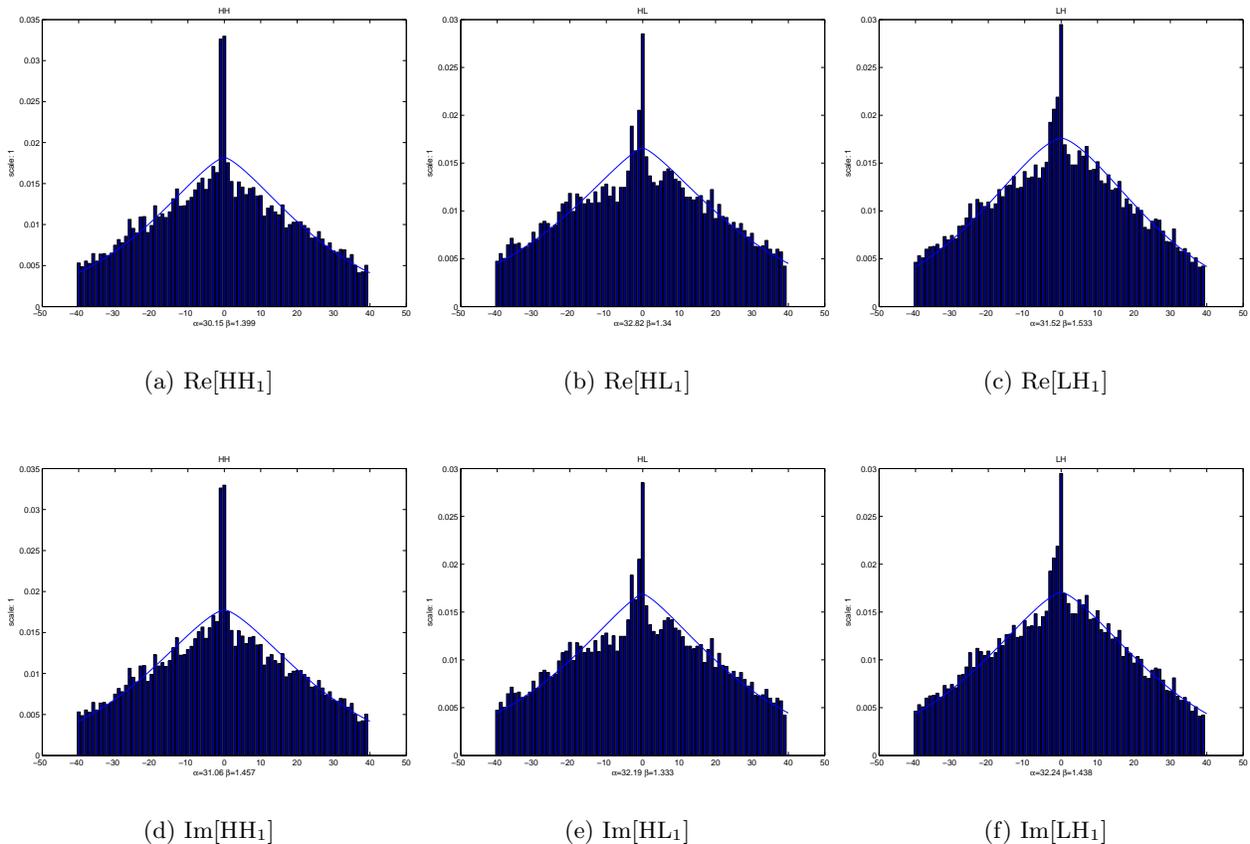


Figure 1. First level of decomposition real and imaginary detail subband coefficient distributions for channel 35 of multispectral dataset *Desert* *after* channel decorrelation. Each has been fit with a GG distribution having the parameters α and β shown. It can be seen that both the real and imaginary components are reasonably approximated by a GG distribution as assumed by the BayesShrink adaptive threshold.

In addition, as baseline comparison, the MATLAB routine `wiener2`, with default parameters, was used to estimate the multichannel signal by independently processing each channel of the signal. All the wavelet-based estimators used a Daubechies length 4 wavelet and three levels of decomposition.

Prior to using the BayesShrink estimator, we must verify that the detail subbands of the channel decorrelated channels satisfy the GG distribution requirement of the BayesShrink adaptive threshold. Histograms of the real and imaginary detail subbands for the first level decomposition of resulting channel 35 are shown in Fig. 1 with the corresponding best fit GG density. It can be seen that both the real and imaginary components do indeed satisfy a GG distribution as assumed by the BayesShrink adaptive threshold. The remaining channels and subbands have similar distributions to those shown, also satisfying the BayesShrink GG distribution requirement.

Channel 35 of the original, observed, and estimated signals is shown in Fig. 2 for $\sigma_n^2 = 1000^2$. As can be determined both visually and from the average and select channel SNR values listed in Table 1, the Soft, Soft-Overcomplete and BayesShrink estimators significantly improve on the noisy observation and `wiener2`. Using the Soft estimator, the average channel SNR increases by 10.73 dB, while the `wiener2` estimator delivers an average gain of only 6.85 dB. Visually, we can easily see that the results provided by the BayesShrink estimator are substantially better than all other estimators for this noise variance. While all wavelet-based estimators recover the structure of the signal reasonably, the BayesShrink estimate is the sharpest and contains minor details not found in the other estimates. Both the Soft-Overcomplete and BayesShrink estimators improve upon the Soft estimator as expected. Using the overcomplete wavelet expansion, the average channel SNR is

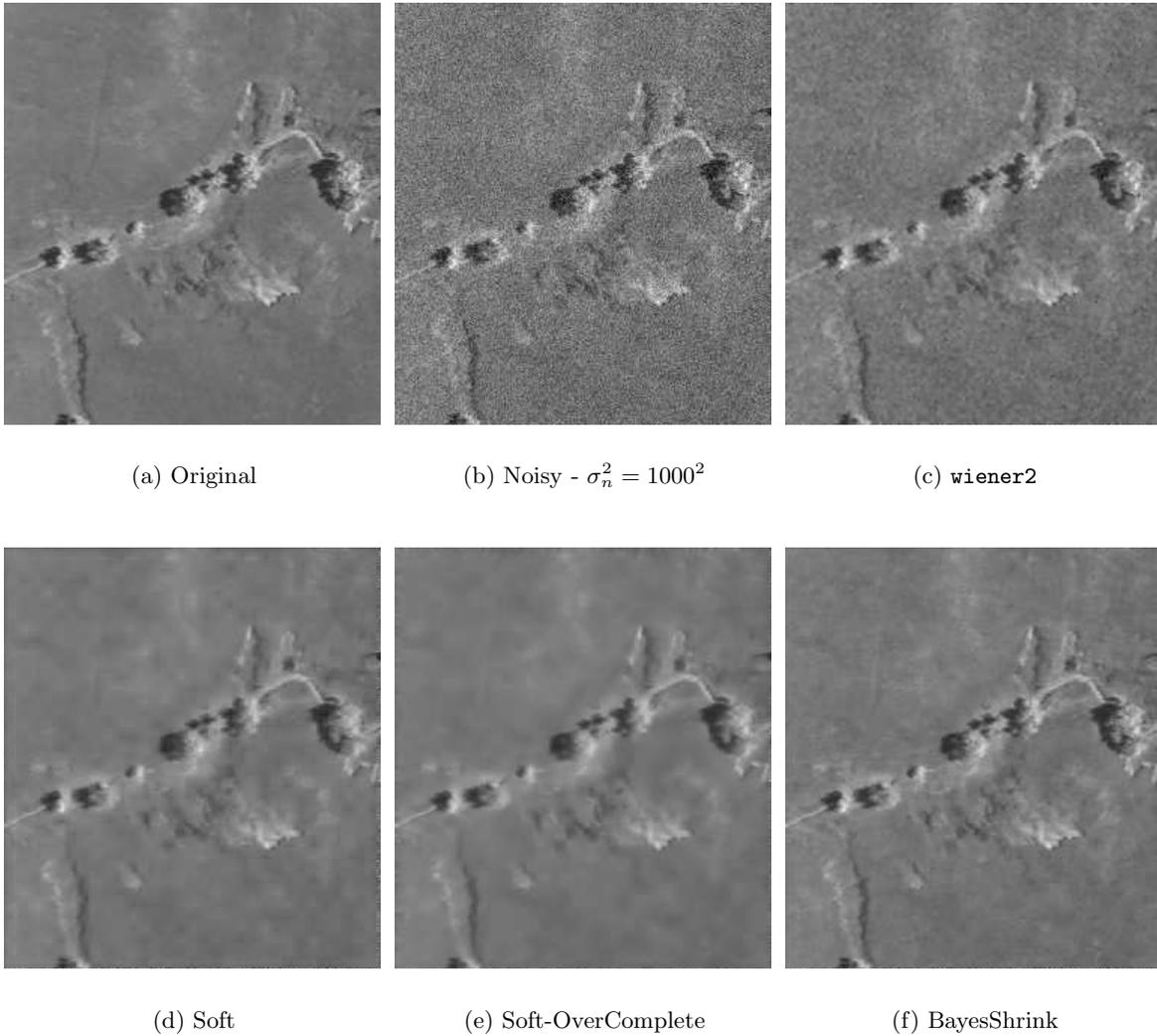


Figure 2. Noisy and estimated versions of channel 35 of the multispectral dataset Desert. BayesShrink produces the best estimate both in terms of SNR and visual quality. The BayesShrink and Soft-Overcomplete estimators both improve estimation quality compared to the Soft estimator with BayesShrink best matching minor channel features and producing the sharpest estimate. SNR values are available in Table 1.

increased by 1.28 dB whereas the BayesShrink adaptive threshold increased the SNR from 22.42 dB by 1.65 dB to 24.07 dB.

4.1. Effect of noise variance on estimator performance

Having demonstrated the additional gains an overcomplete wavelet expansion or adaptive wavelet-domain threshold provide for fixed noise variance of $\sigma_n^2 = 1000^2$, we will now show that similar performance is achieved across a wide range of corrupting noise variance values. As noise variance increases, the maximum possible SNR gain increases as well. A robust estimation scheme will yield a larger SNR gain in the presence of a higher noise level. To quantify performance as a function of noise variance, observations were generated for various noise levels ranging from low-noise ($\sigma_n^2 = 100^2$) to high-noise ($\sigma_n^2 = 2500^2$). At each noise level, the desired signal was estimated from the noisy observation using the Soft, Soft-Overcomplete, BayesShrink, and `wiener2` estimation schemes.

Channel	Noisy	wiener2	Soft	Soft-Overcomplete	BayesShrink
Ave of All	11.69	18.54	22.42	23.70	24.07
10	10.50	17.50	22.04	23.34	23.70
20	12.53	19.24	22.67	23.77	25.42
30	14.05	20.32	22.01	22.85	24.89
35	13.74	20.40	23.31	24.37	26.14
40	13.21	20.26	24.48	26.01	26.04

Table 1. Average and selected channel SNR values in decibels for observed and estimated versions of the multispectral dataset Desert. On average, the Soft estimator yields an additional 3.88 dB of SNR over the `wiener2` estimator, while Soft-Overcomplete and BayesShrink provide 5.16 dB and 5.53 dB, respectively.

σ_n^2	Noisy	wiener2	Soft	Soft-Overcomplete	BayesShrink
100^2	31.67	33.29	32.76	33.68	36.94
500^2	17.69	23.34	25.51	26.61	28.27
1000^2	11.67	18.54	22.43	23.70	24.08
1500^2	8.14	15.47	20.38	21.77	21.52
2000^2	5.64	13.20	18.79	20.28	19.62
2500^2	3.70	11.39	17.45	19.01	18.09

Table 2. Average channel SNR values in decibels for observed and estimated versions of the multispectral dataset Desert at various noise variances. The best-performing estimator is BayesShrink for low to moderate noise levels ($\sigma_n^2 < 1250^2$) and Soft-Overcomplete for high noise levels ($\sigma_n^2 > 1250^2$). While the Soft estimator yields higher SNR than `wiener2`, it cannot match the performance of the more sophisticated Soft-Overcomplete and BayesShrink estimators.

The average channel SNR gain for each estimation method at the various noise levels is shown in Fig. 3 and select SNR values are included in Table 2. We can see that all three wavelet-based estimators yield SNR gains superior to `wiener2`, improving the SNR by several decibels at moderate to high noise levels. For all noise variances tested, both the Soft-Overcomplete and BayesShrink estimators produce estimates with higher SNR than the Soft estimator.

For low to moderate noise levels, BayesShrink provides the highest SNR gain of all estimators tested. We can see from Fig. 3 that the additional gain it yields falls off with noise level, becoming minimal for $\sigma_n^2 > 2000^2$. The reason for this is that the threshold levels used by the BayesShrink and Soft estimators are both functions of the noise standard deviation that therefore increase with noise level. Eventually, both thresholds will be larger than all wavelet-domain coefficients, and both estimates will become identically zero. At low noise levels, $\sigma_n^2 < 500^2$, the additional gain from using the BayesShrink estimator is particularly impressive. At $\sigma_n^2 = 500^2$ SNR gain improves from 7.82 dB by 35% to 10.58 dB using the BayesShrink estimator instead of the Soft estimator. At very low noise levels, performance increases of more than 100% are achieved by the BayesShrink estimator. High noise levels show a comparatively little gain of less than 5% over the Soft estimator.

In contrast to those of the BayesShrink estimator, the additional gains provided by the Soft-Overcomplete estimator are more constant across the noise variances tested. We can see from Fig. 3 that using an overcomplete wavelet expansion results in an additional 1 to 1.5 dB of SNR compared to using a critically sampled 2-D DWT. These additional gains, although desirable, require the use of a computationally expensive overcomplete wavelet expansion, which may be impractical for many applications due to the generally large size of multichannel datasets. Furthermore, the Soft-Overcomplete estimator only outperforms the BayesShrink estimator for high noise levels, at which the estimate will generally be of poor quality regardless of the estimation method.

4.2. Effect of channel count on estimator performance

For a finite, but large number of channels, the DFT only provides approximate channel decorrelation. Therefore, it is of interest to quantify the effect the number of signal channels has on estimator performance. We accomplish this by considering all possible channel counts ($M = 2, 3, \dots, 48$) at a fixed noise variance of $\sigma_n^2 = 1000^2$. For each

Channel Count	Noisy	wiener2	Soft	Soft-Overcomplete	BayesShrink
10	11.27	18.22	18.56	19.97	21.89
30	11.66	18.55	20.27	21.44	23.49
40	11.59	18.48	20.55	21.79	23.77
48	11.66	18.54	21.03	22.22	24.08

Table 3. Average channel SNR values in decibels for observed and estimated versions of the multispectral dataset Desert using various numbers of channels at a fixed noise variance of $\sigma_n^2 = 1000^2$. The BayesShrink estimator performs best of all methods followed by the Soft-Overcomplete, the Soft, and finally the `wiener2` estimator.

channel count, M -channel desired and observed signals were created from the original 48-channel dataset, and the observed signal was then processed with the Soft, Soft-Overcomplete, BayesShrink, and `wiener2` estimators.

The results are summarized in Fig. 4, which plots the average SNR gain as a function of channel count for the various estimates. Select SNR values are listed in Table 3 for each of the signals. Naturally, the number of channels does not greatly influence the results of the `wiener2` estimate as it neglects all interchannel correlation and processes each channel as an independent signal. The wavelet-based estimators, however, exploit this information, and as a result, additional channels lead to higher SNR. Although the overall SNR increases as channels are added, the additional gain provided from an overcomplete wavelet expansion or the BayesShrink adaptive threshold remains approximately constant. This is since after decorrelating the channels via the DFT, we assume there is no residual channel correlation and process the resulting channels independently. Any additional SNR provided by BayesShrink or Soft-Overcomplete is therefore due to the additional SNR that the respective estimator yields on each decorrelated channel and not the number of channels.

5. CONCLUSION

We have demonstrated the value of an overcomplete wavelet expansion and adaptive wavelet-domain thresholding to multichannel signal estimation. Optimal estimation of a 2-D multichannel signal relies on the ability to decorrelate the signal in both channel and space, which requires second-order signal statistics to be known. A recently proposed wavelet-based estimation technique for 2-D multichannel signals removes this statistical requirement by combining a DFT and an orthonormal 2-D DWT to approximately decorrelate the signal in channel and space. These approximately-decorrelated coefficients are then denoised using a wavelet-domain thresholding operation, such as a hard or soft threshold.

Since the proposed multichannel estimation algorithm utilizes a 2-D DWT, estimation results can be improved using an overcomplete wavelet expansion. SNR improvements of 1 to 1.5 dB were demonstrated for a multispectral dataset using an overcomplete wavelet expansion versus an orthonormal wavelet expansion. These gains were fairly constant across various input SNR levels and number of signal channels.

Using the BayesShrink adaptive wavelet threshold to denoise the decorrelated coefficients was shown to improve the estimate SNR of the multispectral dataset by more than 4 dB for low noise levels. The additional gain provided falls off with increased noise level, but remains approximately constant regardless of the number of channels. It was found that the BayesShrink adaptive threshold is applicable to multichannel signal estimation provided that the detail subbands of the original signal obey a generalized Gaussian distribution.

Although it was not specifically explored in this paper, one would believe that a combination of an overcomplete wavelet expansion and an adaptive wavelet-domain threshold would yield superior results than either alone. While this would require using the computationally costly overcomplete expansion, the potentially improved results may be worth these added costs. This is particularly true in applications such as medical imaging where estimation quality is of extreme importance.

REFERENCES

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

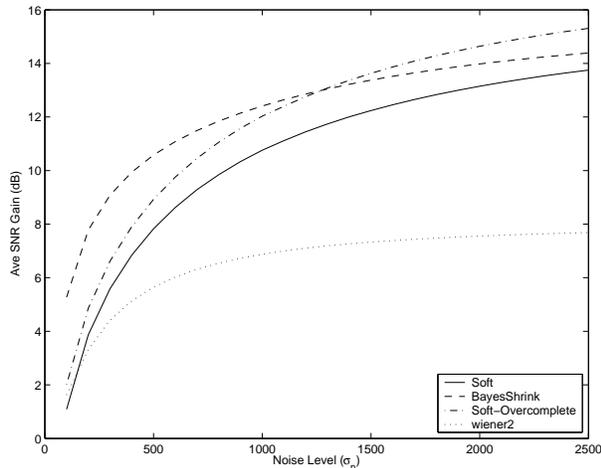


Figure 3. Comparison of Soft, Soft-Overcomplete, BayesShrink, and `wiener2` estimates of the multispectral dataset Desert for various noise powers. Graph is of average SNR gain for all channels in the dataset. The BayesShrink estimator provides the best results for input noise power up to $\sigma_n^2 \approx 1250^2$, at which point the Soft-Overcomplete estimator performs slightly better. Both methods outperform the Soft estimator for all noise powers.

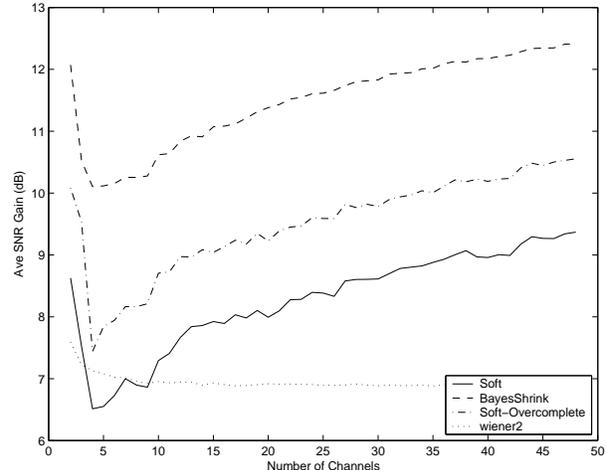


Figure 4. Comparison of Soft, Soft-Overcomplete, BayesShrink and `wiener2` estimates of the multispectral dataset Desert using various numbers of channels at a fixed noise variance of $\sigma_n^2 = 1000^2$. Graph is of average SNR gain for all tested channels in the dataset. The BayesShrink estimator performs best of all methods followed by the Soft-Overcomplete estimator and finally the Soft estimator.

2. D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory* **41**, pp. 613–627, May 1995.
3. S. G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1999.
4. D. Wei and A. C. Bovik, *Wavelet Denoising for Image Enhancement*, pp. 117–123. Academic Press, San Diego, 2000.
5. S. G. Chang, B. Yu, and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” *IEEE Transactions on Image Processing* **9**, pp. 1522–1531, September 2000.
6. S. G. Chang, B. Yu, and M. Vetterli, “Image denoising via lossy compression and wavelet thresholding,” in *Proceedings of IEEE Int. Conf. Image Processing*, **1**, pp. 604–607, November 1997.
7. S. G. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Transactions on Image Processing* **9**, pp. 1532–1546, September 2000.
8. I. Atkinson, “Wavelet-based near-optimal estimation of multichannel two-dimensional signals,” Master’s thesis, University of Illinois at Urbana-Champaign, 2003.
9. I. Atkinson, F. Kamalabadi, S. Mohan, and D. L. Jones, “Wavelet-based 2-D multichannel signal estimation,” in *Proceedings of IEEE Int. Conf. Image Processing*, 2003. in press.
10. M. Lang, H. Guo, J. Odegard, C. Burrus, and R. Wells, “Noise reduction using an undecimated discrete wavelet transform,” *IEEE Signal Processing Letters* **3**, pp. 10–12, January 1996.
11. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, Upper Saddle River, 1995.
12. S. G. Mallat, “A theory for multiresolutional signal decomposition: The wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, pp. 674–693, July 1989.
13. M. N. Do and M. Vetterli, “Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance,” *IEEE Transactions on Image Processing* **11**, pp. 146–158, February 2002.
14. H. Stark and J. Woods, *Probability and Random Processes with Applications to Signal Processing*, Prentice Hall, Upper Saddle River, 3rd ed., 2002.
15. J. W. Brewer, “Kronecker products and matrix calculus in system theory,” *IEEE Transactions on Circuits and Systems* **CAS-25**, pp. 772–781, September 1978.