Wavelet-Based Deconvolution for Ill-Conditioned Systems

Ramesh Neelamani, Hyeokho Choi, and Richard Baraniuk [*]

Corresponding author: Ramesh Neelamani

Department of Electrical and Computer Engineering, Rice University

6100 South Main Street, Houston, TX 77005–1892

Email: {neelsh,choi,richb}@ece.rice.edu, Web: www.dsp.rice.edu, Fax: 713 348 6196

EDICS Numbers: 2-MWAV (Wavelets), 2-REST (Restoration)

Submitted to the IEEE Transactions on Image Processing, February 2000

Abstract— We propose a hybrid approach to wavelet-based deconvolution that comprises Fourier-domain system inversion followed by wavelet-domain noise suppression. In contrast to other wavelet-based deconvolution approaches, the algorithm employs a *regularized inverse filter*, which allows it to operate even when the system is non-invertible. Using a mean-square-error (MSE) metric, we strike an optimal balance between Fourier-domain regularization (matched to the convolution operator) and wavelet-domain regularization (matched to the signal/image). Theoretical analysis reveals that the optimal balance is determined by the Fourier-domain operator structure and the economics of the wavelet-domain signal representation. The resulting algorithm is fast ($O(N \log N)$ complexity for signals/images of N samples) and is well-suited to data with spatially-localized phenomena such as edges and ridges. In addition to enjoying asymptotically optimal rates of error decay for certain systems, the algorithm also achieves excellent performance at fixed data lengths. In real data experiments, the algorithm outperforms the conventional time-invariant Wiener filter and other wavelet-based image restoration algorithms in terms of both MSE performance and visual quality.

[*]This work was supported by the National Science Foundation grant CCR–99–73188, DARPA/AFOSR grant F49620–97–1–0513, ONR grant N00014–99–1–0219, Texas Instruments, and the Rice Consortium for Computational Seismic/Signal Interpretation.

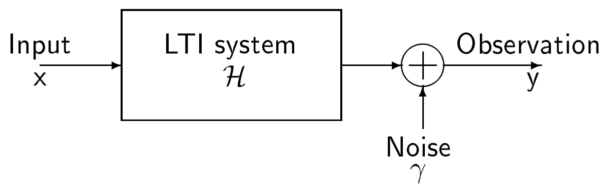


Figure 1: *Convolution model setup. The desired data x is first corrupted by convolution with the linear time-invariant (LTI) system $\mathcal{H}$, and then by the additive white Gaussian (AWGN) noise γ before being observed as y.*

1 Introduction

Deconvolution is a recurring theme in a wide variety of signal and image processing problems, from channel equalization [1] to image restoration [2]. For example, practical satellite images are often blurred due to limitations such as aperture effects of the camera, camera motion, or atmospheric turbulence. Deconvolution becomes necessary if we wish a crisp, deblurred image for viewing or further processing.

1.1 Problem description

In this paper, we treat the classical deconvolution scenario (see Figure 1). Two degradations corrupt our observation y of the desired data x: convolution with an linear time-invariant LTI system $\mathcal{H}$ having impulse response h and additive white Gaussian (AWGN) noise γ of variance σ^2 . In 1-d, we have

$$y(t_n) := (x \circledast h)(t_n) + \gamma(n), \qquad n = 0, \ldots, N-1, \tag{1}$$

where t_n denotes the discrete sample points; and $\circledast$ denotes circular convolution, which is assumed for simplicity but without loss of generality. Given y and h, we seek to estimate x.

In the discrete Fourier transform (DFT) domain, we equivalently have

$$Y(f_n) = H(f_n)\, X(f_n) + \Gamma(f_n), \qquad n = 0, \ldots, N-1, \tag{2}$$

with Y, H, X and Γ the respective length-N DFTs and $f_n := \frac{2\pi n}{N}$ the normalized DFT frequencies. The problem formulation trivially extends to multidimensional data.

If the system frequency response H has no zeros, then an unbiased estimate of x can be obtained through

$$\widetilde{X}(f_n) \quad := \quad H^{-1}(f_n)\, Y(f_n)$$

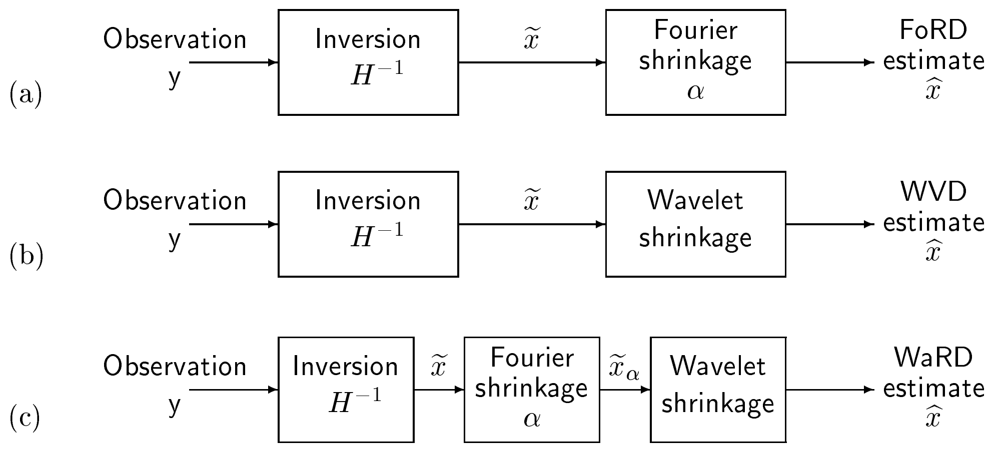


Figure 2: *Three different deconvolution strategies. (a) Formally, Fourier-domain regularized deconvolution (FoRD) estimates the signal in the presence of the noise colored by H^{-1} using Fourier-domain shrinkage. The amount of Fourier-domain shrinkage is controlled by the regularization parameter α. Wiener deconvolution is a special case of FoRD that shrinks according to the signal-to-noise-ratio (SNR) at each frequency ($\alpha = 1$). (b) Wavelet-vaguelette deconvolution (WVD) estimates the signal in the presence of the noise colored by H^{-1} using wavelet-domain shrinkage. (c) Our proposed wavelet-based regularized deconvolution (WaRD) exploits both the FoRD and WVD concepts to minimize the distortion of spatially localized features in the signal.*

$$= \quad X(f_n) + H^{-1}(f_n)\,\Gamma(f_n). \tag{3}$$

However, if the system is ill-conditioned, i.e., if $H(f_n)$ is small at any f_n, then the noise is enormously amplified during inversion to yield an extremely noisy, useless estimate.

1.2 Fourier-domain regularized deconvolution (FoRD)

Noise amplification can be alleviated by using an approximate, regularized inverse instead of a pure inverse. Regularization aims to provide a better solution by reducing noise in exchange for some bias in the estimate [3]; regularization becomes essential in situations involving ill-conditioned systems.

The LTI Wiener deconvolution filter is a classical example of what we will term *Fourier-domain regularized deconvolution* (FoRD, see Figure 2(a)). Formally, the estimation procedure used by the LTI Wiener deconvolution filter can be understood as inverting the convolution operator using H^{-1} to obtain a noisy estimate $\widetilde{x}$ followed by shrinkage of each frequency component f_n of $\widetilde{x}$

according to the signal-to-noise-ratio (SNR) at f_n (shrink less/more when the SNR is high/low).[1]

The Fourier transform diagonalizes the convolution operator $\mathcal{H}$; hence the Fourier domain is ideally suited to represent the colored noise $H^{-1}\Gamma$ in $\widetilde{X}$ (see (3)). Consequently, the LTI Wiener deconvolution filter, which employs Fourier-domain shrinkage, can precisely identify and attenuate the noise that gets amplified during inversion of $\mathcal{H}$, thereby fully exploiting the structure of the blurring system. In fact, when the input signal can be modeled as wide-sense stationary (WSS) and Gaussian, the LTI Wiener deconvolution filter is MSE-optimal over *all* estimators.

However, the Fourier domain is not well-suited for representing many common signals and images that contain spatially localized phenomena such as edges, because the Fourier basis functions have support that extend over the entire spatial domain.[2] Scalar processing of the Fourier components, employed by FoRD, lacks spatial selectivity; consequently, important spatially localized features such as edges and ridges become distorted during FoRD. Such distortions are reflected as ringing around the localized features.

1.3 Wavelets and wavelet-vaguelette deconvolution (WVD)

The wavelet transform is an invaluable tool for dealing with signals and images containing spatially localized features such as edges and ridges. Wavelets provide economical representations for these signals and in particular those belonging to Besov spaces [4], i.e., the wavelet expansion captures most of the signal energy using a few large wavelet coefficients. This property has been leveraged into powerful, spatially adaptive, signal estimation algorithms that are based on simply shrinking the wavelet coefficients of the noisy signal [5, 6].

Motivated by the economy of wavelet representations, Donoho proposed the wavelet-vaguelette decomposition algorithm to solve a special class of linear inverse problems [7]. With a slight abuse of notation, we refer to the wavelet-vaguelette decomposition applied to deconvolution as *wavelet-vaguelette deconvolution* (WVD). In contrast to FoRD, WVD employs wavelet-domain shrinkage to estimate the signal in the presence of colored noise $H^{-1}\Gamma$ (see Figure (3) and 2(b)).

WVD exploits the economical wavelet representation of signals to effectively identify and

[1]The inversion and shrinkage is performed jointly in practice; hence FoRD and Wiener deconvolution are applicable even when the system $\mathcal{H}$ is not invertible. All through this paper, we have referred to any form of attenuation, i.e., multiplication by a number between 0 and 1, as *shrinkage*.

[2]Such signals cannot be modeled as WSS.

estimate the signal. In fact, for special classes of blurring operators such as the Radon transform, WVD exhibits asymptotically (as $N \to \infty$) near-optimal rates of error decay for a wide class of input signals [7].

However, the wavelet transform does not diagonalize the convolution operator $\mathcal{H}$. Consequently, the noise frequency components amplified during inversion of $\mathcal{H}$ corrupt many wavelet coefficients. For example, for a box-car impulse response[3] h — a common model for image blurring due to camera motion [2] — the noisy estimate $\widetilde{x}$ obtained after system inversion has infinite noise variance at all wavelet scales. Thus, even though wavelets provide an efficient input signal representation, signal estimation using scalar operations in the wavelet domain is futile and results in a zero signal as the estimate.

1.4 Wavelet-based regularized deconvolution (WaRD)

Motivated by the fact that the Fourier domain matches the convolution operator while the wavelet domain matches a large class of potential input signals, we propose an improved hybrid *wavelet-based regularized deconvolution* (WaRD) algorithm suitable for use with any ill-conditioned system. The basic idea is simple: employ the best of *both* FoRD and WVD processing (see Figure 2(c)). In this tandem processing, Fourier-domain regularization adapted to the convolution operator partially controls noise amplification. However, we use it sparingly to keep the accompanying smearing distortions to a minimum; the bulk of the noise removal and signal estimation is achieved using wavelet shrinkage. Figure 3 illustrates the superior overall visual quality and the mean square error (MSE) of the WaRD estimate as compared to the Wiener estimate.

By optimizing over an MSE metric, we will find that the optimal balance between local processing with the wavelet basis and global processing with the Fourier basis is determined by both the Fourier-domain convolution operator structure and the economy of wavelet domain signal representation.

1.5 Related work

One extreme of our Fourier/wavelet balance is to perform no Fourier-domain regularization; this is equivalent to the WVD approach of Donoho [7] and the mirror-wavelet basis approach of

[3]The $Q \times Q$-point box-car impulse response is defined as $h(n,m) = \frac{1}{Q^2}$ for $1 \leq n,m \leq Q$; and $h(n,m) = 0$ otherwise.

Figure 3: *(a) Original Cameraman image* x *(*256×256 *samples). (b) Observed image* y *(blurred signal-to-noise ratio (BSNR) = 40 dB). Smoothed by a 9×9-point box-car blur + noise. (c) Wiener filter (FoRD with* $\alpha = 1$*) estimate (SNR = 20.7 dB, improvement in SNR (ISNR) = 5.5 dB). The* $|X(f_n)|^2$ *required by the Wiener filter was estimated using the iterative technique of [8]. The ripples in the image result because the underlying Fourier basis elements have support over the entire spatial domain. (d) WaRD with* $\alpha = 0.1$ *(SNR = 22.5 dB, ISNR = 7.3 dB). In contrast to the Wiener estimate, the smooth regions and most edges are well preserved in the WaRD estimate, thanks to the spatially-localized wavelet basis functions. However, some faint features such as parts of the grass are lost during wavelet-domain estimation. The WVD algorithm is not applicable in this case since the box-car blurring operator used in the example is non-invertible. For the same experiment, the multiscale Kalman filter of [9] provides an ISNR of 6.7 dB. See Section 7 for more details.*

Kalifa, Mallat, and Rougé [10]. We will show in Section 5 that WaRD subsumes WVD and thus WaRD possesses the same asymptotically (as $N \to \infty$) near-optimal error decay rates as WVD for special operators such as the Radon transform. However, at any fixed sample-size N, WaRD will outperform WVD. Furthermore, unlike WVD, WaRD is applicable to any convolution operator.

The mirror-wavelet basis approach of Kalifa et al. [10] adapts to the frequency response of the convolution operator $\mathcal{H}$. Though the adapted basis improves upon the WVD, it is not effective for all types of ill-conditioned systems. For example, when $\mathcal{H}$ has a box-car impulse response h, adapting to the sinc frequency response H using wavelets fails.

Nowak [11] have employed an under-regularized system inverse and subsequently used wavelet-domain signal estimation. However, they did not address the implications of using the regularization and the choice of the optimal amount of regularization.

Banham and Katsaggelos [9] apply a multiscale Kalman filter to the deconvolution problem. Their approach employs an under-regularized constrained-least-squares prefilter to reduce the support of the state vectors in wavelet domain, thereby improving the computational efficiency of the multiscale restoration filter. The amount of regularization chosen for each wavelet scale is the lower bound that allows for reliable edge classification.

While similar in spirit to the multiscale Kalman filter approach, in WaRD the amount of regularization is chosen to optimize the overall MSE performance of the deconvolved estimate. In addition, WaRD employs simple shrinkage on the wavelet coefficients of an over-complete wavelet basis in contrast to more complicated prediction on edge and non-edge quad-trees over an orthonormal wavelet basis [9]. As mentioned in Figure 3, WaRD outperforms the multiscale Kalman technique in terms of the improvement in signal-to-noise-ratio (ISNR).

There is also a vast amount of literature on iterative techniques [12, 13]. In this paper, we have focused only on non-iterative deconvolution techniques.

1.6 Paper organization

After discussing regularization in more depth in Section 2, we briefly review wavelet transforms and their properties in Section 3. We outline previous wavelet-based deconvolution techniques in Section 4. We present our improved WaRD scheme in Section 5, elaborate on its implementation in Section 6. Illustrative examples lie in Section 7. We conclude by summarizing our work and sketching future directions in Section 8.

2 Fourier-domain Regularized Deconvolution (FoRD)

2.1 The FoRD algorithm

Given the general deconvolution problem from the Section 1, FoRD can be understood formally as follows (see Figure 2(a)):

1. **Pure inversion:** Treat the y with H^{-1} to obtain a noisy, unbiased estimate $\widetilde{x}$ of the input signal x as in (3). This necessarily amplifies the noise components at frequencies where $|H(f_n)|$ is small.

2. **Fourier-domain signal estimation:** Shrink each frequency component of the noisy signal $\widetilde{x}$ using frequency-dependent weights

$$R_\alpha(f_n) \;:=\; \frac{|H(f_n)|^2\,|P_x(f_n)|}{|H(f_n)|^2\,|P_x(f_n)| \;+\; \alpha\,\sigma^2}, \tag{4}$$

   where $P_x(f_n)$ is the power spectral density (PSD) of the input signal.[4] This yields the input signal estimate $\widehat{X}_{\text{FoRD}}(f_n) := R_\alpha(f_n)H^{-1}(f_n)Y(f_n)$.[5]

The parameter α, called the *regularization parameter*, controls the tradeoff between the amount of noise suppression and the amount of signal distortion. Setting $\alpha = 0$ gives an unbiased but noisy estimate. LTI Wiener deconvolution corresponds to $\alpha = 1$ [14]. Setting $\alpha = \infty$ completely suppresses the noise, but also completely distorts the signal ($\widehat{x}_{\text{FoRD}} = 0$).

2.2 Optimality of FoRD

For Gaussian wide-sense-stationary signals, the LTI Wiener deconvolution provides the globally MSE-optimal estimate for the input, since the Fourier domain provides the ideal representation for both the colored noise after inversion and the signal of interest.

[4] This assumes the signal x to be a stationary random process. We will rather assume x to be deterministic and substitute $P_x(f_n) = |X(f_n)|^2$ in (4).

[5] Conceptually, it is easier to understand FoRD as a two step process. However, in practice, Steps 1 and 2 are implemented jointly. Since $R_\alpha(f_n)H^{-1}(f_n)$ is defined even when H contains zeros in its frequency response, FoRD is always applicable.

2.3 Drawbacks of FoRD

The LTI Wiener filter does not provide a good estimate when the input signal contains spatially localized phenomena such as edges. Although the Fourier domain remains the ideal domain to represent the colored noise $H^{-1}\Gamma$ (see (3)), it is not well-suited to represent the signal x. Since the supports of the Fourier basis functions extend over the entire spatial domain, scalar operations on the Fourier coefficients lack spatial adaptivity. Consequently, important spatial components of the input signal such as edges and ridges become distorted.

2.4 Alternative solutions

Deconvolution techniques must take the spatial variations of the signal into account to produce the best possible results. One such technique is the best linear estimator, the time-varying or matrix version of the Wiener inverse [15]. However, the time-varying Wiener filter is impractical, because it is not only computationally intensive ($O(N^3)$), but more importantly, requires knowledge of the input signal cross-correlation matrix. Robust estimation of the input signal cross-correlation matrix from the blurred and noisy observation y turns out to be formidable even if matrix is approximated as block circulant [16].

3 Background on Wavelets and Signal Estimation

3.1 Wavelet transform

The joint time-frequency analysis of the wavelet transform efficiently captures spatially varying features in a signal. The discrete wavelet transform (DWT) represents a 1-d signal x in terms of shifted versions of a low-pass scaling function ϕ and shifted and dilated versions of a prototype bandpass wavelet function ψ [17, 18]. For special choices of ϕ and ψ, the functions

$$\psi_{j,k}(t) \;:=\; 2^{j/2}\,\psi\big(2^j t - k\big)\,, \tag{5}$$

$$\phi_{j,k}(t) \;:=\; 2^{j/2}\,\phi\big(2^j t - k\big)\,, \qquad \text{with } j,k \in \mathbb{Z} \tag{6}$$

form an orthonormal basis. A finite-resolution approximation x_J to x is given by

$$x_J(t) \;=\; \sum_k u_{j_0,k}\,\phi_{j_0,k}(t) \;+\; \sum_{j=j_0}^{J}\sum_k w_{j,k}\,\psi_{j,k}(t), \tag{7}$$

with wavelet coefficients $u_{j,k} := \int x(t)\,\phi^*_{j,k}(t)\,dt$ and $w_{j,k} := \int x(t)\,\psi^*_{j,k}(t)\,dt$. The parameter J controls the resolution of the wavelet reconstruction x_J of x; in fact, $x_\infty = x$.

The wavelet transform can be extended to represent sampled signals as well. For a discrete-time signal with N samples, the N wavelet coefficients $\{u_{j_0,k}, w_{j,k}\}$ can be easily computed using a filter bank consisting of low-pass filters, high-pass filters, and decimators. Due to the special filter bank structure, the forward and inverse wavelet transform can be computed in $O(N)$ operations [18]. For simplicity of discussion alone, we will use the periodic DWT, which employs circular convolutions in its filter bank.[6] For brevity, we will refer to the set of scaling and wavelet coefficients collectively as $\{\theta_{j,k}\} := \{u_{j_0,k}, w_{j,k}\}$. Multidimensional DWTs are easily obtained by alternately wavelet-transforming along each dimension [17].

3.1.1 Multiresolution and time-frequency localization of wavelets. The wavelet transform provides a multiscale representation of a signal, i.e., the wavelet coefficients capture the signal features at different resolution levels. In (7), j indexes the *scale* or the resolution of analysis — large j corresponds to higher resolution of analysis, while small j corresponds to the coarse scale or lowest resolution of analysis. The scale $j = J$ corresponds to the finest scale or highest resolution of analysis. In (7), k indexes the spatial location of analysis. For a wavelet $\psi(t)$ centered at time zero and frequency f_0, the wavelet coefficient $w_{j,k}$ measures the signal content around time $2^{-j}\,k$ and around frequency $2^j\,f_0$. Thus, wavelets exhibit simultaneous spatial and frequency localization.

3.1.2 Wavelets as unconditional bases. In essence, the unconditional basis property of wavelets means that signals are characterized by only the amplitudes of their wavelet coefficients [4, 7]. Wavelets provide an unconditional basis for spaces such as Besov spaces which contain signals and images with edges and ridges [4]. L_p spaces (for $1 < p < \infty$) and Sobolev spaces and also belong to the Besov scale. In contrast, the Fourier basis is not an unconditional basis for such a wide class of signal spaces.

The implications of this abstract notion of unconditional basis are extremely appealing. Unconditional basis provide economical signal representations by capturing most of the signal energy in just a few large coefficients [19]. For example, the wavelet coefficients of signals in a Besov space decay exponentially with scale. Such economical representations are desirable for signal estimation and compression using non-linear approximation [20, 21]. Furthermore, the unconditional basis property also ensures that simple scalar operations are sufficient to achieve near-optimal

[6] For deconvolving non-circular convolution operators, a non-periodic DWT can be employed [17, 18].

estimation.

3.2 Signal estimation by wavelet shrinkage

Wavelets provide a natural and effective solution to the problem of signal estimation in the presence of white noise [5, 22]. Many real-world signals have economical wavelet-domain representations in which a few large wavelet coefficients capture most of the signal energy [17, 23]. However, since a wavelet transform is orthonormal and linear, the energy of white noise remains scattered over all of the wavelet coefficients. This disparity between the signal and noise representation in the wavelet domain has been exploited in a number of powerful, near-optimal, signal estimation techniques based on simply shrinking the wavelet coefficients of the noisy signal.[7] Signal estimation using wavelet shrinkage is a spatially adaptive process (it smoothes more in the smooth regions of the signal) well-suited to signals with edges and other singularities.

Wavelet domain signal estimation techniques based on shrinkage can also be extended to estimate signals in the presence of colored noise. The optimality of such a wavelet-based approach now becomes dependent on the coloring of the noise in addition to the signal class [24].

4 Wavelet-Vaguelette Deconvolution (WVD)

Donoho [7] studied the application of wavelets to a special class of linear inverse problems and proposed the wavelet-vaguelette decomposition. We will call wavelet-vaguelette decomposition applied to deconvolution as wavelet-vaguelette deconvolution (WVD).

4.1 The WVD algorithm

WVD consists of the following steps (see Figure 2(b)):

1. **Pure inversion:** As in FoRD, obtain a noisy, unbiased estimate $\widetilde{x}$ of the input signal as in (3).

2. **Wavelet-based signal estimation:** In contrast to FoRD's Fourier-domain shrinkage, WVD estimates the signal from the noisy $\widetilde{x}$ by shrinking each wavelet coefficient of the

[7]The optimality is in terms of the rate of error decay as increasingly denser observation samples are obtained.

noisy signal.[8] The variance of noise corrupting all wavelet coefficients in a particular scale is the same, but varies with different scales [24]. Hence scale-dependent shrinkage is employed to estimate the signal wavelet coefficients. The inverse DWT then yields the WVD estimate from the estimated signal wavelet coefficients.

In contrast to FoRD, WVD can be viewed as spatially varying regularization provided by the spatial adaptivity of wavelets.

4.2 Optimality of WVD

Donoho [7] showed that a WVD deconvolution approach is near-optimal to recover a wide class of signals (e.g., those in Besov spaces) when the linear operator $\mathcal{H}$ is dilation-homogeneous, i.e.,

$$h(t) \circledast y(at) = a^{-\nu} \, |a| \, h(at) \circledast y(at), \tag{8}$$

for some exponent ν. The Radon transform is an important example of such an operator. To the best of our knowledge, the optimality of WVD for general dilation-inhomogeneous operators is unknown. Using wavelet-based techniques, Nowak et al. [11] observed impressive results for some common LTI operators as well. Kalifa et al. [10] has advocated a similar philosophy and obtained excellent performances in satellite image recovery.

Consider the example of piecewise polynomial functions to gain insight into the advantages of the WVD approach. Choose a wavelet system with basis functions whose number of zero-moments is greater than or equal to the degree of all the polynomial segments. If the support of any wavelet basis function lies within any polynomial piece, then the corresponding wavelet coefficient is zero. Because the support of only $O(\log N)$ wavelet basis functions do not completely lie within any polynomial piece, at most $O(\log N)$ wavelet coefficients are non-zero. Thus the wavelet transform represents a length-N piecewise polynomial signal economically. Assuming an invertible low-pass $\mathcal{H}$ whose frequency response decays polynomially, $H(f) = f^{-\nu}$, the error-per-sample in a WVD system decays rapidly as $N^{\frac{-1}{2\nu+1}} \log N$ as denser samples ($N \to \infty$) of the underlying continuous-time observations are obtained [25] because wavelets economically represent a piece-

[8] Note that different shrinkage techniques such as hard thresholding or soft thresholding can be used on the noisy wavelet coefficients, but the philosophy remains the same: Use wavelet-domain estimation instead of Fourier-domain estimation.

wise polynomial function; this is significantly faster than the rate $\frac{-1}{2\nu+2}$ achieved by Fourier-domain shrinkage.

4.3 Drawbacks of WVD

Such a strictly wavelet-based deconvolution approach has its limitations. These can be understood by focusing on the variance of the colored noise $H^{-1}\Gamma$ (see (3)) at different wavelet scales. The wavelet basis function frequency response at any scale j is nearly band-limited to $[f_{2^{j-1}}, f_{2^j})$ and almost constant within this band. If the zeros in the frequency response H are of sufficiently low order and the zeros occur at normalized frequency 0 or 0.5, then the noise variance σ_j^2 at scale j is well-approximated by

$$\sigma_j^2 \approx \frac{1}{2^{(j-1)}} \sum_{n=2^{(j-1)}}^{2^{(j)}-1} \frac{1}{|H(f_n)|^2} \sigma^2, \tag{9}$$

where $H(f_n)$ are the discrete Fourier coefficients of the blurring operator $\mathcal{H}$, and σ^2 is the variance of the AWGN γ. From (9), it is clear that if $H(f_n)$ is small at any isolated frequency f_n, then the noise variance in the corresponding wavelet scale is extremely high. This renders ineffective the task of extracting the signal coefficients from this scale using any form of scalar shrinking of the wavelet coefficients (see Step 2 in Section 4) . In general, the frequency response H^{-1} outside $[f_{2^{j-1}}, f_{2^j-1}]$ influences the noise variance, but to a lesser extent. So, if the frequency response H has a zero whose order is greater than the number of vanishing moments of the wavelet system, or if the frequency response H goes to zero at any other arbitrary frequency, then the noise variance is infinite at all wavelet scales. In such a case, WVD provides a zero estimate.

4.4 Best-basis solution improves performance

In cases when expression (9) holds, the noise variance at scale j is primarily influenced by the singularities of H^{-1} that lie in the corresponding frequency band. In such a case, Kalifa et al. [10] advocate adapting the wavelet basis to H^{-1} to improve on the performance of WVD. The adaptation is achieved by using a *mirror-wavelet basis* with a time-frequency tiling structure different from that of conventional wavelets. Such a tiling isolates the frequency where the convolution operator H approaches zero by using wavelet basis functions with narrower frequency responses. This attenuates the noise in many wavelet coefficients, thereby facilitating better estimation of the mirror-wavelet coefficients of the original signal.

However, in general it is impossible to obtain a set of basis functions that can isolate all such singular frequencies in H^{-1}, because the wavelet basis functions are not exactly band-limited. For example, no adapted wavelet basis scheme can provide a satisfactory solution when the blurring operator has a box-car system impulse response, since the sinc frequency response H has multiple zeros. In such a case, the variance of colored noise $H^{-1}\,\Gamma$ is infinite in all wavelet scales, rendering the signal estimation in any adapted wavelet basis domain ineffective.

5 Wavelet-based Regularized Deconvolution (WaRD)

To solve the deconvolution problem, we propose a wavelet-based regularized deconvolution (WaRD) algorithm that simultaneously exploits the Fourier-domain representation of the convolution operator and the spatially adaptive wavelet-domain representation of the signal. The Fourier domain is ideal for identifying and attenuating noise components amplified during inversion of $\mathcal{H}$, since it diagonalizes the convolution operator. However, since the supports of Fourier basis elements extend over the entire spatial domain, the energy of signals containing spatially localized phenomena such as edges spreads over many Fourier coefficients. Solely employing Fourier-domain shrinkage, as in FoRD, thus results in significant distortions reflected as ringing around the edges in the estimate. In contrast, the WVD exploits the economical signal representation in the wavelet domain. However, noise components that are severely amplified during system inversion corrupt many wavelet coefficients, thereby limiting the effectiveness of wavelet-domain signal estimation. In WaRD, a small amount of Fourier-domain regularization, which attenuates the severely amplified Fourier noise components, is used to substantially reduce the wavelet-domain noise variance without imparting significant signal distortion.

5.1 The WaRD algorithm

The WaRD algorithm consists of the following steps (see Figure 2(c)):

1. **Pure inversion:** Similar to FoRD and WVD, obtain a noisy, unbiased estimate $\widetilde{x}$ of the input signal as in (3).

2. **Fourier-domain noise attenuation:** Employ a small amount of Fourier-domain shrinkage using weights $R_\alpha(f_n)$ (see (4)) to achieve partial noise attenuation in the estimate $\widetilde{x}_\alpha$

$$\widetilde{X}_\alpha(f_n) \quad := \quad R_\alpha(f_n)\,\widetilde{X}(f_n)$$

$$= \quad R_\alpha(f_n)\, X(f_n) + \left(\frac{1}{H(f_n)}\right) R_\alpha(f_n)\, \Gamma(f_n), \tag{10}$$

where $\widetilde{X}_\alpha$ and $\widetilde{X}$ are the length-N DFTs of $\widetilde{x}_\alpha$ and $\widetilde{x}$ respectively (see Figure 2(c))). Here the regularization parameter α, which controls the amount of Fourier-domain shrinkage, is typically much smaller than that used to obtain a Wiener estimate (i.e., $\alpha \ll 1$). Even a small amount of regularization ensures that the severely amplified Fourier noise components in $\widetilde{x}$ are significantly attenuated. Section 5.5 discusses the choice of α in greater depth. Steps 1 and 2 together constitute a FoRD system (see Figure 2(a)), but with a small α.

3. **Wavelet-domain signal estimation:** Since the estimate $\widetilde{x}_\alpha$ in (10) still contains some residual noise, as in Step 2 of the WVD algorithm, we shrink the wavelet coefficients of $\widetilde{x}_\alpha$ at each scale according to the noise variance at that scale to obtain the WaRD estimate. Wavelet-domain signal estimation remains effective, since the noise corrupting the wavelet coefficients is not excessive, thanks to Step 2.

5.2 Tradeoff: Distortion vs. noise attenuation

Since Fourier-domain noise reduction comes at the cost of signal distortion, the amount of Fourier-domain regularization needs to be controlled. This raises the question: how to choose the best value for the regularization parameter α? The bias-variance tradeoff is clear: On one hand, since Fourier-domain shrinkage smears non-stationary signal features such as edges (bias), α should be as small as possible. On the other hand, large α prevents excessive noise amplification during inversion (variance), which aids the wavelet-domain signal estimation.

We wish to determine the *optimal* regularization parameter $\alpha^\star$ for the WaRD system by minimizing the overall $\mathrm{MSE}(\alpha)$ with respect to α. This optimal regularization parameter $\alpha^\star$ will balance the amount of Fourier-domain and wavelet-domain noise reduction.

5.3 Cost function

Since it is not possible to analytically express the exact $\mathrm{MSE}(\alpha)$ of a WaRD system with regularization parameter α conveniently due to WaRD's inherent non-linearity, we propose an approximation $\widetilde{\mathrm{MSE}}(\alpha)$. The cost function $\widetilde{\mathrm{MSE}}(\alpha)$ comprises the distortion error due to the Fourier-domain regularized inversion and error incurred during wavelet-domain signal estimation. The cost function assumes that *oracle hard thresholding* T is employed during signal estimation in the wavelet domain [5, 6]. Oracle thresholding keeps a noisy wavelet coefficient only if the signal power in that coefficient is greater than the noise power; otherwise, the coefficient is set to zero. Defining $\theta_{j,k}$ as the wavelet coefficients of the input signal x, $\widetilde{\theta}^{\alpha}_{j,k}$ as the wavelet coefficients of the noisy signal estimate $\widetilde{x}_{\alpha}(f_n)$ obtained after partially regularized inversion in (10), and $\sigma_j(\alpha)$ as the standard deviation of the noise $\frac{R_\alpha(f_n)}{H(f_n)}\Gamma(f_n)$ from (10) at wavelet scale j, we set

$$T\left(\widetilde{\theta}^{\alpha}_{j,k}\right) \quad := \quad \begin{cases} \widetilde{\theta}^{\alpha}_{j,k}, & \text{if } |\theta_{j,k}| > \sigma_j(\alpha) \\ 0, & \text{if } |\theta_{j,k}| \leq \sigma_j(\alpha). \end{cases} \tag{11}$$

Oracle thresholding assumes that the signal under consideration (and hence $\theta_{j,k}$) is known.

Our cost function $\widetilde{\mathrm{MSE}}(\alpha)$ is given by

$$\widetilde{\mathrm{MSE}}(\alpha) \quad := \quad \sum_{n=0}^{N-1} [1 - R_\alpha(f_n)]^2 \, |X(f_n)|^2 \; + \; \sum_{j,k} \min\left(|\theta_{j,k}|^2, \, \sigma_j^2(\alpha)\right). \tag{12}$$

The first term is an estimate of the distortion in the input signal due to regularized inversion [26]. This distortion error is an increasing function of α. The second term is an estimate of the error due to wavelet-domain oracle thresholding [6]. This thresholding error is a decreasing function of α, since the noise variance $\sigma_j^2(\alpha)$ decreases as α increases.

5.4 Accuracy of the cost function

The cost function $\widetilde{\mathrm{MSE}}(\alpha)$ closely approximates the exact $\mathrm{MSE}(\alpha)$ of the WaRD system. To see this, consider the following two cases:

Case (1) $|\theta_{j,k}| > \sigma_j(\alpha)$: The contribution due to the oracle thresholding error term (second term) in $\widetilde{\mathrm{MSE}}(\alpha)$ is $\sigma_j^2(\alpha)$ (see (11) and (12)). This is the *exact* MSE incurred during the estimation of $\theta_{j,k}$ in the WaRD system.

Case (2) $|\theta_{j,k}| \leq \sigma_j(\alpha)$: The cost that $\widetilde{\mathrm{MSE}}(\alpha)$ associates with the estimation of $\theta_{j,k}$ is $|\theta_{j,k} - \theta^{\alpha}_{j,k}|^2 + |\theta_{j,k}|^2$, while the exact incurred error is $|\theta_{j,k}|^2$. The distortion is determined by the SNR and the regularization parameter (see (14) in Appendix A). At coarse scales, the distortion relative to the signal energy is small, because the SNR is typically large. At fine scales, the region of interest is around small α's; for small α's, the distortion relative to the signal energy is again small. Consequently, $\widetilde{\mathrm{MSE}}(\alpha) \approx \mathrm{MSE}(\alpha)$.

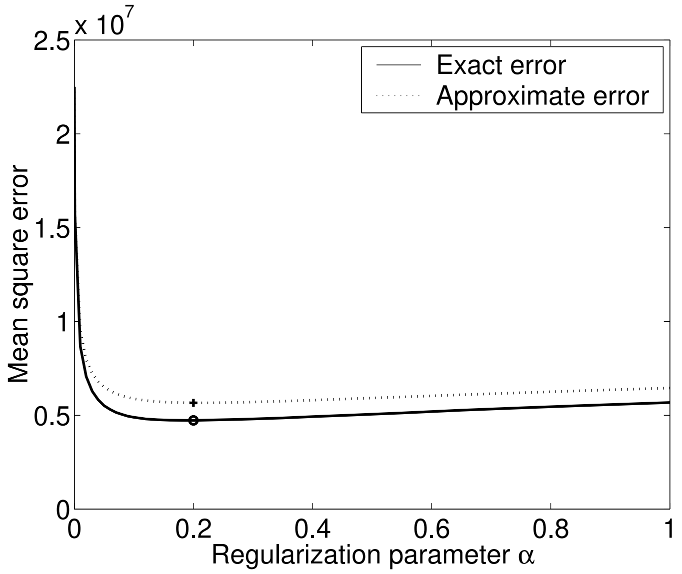


Figure 4: $\widetilde{\mathrm{MSE}}(\alpha)$ *vs. MSE*(α) *as a function of the regularization parameter* α *for the experimental setup described in Section 7. Clearly, the approximate error* $\widetilde{\mathrm{MSE}}(\alpha)$ *(dotted line) closely matches the exact MSE*(α) *incurred using oracle thresholding (solid line) for different values of* α*. The respective global minima,* $\alpha^\star$ *and* $\widetilde{\alpha}^\star$*, are marked by "$\circ$" and "$+$" respectively. Since* $\alpha^\star$ *and* $\widetilde{\alpha}^\star$ *agree closely, we will minimize* $\widetilde{\mathrm{MSE}}(\alpha)$ *to determine the regularization parameter that balances the amount of Fourier-domain and wavelet-domain noise reduction in the WaRD system.*

Even in the worst case, $\mathrm{MSE}(\alpha) \leq \widetilde{\mathrm{MSE}}(\alpha) \leq 2 \times \mathrm{MSE}(\alpha)$, since $|\theta_{j,k} - \theta_{j,k}^{\alpha}|^2 \leq |\theta_{j,k}|^2$. Figure 4 confirms that the cost function $\widetilde{\mathrm{MSE}}(\alpha)$ closely approximates the actual error $\mathrm{MSE}(\alpha)$ incurred using oracle thresholding for the experimental setup consisting of the input image, blurring function, and noise level described in Section 7. Hence we will minimize $\widetilde{\mathrm{MSE}}(\alpha)$ to find the regularization parameter $\widetilde{\alpha}^\star$ that balances Fourier-domain regularized inversion and wavelet-domain signal estimation.[9]

5.5 Optimal α for each scale

In the previous sections, we assumed a single Fourier-domain regularized inverse with regularization parameter α common for all wavelet scales. An interesting generalization is to employ a different Fourier-domain regularized inverse with regularization parameter α_j at each wavelet

[9]Ideally, we would like to determine $\alpha^\star$ that minimizes the exact $\mathrm{MSE}(\alpha)$, but the exact $\mathrm{MSE}(\alpha)$ cannot be quantified conveniently.

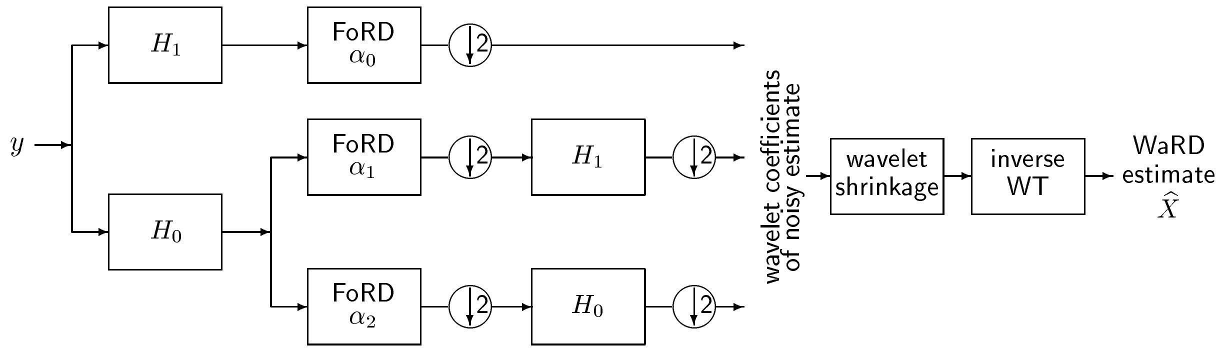


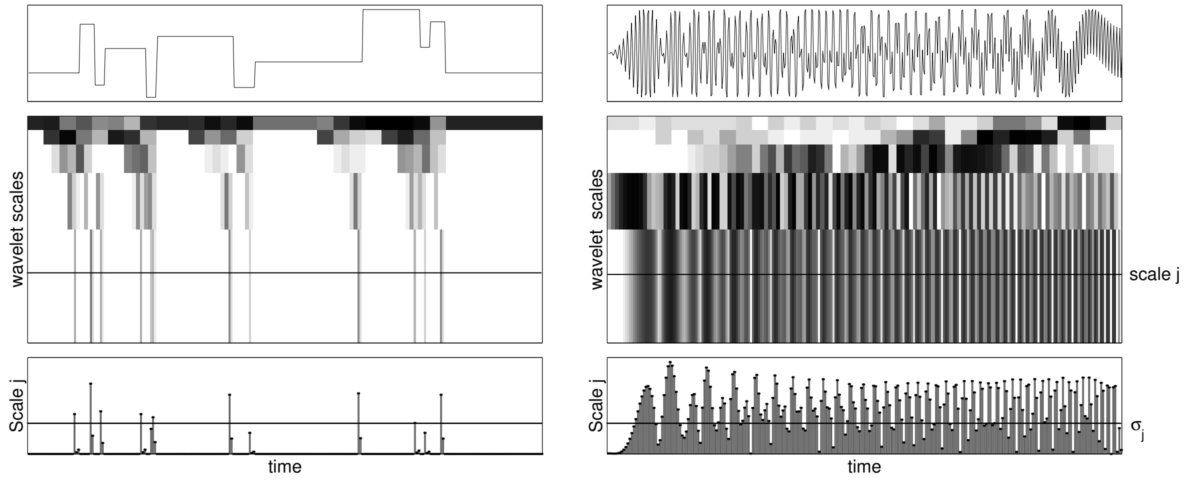
Figure 5: *Different regularization parameters for each wavelet scale. Blocks H_0 and H_1 denote the low-pass and the high-pass filters of the usual wavelet filter bank [18, pp. 35]. FoRD, with different amounts of Fourier shrinkage, is used in the different wavelet scales. Using different α_j for each scale makes the cost function $\widetilde{\mathrm{MSE}}$ separable with respect to each α_j.*

scale j (see Figure 5). This generalization makes the cost function $\widetilde{\mathrm{MSE}}$ separable with respect to the regularization parameter at each scale, thereby simplifying the solution for the optimal α_j for each scale. By minimizing the cost function with respect to the α_j, we show in Appendix A that the optimal regularization parameter $\widetilde{\alpha}_j^\star$ for scale j satisfies

$$\widetilde{\alpha}_j^\star = \frac{1}{2^j} \# \left\{ |\theta_{j,k}| > \sigma_j(\widetilde{\alpha}_j^\star) \right\}, \tag{13}$$

with $\# \left\{ |\theta_{j,k}| > \sigma_j(\widetilde{\alpha}_j^\star) \right\}$ the number of wavelet coefficients $\theta_{j,k}$ larger than the noise standard deviation $\sigma_j(\widetilde{\alpha}_j^\star)$. In words, (13) means that *the optimal regularization parameter equals the proportion of the input signal wavelet coefficients larger in magnitude than the noise standard deviation.*

Figure 6 captures the intuitive nature of this result. For signals with sparse wavelet-domain representations (see Figure 6(a)), (13) advocates a small amount of Fourier regularization ($\widetilde{\alpha}_j^\star \ll 1$), since the optimal $\widetilde{\alpha}_j^\star$ is controlled by the number of wavelet coefficients exceeding the noise standard deviation in magnitude. Thus, (13) suggests that most of the noise attenuation in the WaRD system should be performed in the wavelet domain, and not in the Fourier domain. For signals with non-sparse wavelet representation (see Figure 6(b)), (13) advocates a large amount of Fourier regularization ($\widetilde{\alpha}_j^\star \approx 1$), which implies that wavelet-domain processing plays a secondary role. Thus, $\widetilde{\alpha}_j^\star$ strikes the optimal balance between Fourier-domain processing and wavelet-domain processing.



(a) (b)

Figure 6: *Effect of the sparsity of the wavelet-domain signal representation on the value of the optimal regularization parameter* $\alpha^\star$ *in WaRD. (a) This1-d signal (top) has a sparse wavelet-domain representation. This is illustrated in the wavelet coefficient time-frequency plot (middle), where darker shades of gray indicate larger magnitudes. The solid line in the middle plot denotes wavelet scale* j. *The plot of the wavelet coefficients at scale* j *(bottom) indicates that most of the signal energy is concentrated in just a few large coefficients. The solid line in the bottom plot denotes a typical noise standard deviation* σ_j. *Since the number of signal wavelet coefficients exceeding* σ_j *is small, the optimal* α_j *for this signal is small* ($\widetilde{\alpha}_j^\star \ll 1$)*, according to (13). Thus, we should rely primarily on wavelet-domain shrinkage in the WaRD estimate. (b) In contrast, this signal does not enjoy an economical wavelet-domain representation, and the number of wavelet coefficients exceeding any typical* σ_j *is large. Hence, according to (13), the optimal* α_j *for this signal is large* ($\widetilde{\alpha}_j^\star \approx 1$)*, implying that we should employ a significant amount of Fourier-domain shrinkage in the WaRD estimate.*

The optimal regularization parameter $\widetilde{\alpha}_j^\star$ should never be set to 0. Multiple values of α_j can satisfy (13). The value $\alpha_j = 0$ satisfies (13) when the noise variance at $\alpha_j = 0$ is greater than each wavelet coefficient energy at scale j. For a WaRD system with $\alpha_j = 0$, all wavelet coefficients at scale j would be shrunk to zero during the wavelet-domain estimation stage in such a case. However, the noise variance can be arbitrarily reduced by increasing α_j so that not all wavelet coefficients are shrunk to zero, and some signal coefficients are salvaged. Hence even if $\alpha_j = 0$ satisfies (13), $\widetilde{\alpha}_j^\star$ should be set to some non-zero value that satisfies (13).[10]

Typically, $\widetilde{\alpha}_j^\star \neq 1$. The value $\alpha_j = 1$ satisfies (13) only when the noise variance at $\alpha_j = 1$ is less than the energy of each wavelet coefficient at scale j. This happens rarely since at least some of the wavelet coefficients are ≈ 0 (particularly true for fine scales).

Finally, we experimentally verify the accuracy of (13) in predicting the MSE-optimal regularization parameters $\alpha_j^\star$ using the input signal, blurring function, and noise level described in Section 7. We assume complete knowledge of the input signal to perform oracle thresholding (see (11)). Table 1 tabulates the regularization parameters $\widetilde{\alpha}_j^\star$'s recommended by (13) for the different wavelet subbands. We found the $\alpha_j^\star$ that minimizes the exact MSE in subband j by searching over all possible values for α_j. The excess error due to using the $\widetilde{\alpha}_j^\star$'s instead of $\alpha_j^\star$'s provides a measure to compare the accuracy of (13). Even for the worst case (1st row of Table 1), the increase in the overall error due to using $\widetilde{\alpha}_j^\star$ instead of $\alpha_j^\star$ is less than 1%. This substantiates our claim that the balance between the Fourier-domain processing and wavelet-domain processing can be struck by exploiting (13).

5.6 Optimality of WaRD

Similar to the WVD, the WaRD has asymptotically near-optimal rates of MSE decay for dilation-homogeneous operators. WVD [7] and mirror-wavelet basis technique [10] are special cases of WaRD[11] with $\alpha = 0$. By construction, WaRD includes the value $\alpha = 0$ in the search space for the $\widetilde{\mathrm{MSE}}$ optimal $\alpha^\star$. From the two cases analyzed in Section 5.4, it easily follows that $\mathrm{MSE} \leq \widetilde{\mathrm{MSE}} \leq 2\mathrm{MSE}$. Since the MSE for a WVD system decays with asymptotically near-optimal rates as the sampling density increases, WaRD also enjoys similar asymptotically

[10] In the rare cases when $\alpha_j = 0$ is the only value that satisfies (13), $\widetilde{\alpha}_j^\star$ should be set to 1.

[11] For comparisons with the mirror-wavelet basis approach [10], the use of a similarly adapted wavelet basis would be required.

Table 1: *Experimental verification that (13) provides good estimates of the MSE-optimal regularization parameter* $\alpha_j^\star$. *The input image, blurring function, and the noise level are as described in Section 7. The first column denotes the decomposition level of the wavelet transform with 5 denoting the finest scale. The second column specifies the three different subbands of the 2D wavelet decomposition, viz., high-pass vertically and horizontally (HH), high-pass vertically and low-pass horizontally (HL), and low-pass vertically and high-pass horizontally (LH). The third column contains the* $\alpha_j^\star$ *that minimizes the exact MSE of subband* j *(determined empirically by searching over all possible values for* α_j*). The fourth column contains the* $\widetilde{\alpha}_j^\star$ *advocated by (13), which minimizes* $\widetilde{\mathrm{MSE}}$ *in subband* j*. The fifth column contains the percentage increase in the MSE of subband* j *due to using* $\widetilde{\alpha}_j^\star$ *rather than* $\alpha_j^\star$*, while the sixth column contains the percentage increase in the overall MSE due to using the* $\widetilde{\alpha}_j^\star$ *rather than* $\alpha_j^\star$*. The percentage change in the MSE provides a measure for the "distance" between the predicted and the exact optimal regularization parameters. Clearly, we see that the* $\widetilde{\alpha}_j^\star$*'s provide near-optimal MSE performance. Even for the worst case (row 1), the increase in the overall MSE is less than 1%. Thus,* $\widetilde{\alpha}_j^\star$ *advocated by (13) accurately strikes the balance between Fourier-domain and wavelet-domain processing in a WaRD system.*

Decomposition level	Subband type j	Exact MSE minimizer, $\alpha_j^\star$	$\widetilde{\alpha}_j^\star$ advocated by (13)	% increase in subband MSE	% increase in total MSE
5	HH	0.06	0.16	6.5	9.6×10^{-1}
	HL	0.14	0.16	4.7×10^{-1}	1.3×10^{-1}
	LH	0.16	0.23	4.7×10^{-1}	9.0×10^{-2}
4	HH	0.16	0.18	5.7×10^{-1}	6.0×10^{-2}
	HL	0.35	0.29	2.3×10^{-1}	2.0×10^{-2}
	LH	0.35	0.34	1.2×10^{-1}	1.0×10^{-2}
3	HH	0.55	0.33	5.4×10^{-1}	2.0×10^{-2}
	HL	0.65	0.5	6.2×10^{-1}	1.0×10^{-2}
	LH	0.75	0.55	6.0×10^{-1}	1.0×10^{-2}
2	HH	0.75	0.84	1.0	4×10^{-4}
	HL	1	0.95	1.1×10^{-2}	3.0×10^{-6}
	LH	0.65	0.93	2.0	5.1×10^{-4}
1	HH	0.55	0.98	3.2	1.6×10^{-4}
	HL	1	1	0	0
	LH	1	1	0	0

near-optimal error decay rates for dilation-homogeneous operators.

However, the optimal $\alpha^\star$ is never 0 at a finite resolution N though it may approach zero with denser sampling , i.e., increasing N. Hence, WaRD will outperform wavelet-based deconvolution methods described in [7, 10, 27] in terms of $\widetilde{\mathrm{MSE}}$ at any given resolution. Thus, WaRD has small sample (i.e., finite resolution N) as well as asymptotic optimality properties.

Further, while the WVD is generally inapplicable when $\mathcal{H}$ is not invertible, WaRD gives excellent estimates even when $\mathcal{H}$ is non-invertible. However, the optimality of WaRD cannot be determined in such cases.

6 WaRD Implementation

The theoretical analysis in Section 5 assumes an ideal WaRD setup with knowledge of several quantities that are typically unavailable in practice. Here we overview the practical aspects of the WaRD algorithm.

6.1 Estimation of σ^2 and $|X(f)|^2$

The variance σ^2 of the additive noise γ and the Fourier spectrum $|X(f)|^2$ of the input signal are unknown in practice and must be estimated from the blurred observation y. The noise variance can be reliably estimated using a median estimator in the finest wavelet scale [22]. To estimate $|X(f_n)|^2$, we employ the iterative Wiener technique of [8]. However, since this estimate is not robust at frequencies where $H(f_n) \approx 0$, we boost the estimated $|X(f_n)|^2$ at these frequencies by adding a small positive constant to the estimate obtained after 10 iterations of the algorithm.

6.2 Choice of wavelet-domain estimation scheme

Oracle thresholding cannot be employed in practice, because it assumes knowledge of the wavelet coefficients of the unknown original signal. Hence empirical wavelet-domain estimation schemes need to be employed as a substitute. The choice of the estimation scheme influences the final performance of the WaRD system significantly. We have found that the wavelet-domain *Wiener-shrink* estimation algorithm [28, 29] outperforms in the MSE sense conventional wavelet-domain estimation schemes that employ hard and soft thresholding. Wiener-shrink estimation works as follows: First obtain a rough estimate of the input signal using a conventional wavelet-domain thresholding technique. Then, use this estimate to obtain a final refined estimate by employing Wiener estimation on each wavelet coefficient.

Wavelet-domain estimation schemes based on the DWT are not shift-invariant, i.e., translations of y will result in different estimates. The redundant, shift-invariant DWT yields significantly improved estimates [30] by averaging over all possible shifts of the observation y. We employ a redundant, shift-invariant DWT with wavelet-domain Wiener-shrink to estimate the input signal.

6.3 Choice of α

The theoretical analysis in Section 5 climaxed with expression (13), which helps us quantitatively understand the influence of the wavelet-domain signal representation on the choice of the optimal regularization parameters in the ideal case. However, the derivation of condition (13) (in Appendix A) assumes oracle thresholding, which is not possible practically. Since the choice of wavelet-domain estimation algorithm influences the value of the optimal regularization parameter α_j, we cannot employ (13). This necessitates the use of empirical techniques to set the regularization parameters α_j's.

First, we note that the improvement gained by using a different α_j for each wavelet scale j (see Figure 5) as compared to using a common α for all the scales is negligible; for the setup described in Section 7, using a common $\alpha = 0.2$ for all scales increased the error by a meager 3%. Hence we empirically determine a common regularization parameter for all wavelet scales.

To determine a good value for α, we exploit our observation that the MSE of a WaRD system changes very gradually for α near and above $\alpha^\star$ (the MSE decreases rapidly when $\alpha \ll \alpha^\star$). Since the MSE, the l_2 norm of the WaRD estimate, and the l_2 norm of the desired signal satisfy the triangle inequality, the region of insignificant change in the MSE coincides with the region of insignificant change in the l_2 norm of the WaRD estimate with α. Hence, from a plot of the l_2 norm of the WaRD estimate versus α, we choose the smallest α from the region of insignificant change in the WaRD estimate norm to obtain near-optimal results (see [25] for more details).

6.4 Computational Cost

The overall computational complexity of the WaRD algorithm is primarily determined by the complexity of calculating the Fourier transform and the redundant DWT; both require $O(N \log N)$ operations, with N the number of samples. Hence, given the regularization parameter and an estimate of $|X(f_n)|^2$, the computational cost of the WaRD algorithm is also $O(N \log N)$.

7 Results

We illustrate the performance of the WaRD algorithm using a 2-d simulation as described by Banham et al. [9]. The input x is the 256×256 *Cameraman* image and the discrete-time system response h is a 2-d, 9×9-point box-car blur. Such a response is commonly used as a model for blurring due to a square scanning aperture such as in a CCD camera [2]. Define the blurred SNR (BSNR) as $10 \log_{10} \left(\|(x \circledast h) - \mathbb{E}[x \circledast h)]\|_2^2 / N\sigma^2 \right)$, where $\mathbb{E}[x \circledast h)]$ denotes the mean value of the blurred image; $N = 256 \times 256$ for this experiment. We set the additive noise variance σ^2 such that the BSNR was 40 dB. Figure 3 illustrates the original x, the observed y, the Wiener filter estimate (FoRD), and the WaRD estimate for $\alpha^\star = 0.1$ (determined empirically as described in Section 6). The WVD [7] and mirror wavelet basis [10] methods are not applicable in this situation, due to the many zeros in frequency response H.

As seen in Figure 3, WaRD clearly outperforms FoRD in overall visual quality and MSE. The WaRD estimate also outperforms the multiscale Kalman estimate proposed by Banham et al. [9]) in terms of improvement in the SNR (ISNR) as well as visual quality (compare Figure 7(d) in Banham et al. [9] with Figure 3(d)). ISNR is defined as $10 \log_{10} \left(\|x - \tilde{y}\|_2^2 / \|x - \widehat{x}\|_2^2 \right)$, with $\widehat{x}$ the estimate, and $\tilde{y}$ denotes the shifted version of the observation that minimizes the ISNR. For the reported experiment, $\tilde{y}$ is cyclically shifted with respect to y by $(4, 4)$ towards the top left corner. Banham et al. report an ISNR of 6.7' dB using their the multiscale Kalman filter. In contrast, the proposed WaRD technique provides an ISNR of 7.3 dB.

The difference in the quality of the estimates obtained using FoRD and WaRD is highlighted in the cross-sections through row 160 of images shown in Figure 7. The FoRD estimate cross-section shown in Figure 7(c) illustrates the failure of the FoRD to adapt to the smooth regions and the edges in the image simultaneously. This lack of spatial-localization reflects as ripples in the FoRD estimate. In contrast, Figure 7(d) clearly illustrates the spatial-adaptivity of WaRD. We observe that the smooth regions and the edges are preserved simultaneously.

8 Conclusions

In this paper, we have proposed an efficient multiscale deconvolution algorithm that optimally combines Fourier-domain regularized inversion and wavelet-domain signal estimation. WaRD can be potentially employed in a wide variety of applications such as satellite imagery, seismic

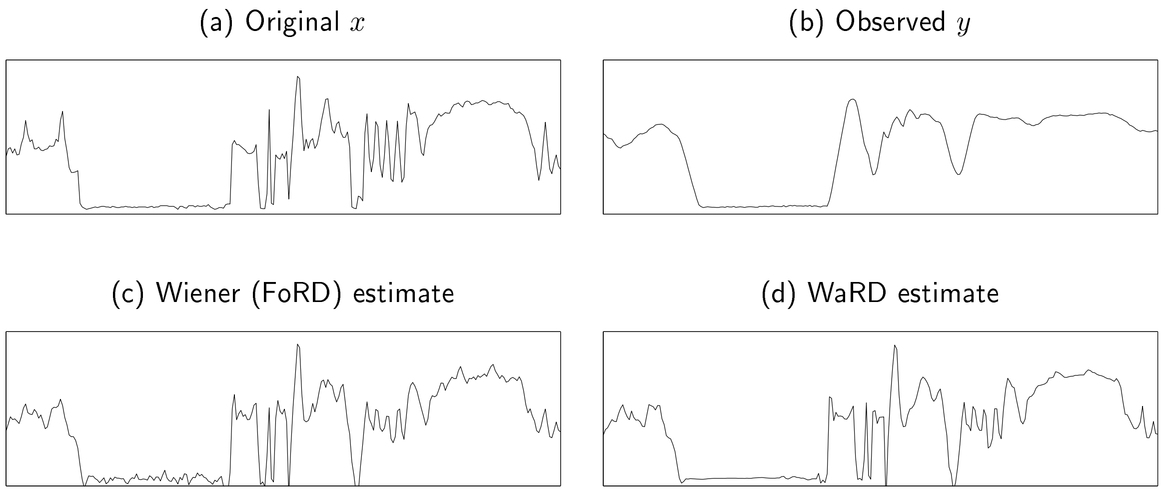


Figure 7: *(a) Cross-section of original image x (row 160 from Figure 3(a)) contains both smooth regions and discontinuities. (b) Cross-section of the blurred and noisy observed image y. (c) Cross-section of estimate obtained using the spatially invariant Wiener filter (FoRD with $\alpha = 1.0$) exhibits ringing artifacts in the smooth regions. (d) Cross-section of the hybrid WaRD estimate shows that smooth regions and edges are simultaneously preserved.*

deconvolution, and channel equalization to obtain enhanced deconvolution estimates.

For spatially varying signals, the WaRD outperforms the LTI Wiener filter and WVD in terms of both visual quality and MSE performance. Since WaRD subsumes WVD, WaRD also enjoys asymptotically near-optimal rates of error decay with increasing samples for convolution operators such as Radon transform. In addition, WaRD also improves on the performance of the WVD at any fixed resolution. Furthermore, WaRD continues to provide a good estimate of the original signal even when the convolution system is non-invertible. The computational complexity of the WaRD algorithm is just $O(N \log N)$, with N is the number of samples.

Theoretical analysis of the ideal WaRD algorithm reveals that the optimal regularization parameter at each wavelet scale is determined by the proportion of distorted input signal wavelet coefficients that exceed the variance of noise colored by Fourier-domain regularized inversion. For finite data samples, inversion without Fourier-domain regularization in a wavelet-based deconvolution system is never optimal. Further, using a regularization parameter $\alpha = 1$, which corresponds to employing a Wiener deconvolution filter for inversion, is also typically sub-optimal for most real-world signals. In essence, the balance between Fourier-domain and wavelet-domain processing is simultaneously determined by the frugality of the wavelet representation of the input signal and the Fourier-domain structure of the convolution operator. All results obtained in this paper trivially extend to colored noise denoising, because deconvolution is equivalent to signal estimation in the presence of colored noise (see (3)).

There are several avenues for future WaRD related research. We have focused on scalar processing during wavelet-domain estimation. However, there exist dependencies between the wavelet coefficients that can be exploited. We are currently working towards combining WaRD concepts with the hidden Markov tree model-based wavelet estimation [31]. We believe that exploiting such inter-dependencies in the wavelet domain will help preserve edges and other spatially localized phenomena better consequently leading to better deconvolution estimates.

An interesting twist to WaRD would be to first exploit the wavelet domain to estimate $x \circledast h$ from the noisy observation y and then invert the convolution operator. This technique, called the vaguelette-wavelet decomposition (VWD), has been studied by Silverman and Abramovich [32]. The salient point of such a technique is that the wavelet-domain estimation now deals with white noise instead of more complicated colored noise. However, now the reconstruction basis is no longer a signal-adapted wavelet-basis, but rather a hybrid basis that is not spatially localized. Like the WVD, this technique is also not applicable when $\mathcal{H}$ is non-invertible. The residual noise after wavelet-domain estimation of $x \circledast h$, but before inversion of the convolution operator, is time-varying. Hence, unlike in WaRD, FoRD may be inappropriate to perform the inversion. Construction of a universally applicable, hybrid deconvolution scheme that lies between WVD and VWD appears both promising and challenging.

In WaRD, we have assumed knowledge of the convolution operator. However, in many cases such as most practical imaging systems, the convolution operator is also unknown. In such blind deconvolution problems, the convolution system must be estimated from the observation. Though WaRD can be used to perform the deconvolution after estimating the convolution system, an interesting open problem is to adapt the WaRD framework to perform the estimation and deconvolution interdependently to obtain better estimates.

A Derivation of the Optimal Regularization Parameters

Our goal is to show (13). We approximate the WaRD estimate MSE as the sum of the distortion incurred during Fourier-domain regularized inversion and the error due to oracle thresholding of the wavelet coefficients (see (12)):

$$\widetilde{\mathrm{MSE}}(\alpha) = \sum_{n=1}^{N} \frac{\alpha^2\sigma^4|X(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha\sigma^2)^2} + \sum_{j=j_0}^{J}\sum_{k=1}^{2^j} \min\left(|\theta_{j,k}|^2, \sigma_j^2(\alpha)\right). \tag{14}$$

The scale $j = j_0$ denotes the coarsest wavelet scale, and $j = J$ denotes the finest scale. Since wavelets form an orthonormal basis, we have

$$\sum_{j=j_0}^{J}\sum_{k=1}^{2^j} |\Psi_{j,k}(f_n)|^2 = 1, \tag{15}$$

where $\Psi_{j,k}(f_n)$ denote the DFT-domain representations for the wavelet basis functions $\psi_{j,k}(t)$. Combining (14) and (15), we obtain

$$\widetilde{\mathrm{MSE}}(\alpha) = \sum_{j=j_0}^{J}\sum_{k=1}^{2^j}\left(\sum_{n=1}^{N} \frac{\alpha^2\sigma^4|X(f_n)|^2|\Psi_{j,k}(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha\sigma^2)^2}\right) + \min\left(|\theta_{j,k}|^2, \sigma_j^2(\alpha)\right). \tag{16}$$

A.1 Different $\widetilde{\alpha}\star$ for each wavelet scale

$\widetilde{\mathrm{MSE}}(\alpha)$ in (16) assumes that a common regularization parameter α is employed for all wavelet scales. More generally, we can employ different regularization parameters α_j for each wavelet scale j (see Figure 5). The cost function can then be expressed as a function of the different regularization parameters as

$$\widetilde{\mathrm{MSE}}(\alpha_{j_0},\ldots,\alpha_{J-1}) = \sum_{j=j_0}^{J}\sum_{k=1}^{2^j}\left(\sum_{n=1}^{N} \frac{\alpha_j^2\sigma^4|X(f_n)|^2|\Psi_{j,k}(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha_j\sigma^2)^2}\right) + \min\left(|\theta_{j,k}|^2, \sigma_j^2(\alpha_j)\right).$$

We will find the optimal regularization parameter at each scale by partially differentiating $\widetilde{\mathrm{MSE}}$ with respect to α_j and setting the derivative to zero.

A.2 Differentiating the distortion terms

Denote the total distortion error (first term in (17)) by

$$\mathrm{D}\left(\alpha_{j_0},\ldots,\alpha_J\right) := \sum_{j=j_0}^{J}\sum_{k=1}^{2^j}\sum_{n=1}^{N} \frac{\alpha_j^2\sigma^4|X(f_n)|^2|\Psi_{j,k}(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha_j\sigma^2)^2}. \tag{17}$$

Taking the partial derivative with respect to α_j, we have

$$\frac{\partial}{\partial\alpha_j}\mathrm{D}\left(\alpha_{j_0},\ldots,\alpha_J\right) = \sum_{k=1}^{2^j}\sum_{n=1}^{N} \frac{2\alpha_j\sigma^4|H(f_n)|^2|X(f_n)|^4|\Psi_{j,k}(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha_j\sigma^2)^3}. \tag{18}$$

For the sake of convenience, define

$$C(\alpha_j) := \sum_{k=1}^{2^j}\sum_{n=1}^{N} \frac{2\sigma^4|H(f_n)|^2|X(f_n)|^4|\Psi_{j,k}(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha_j\sigma^2)^3}; \tag{19}$$

thus,

$$\frac{\partial}{\partial \alpha_j} \mathrm{D}\left(\alpha_{j_0}, \ldots, \alpha_J\right) = \alpha_j\, C(\alpha_j). \tag{20}$$

A.3 Differentiating the oracle thresholding terms

Denote the total oracle thresholding error (second term in (17)) by

$$\mathrm{O}\left(\alpha_{j_0}, \ldots, \alpha_J\right) := \sum_{j=j_0}^{J} \sum_{k=1}^{2^j} \min\left(|\theta_{j,k}|^2, \sigma_j^2(\alpha_j)\right). \tag{21}$$

Taking the partial derivative of $\mathrm{O}\left(\alpha_{j_0}, \ldots, \alpha_J\right)$ with respect to α_j, we have

$$\begin{aligned} \frac{\partial}{\partial \alpha_j} \mathrm{O}\left(\alpha_{j_0}, \ldots, \alpha_J\right) &= \frac{\partial}{\partial \alpha_j} \sum_{k=1}^{2^j} \min\left(|\theta_{j,k}|^2, \sigma_j^2(\alpha_j)\right) \\ &= \frac{\partial}{\partial \alpha_j} \left(\sum_{|\theta_{j,k}| \geq \sigma_j(\alpha_j)} \sigma_j^2(\alpha_j) \right) \\ &= \#\left\{|\theta_{j,k}| > \sigma_j(\alpha_j)\right\} \frac{d\sigma_j^2(\alpha_j)}{d\alpha_j}, \end{aligned} \tag{22}$$

with $\#\left\{|\theta_{j,k}| > \sigma_j(\alpha_j)\right\}$ the number of wavelet coefficients greater than the noise standard deviation $\sigma_j(\alpha_j)$ at scale j.

The noise variance at scale j is equal to

$$\sigma_j^2(\alpha_j) = \sum_{n=1}^{N} \frac{|H(f_n)|^2 |X(f_n)|^4 |\Psi_{j,k}(f_n)|^2}{(|H(f_n)|^2 |X(f_n)|^2 + \alpha_j \sigma^2)^2} \sigma^2. \tag{23}$$

Since the noise variance σ_j^2 is independent of the location k within a given scale, we can rewrite (23) as

$$\sigma_j^2(\alpha_j) = \frac{1}{2^j} \sum_{k=1}^{2^j} \sum_{n=1}^{N} \frac{|H(f_n)|^2 |X(f_n)|^4 |\Psi_{j,k}(f_n)|^2}{(|H(f_n)|^2 |X(f_n)|^2 + \alpha_j \sigma^2)^2} \sigma^2. \tag{24}$$

Differentiating $\sigma_j^2(\alpha_j)$ with respect to α_j gives

$$\frac{d\sigma_j^2(\alpha_j)}{d\alpha_j} = -\frac{1}{2^j} C(\alpha_j). \tag{25}$$

Combining (22) and (25) yields

$$\frac{\partial}{\partial \alpha_j} \left(\mathrm{O}\left(\alpha_{j_0}, \ldots, \alpha_J\right)\right) = -\frac{1}{2^j} \#\left\{|\theta_{j,k}| > \sigma_j(\alpha_j)\right\} C(\alpha_j). \tag{26}$$

Now, using (20) and (26), we have

$$\begin{aligned} \frac{\partial}{\partial \alpha_j} \left(\widetilde{\mathrm{MSE}}(\alpha_{\tilde{j}_0}, \alpha_{j_0}, \ldots, \alpha_{J-1})\right) &= \alpha_j\, C(\alpha_j) - \frac{1}{2^j} \#\left\{|\theta_{j,k}| > \sigma_j(\alpha_j)\right\} C(\alpha_j) \\ &= \left(\alpha_j - \frac{1}{2^j} \#\left\{|\theta_{j,k}| > \sigma_j(\alpha_j)\right\}\right) C(\alpha_j). \end{aligned} \tag{27}$$

Finally, setting (27) to zero, we obtain the optimal regularization parameter of (13).

Acknowledgements

We thank R. Nowak for many productive discussions. A big thanks is also due to J. Romberg for helping with the WaRD implementation.

References

[1] J. G. Proakis, *Digital Communications.* McGraw-Hill, 1995.

[2] A. K. Jain, *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1989.

[3] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems.* Washington, D.C.: V. H. Winston & Sons, 1977.

[4] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. 38, pp. 719–746, Mar. 1992.

[5] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[6] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data," in *Different Perspectives on Wavelets*, vol. 47 of *Proc. Symp. Appl. Math.*, pp. 173–205, American Mathematical Society, 1993.

[7] D. L. Donoho, "Nonlinear solution of linear inverse problems by Wavelet-Vaguellete Decomposition," *App. Comp. Harmonic Anal.*, vol. 2, pp. 101–126, 1995.

[8] A. D. Hillery and R. T. Chin, "Iterative Wiener filters for image restoration," *IEEE Trans. Signal Processing*, vol. 39, pp. 1892–1899, Aug. 1991.

[9] M. R. Banham and A. K. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. Image Processing*, vol. 5, pp. 619–634, April 1996.

[10] J. Kalifa, S. G. Mallat, and B. Rougé, "Minimax solution of inverse problems and deconvolution by mirror wavelet thresholding," in *SPIE conference on Wavelet applications in signal and image processing VII*, vol. 3813, (Denver,CO), pp. 42–57, Jul. 1999.

[11] R. D. Nowak and M. J. Thul, "Wavelet-Vaguelette restoration in photon-limited imaging," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '98*, (Seattle, WA), pp. 2869–2872, 1998.

[12] A. K. K. (Ed.), *Digital Image Restoration.* New York: Springer-Verlag, 1991.

[13] Y. Wan and R. Nowak, "A bayesian multiscale approach to joint image restoration and edge detection," in *SPIE conference on Wavelet applications in signal and image processing VII*, vol. 3813, (Denver,CO), Jul. 1999.

[14] K. R. Castleman, *Digital Image Processing.* New Jersey: Prentice Hall, 1996.

[15] S. J. Orfanidis, *Optimum Signal Processing, An Introduction.* New York, NY: Macmillan Publishing Company, 1984.

[16] H. L. G. M. R. Banham, N. P. Galatsanos and A. K. Katsaggelos, "Multichannel restoration of single channel images using a wavelet-based subband decompostion," *IEEE Trans. Image Processing*, vol. 3, pp. 821–833, Nov. 1994.

[17] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[18] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer.* Prentice-Hall, 1998.

[19] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *App. Comp. Harmonic Anal.*, vol. 1, pp. 100–115, Dec 1993.

[20] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Inform. Theory*, vol. 38, pp. 719–746, 1992.

[21] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2435–2476, Oct. 1998.

[22] D. L. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, pp. 1200–1224, Dec. 1995.

[23] A. Cohen and J. P. D'Ales, "Nonlinear approximation of random functions," *SIAM J. App. Math.*, vol. 57, pp. 518–540, Apr. 1997.

[24] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Royal Stat. Soc. B*, no. 59, pp. 319–351, 1997.

[25] R. Neelamani, "Wavelet-based deconvolution for ill-conditioned systems," M.S. thesis, Dept. of ECE, Rice Univ., in *www-dsp.rice.edu/publications/*, May 1999.

[26] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image processing and their relation," *IEEE Trans. Image Processing*, vol. 1, pp. 322–336, Jul. 1992.

[27] R. D. Nowak, "A fast wavelet-vaguelette algorithm for discrete LSI problems," tech. rep., Michigan State University, Aug. 1997.

[28] S. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empirical Wiener filtering," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 3169, pp. 389–399, 1997.

[29] H. Choi and R. G. Baraniuk, "Analysis of wavelet domain Wiener filters," in *IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, (Pittsburgh), Oct. 1998.

[30] R. Coifman and D. Donoho, "Translation invariant denoising," in *Wavelets and Statistics* (A. Antoniades, ed.), Springer, 1995.

[31] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, Apr. 1998. (Special Issue on Wavelets and Filter Banks).

[32] F. Abramovich and B. W. Silverman, "Wavelet decomposition approaches to statistical inverse problems," *Biometrika*, vol. 85, pp. 115–129, Oct. 1998.