

ForWaRD: Fourier-Wavelet Regularized Deconvolution for Ill-Conditioned Systems

Ramesh Neelamani, *Member, IEEE*, Hyeokho Choi, *Member, IEEE*, and Richard Baraniuk, *Fellow, IEEE*

Abstract—We propose an efficient, hybrid *Fourier-wavelet regularized deconvolution* (ForWaRD) algorithm that performs noise regularization via scalar shrinkage in both the Fourier and wavelet domains. The Fourier shrinkage exploits the Fourier transform’s economical representation of the colored noise inherent in deconvolution, whereas the wavelet shrinkage exploits the wavelet domain’s economical representation of piecewise smooth signals and images. We derive the optimal balance between the amount of Fourier and wavelet regularization by optimizing an approximate mean-squared error (MSE) metric and find that signals with more economical wavelet representations require less Fourier shrinkage. ForWaRD is applicable to all ill-conditioned deconvolution problems, unlike the purely wavelet-based *wavelet-vaguelette deconvolution* (WVD); moreover, its estimate features minimal ringing, unlike the purely Fourier-based Wiener deconvolution. Even in problems for which the WVD was designed, we prove that ForWaRD’s MSE decays with the optimal WVD rate as the number of samples increases. Further, we demonstrate that over a wide range of practical sample-lengths, ForWaRD improves on WVD’s performance.

Index Terms—Deblurring, deconvolution, restoration, wavelet-vaguelette, wavelets.

I. INTRODUCTION

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

A. Problem Statement

In this paper, we treat the classical discrete-time deconvolution problem. The problem setup and solutions are described in one dimension (1-D), but everything extends directly to higher

Manuscript received October 21, 2002; revised March 27, 2003. This work was supported by the National Science Foundation under Grant CCR-99-73188, the Air Force Office of Scientific Research under Grant F49620-01-1-0378, the Office of Naval Research under Grant N00014-02-1-0353, the Defense Advanced Research Projects Agency under Grant F30602-00-2-0557, and the Texas Instruments Leadership Universities Program. The associate editor coordinating the review of this paper and approving it for publication was Dr. Chong-Yung Chi.

R. Neelamani was with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005-1892 USA. He is now with the Upstream Research Company, ExxonMobil, Houston, TX 77252 USA.

The authors are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005-1892 USA (e-mail: choi@ece.rice.edu; richb@ece.rice.edu).

Digital Object Identifier 10.1109/TSP.2003.821103

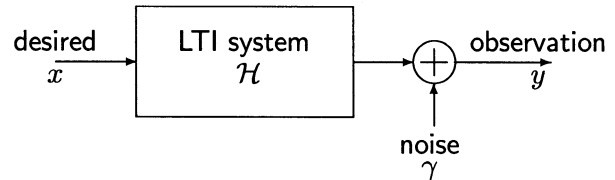


Fig. 1. Convolution model setup. The observation y consists of the desired signal x first degraded by the linear time-invariant (LTI) convolution system \mathcal{H} and then corrupted by zero-mean additive white Gaussian noise (AWGN) γ .

dimensions as well. The observed samples $y(n)$ consist of unknown desired signal samples $x(n)$ first degraded by circular convolution (denoted by \otimes) with a known impulse response $h(n)$ from a linear time-invariant (LTI) system \mathcal{H} and then corrupted by zero-mean additive white Gaussian noise (AWGN) $\gamma(n)$ with variance σ^2 (see Fig. 1)

$$\begin{aligned} y(n) &:= \mathcal{H}x(n) + \gamma(n), \quad n = 0, \dots, N-1 \\ &:= (h \otimes x)(n) + \gamma(n). \end{aligned} \quad (1)$$

Given y and h , we seek to estimate x .

A naive deconvolution estimate \tilde{x} is obtained using the operator inverse \mathcal{H}^{-1} as¹

$$\tilde{x}(n) := \mathcal{H}^{-1}y(n) = x(n) + \mathcal{H}^{-1}\gamma(n). \quad (2)$$

Unfortunately, the variance of the colored noise $\mathcal{H}^{-1}\gamma$ in \tilde{x} is large when \mathcal{H} is ill conditioned. In such a case, the mean-squared error (MSE) between x and \tilde{x} is large, making \tilde{x} an unsatisfactory deconvolution estimate.

In general, deconvolution algorithms can be interpreted as estimating x from the noisy signal \tilde{x} in (2). In this paper, we focus on simple and fast estimation based on *scalar shrinkage* of individual components in a suitable transform domain. Such a focus is not restrictive because transform-domain scalar shrinkage lies at the core of many traditional [3], [4] and modern [2], [5] deconvolution approaches.

B. Transform-Domain Shrinkage

Given an orthonormal basis $\{b_k\}_{k=0}^{N-1}$ for \mathbb{R}^N , the naive estimate \tilde{x} from (2) can be expressed as

$$\tilde{x} = \sum_{k=0}^{N-1} (\langle x, b_k \rangle + \langle \mathcal{H}^{-1}\gamma, b_k \rangle) b_k. \quad (3)$$

¹For noninvertible \mathcal{H} , we replace \mathcal{H}^{-1} by its pseudo-inverse and x by its orthogonal projection onto the range of \mathcal{H} in (2) [2]. The estimate \tilde{x} in (2) continues to retain all of the information that y contains about x .

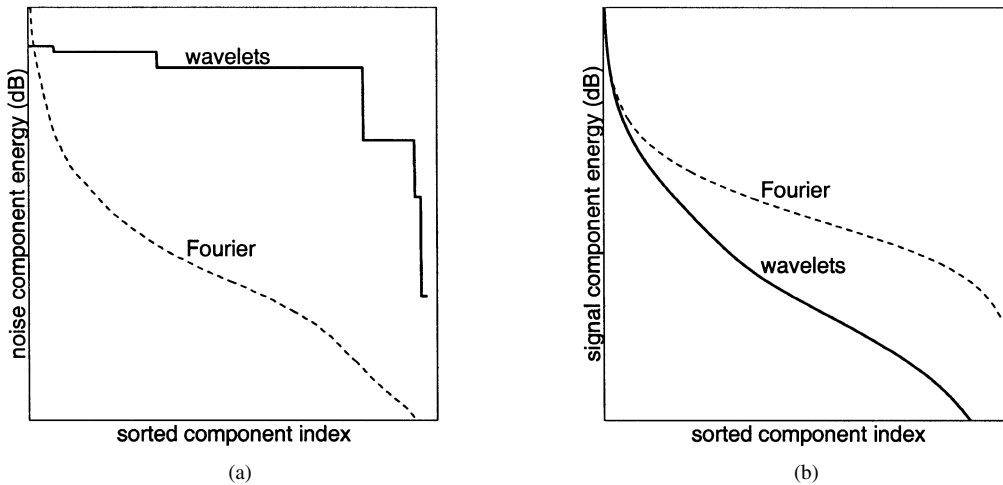


Fig. 2. Economy of Fourier versus wavelet representations. (a) Energies in decibels (dB) of the Fourier and wavelet components of the noise $\mathcal{H}^{-1}\gamma$ colored by the pseudo-inverse of a 2-D 9×9 box-car smoothing operator. The components are sorted in descending order of energy from left to right. The colored noise energy is concentrated in fewer Fourier components than wavelet components. (b) Energies of the Fourier and wavelet components of the *Cameraman* image x . The signal energy is concentrated in fewer wavelet components than Fourier components.

An improved estimate \tilde{x}_λ can be obtained by simply shrinking the k th component in (3) with a scalar λ_k , $0 \leq \lambda_k \leq 1$ [6]:

$$\tilde{x}_\lambda := \sum_{k=0}^{N-1} (\langle x, b_k \rangle + \langle \mathcal{H}^{-1}\gamma, b_k \rangle) \lambda_k b_k \quad (4)$$

$$=: x_\lambda + \mathcal{H}^{-1}\gamma_\lambda. \quad (5)$$

The $x_\lambda := \sum_k \langle x, b_k \rangle \lambda_k b_k$ denotes the *retained part* of the signal x that the shrinkage preserves from (2), whereas $\mathcal{H}^{-1}\gamma_\lambda := \sum_k \langle \mathcal{H}^{-1}\gamma, b_k \rangle \lambda_k b_k$ denotes the *leaked part* of the colored noise $\mathcal{H}^{-1}\gamma$ that the shrinkage fails to attenuate. Clearly, we should set $\lambda_k \approx 1$ if the variance $\sigma_k^2 := \mathbb{E}(|\langle \mathcal{H}^{-1}\gamma, b_k \rangle|^2)$ of the k th colored noise component is small relative to the energy $|\langle x, b_k \rangle|^2$ of the corresponding signal component and set $\lambda_k \approx 0$ otherwise. The shrinkage by λ_k can also be interpreted as a form of *regularization* for the deconvolution inverse problem [4].

The tradeoff associated with the choice of λ_k is easily understood: If $\lambda_k \approx 1$, then most of the k th colored noise component leaks into \tilde{x}_λ with the corresponding signal component; the result is a distortion-free but noisy estimate. In contrast, if $\lambda_k \approx 0$, then most of the k th signal component is lost with the corresponding colored noise component; the result is a noise-free but distorted estimate. Since the variance of the leaked noise $\mathcal{H}^{-1}\gamma_\lambda$ in (5) and the energy of the lost signal $x - x_\lambda$ comprise the MSE of the shrunk estimate \tilde{x}_λ , judicious choices of the λ_k 's help lower the estimate's MSE.

However, an important fact is that for a given transform domain, even with the best possible λ_k 's, the estimate \tilde{x}_λ 's MSE is lower bounded by [5], [7], [8]

$$\frac{1}{2} \sum_{k=0}^{N-1} \min(|\langle x, b_k \rangle|^2, \sigma_k^2). \quad (6)$$

From (6), \tilde{x}_λ has small MSE only when most of the signal energy ($= \sum_k |\langle x, b_k \rangle|^2$) and colored noise energy ($= \sum_k \sigma_k^2$) is captured by just a few transform-domain coefficients—we term such a representation *economical*—and when the energy-capturing coefficients for the signal and noise are different. Other-

wise, the \tilde{x}_λ is either excessively noisy due to leaked noise components or distorted due to lost signal components.

Traditionally, the Fourier domain (with sinusoidal b_k 's) is used to estimate x from \tilde{x} . For example, the LTI Wiener deconvolution filter corresponds to (4) with each λ_k determined by the k th component's signal-to-noise ratio [3], [4]. The strength of the Fourier basis is that it most economically represents the colored noise $\mathcal{H}^{-1}\gamma$ [see Fig. 2(a) and Section III-B for details]. However, the weakness of the Fourier domain is that it does not economically represent signals x with singularities such as images with edges [see Fig. 2(b)]. Consequently, as dictated by the MSE bound in (6), any estimate obtained via Fourier shrinkage is unsatisfactory with a large MSE; the estimate is either noisy or distorted for signals x with singularities [see Fig. 4(c), for example].

Recently, the wavelet domain (with shifts and dilates of a mother wavelet function as b_k 's) has been exploited to estimate x from \tilde{x} , for example, Donoho's *wavelet-vaguelette deconvolution* (WVD) [5]. The strength of the wavelet domain is that it economically represents classes of signals containing singularities that satisfy a wide variety of local smoothness constraints, including piecewise smoothness and Besov space smoothness [see Fig. 2(b) and Section V-B for details]. However, the weakness of the wavelet domain is that it typically does not economically represent the colored noise $\mathcal{H}^{-1}\gamma$ [see Fig. 2(a)]. Consequently, as dictated by the MSE bound (6), any estimate obtained via wavelet shrinkage is unsatisfactory with a large MSE; the estimate is either noisy or distorted for many types of \mathcal{H} .

Unfortunately, no single transform domain can economically represent both the noise colored by a general \mathcal{H}^{-1} and signals from a general smoothness class [5]. Hence, deconvolution techniques employing shrinkage in a single transform domain cannot yield adequate estimates in many deconvolution problems of interest.

C. Fourier-Wavelet Regularized Deconvolution (ForWaRD)

In this paper, we propose a deconvolution scheme that relies on *tandem* scalar processing in both the Fourier domain,

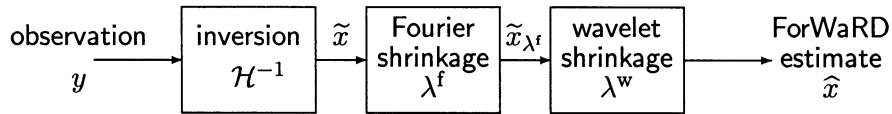


Fig. 3. Fourier-wavelet regularized deconvolution (ForWaRD). ForWaRD employs a small amount of Fourier shrinkage (most $\lambda_k^f \approx 1$) to partially attenuate the noise amplified during operator inversion. Subsequent wavelet shrinkage (determined by λ^w) effectively attenuates the residual noise.

which economically represents the colored noise $\mathcal{H}^{-1}\gamma$, and the wavelet domain, which economically represents signals x from a wide variety of smoothness classes. Our hybrid *Fourier-Wavelet Regularized Deconvolution* (ForWaRD) technique estimates x from \tilde{x} by first employing a small amount of scalar Fourier shrinkage λ^f and then attenuating the leaked noise with scalar wavelet shrinkage λ^w (see Fig. 3) [9], [10].

Here is how it works: During operator inversion, some Fourier coefficients of the noise γ are significantly amplified; just a small amount of Fourier shrinkage (most $\lambda_k^f \approx 1$) is sufficient to attenuate these amplified Fourier noise coefficients with minimal loss of signal components. The leaked noise $\mathcal{H}^{-1}\gamma_{\lambda^f}$ that Fourier shrinkage λ^f fails to attenuate [see (5)] has significantly reduced energy in all wavelet coefficients, but the signal part x_{λ^f} that Fourier shrinkage retains continues to be economically represented in the wavelet domain. Hence, subsequent wavelet shrinkage effectively extracts the retained signal x_{λ^f} from the leaked noise $\mathcal{H}^{-1}\gamma_{\lambda^f}$ and provides a robust estimate.

For an idealized ForWaRD system, we will derive the optimal balance between the amount of Fourier shrinkage and wavelet shrinkage by optimizing over an approximate MSE metric. We will find that signals with more economical wavelet representations require less Fourier shrinkage.

Fig. 4 illustrates the superior overall visual quality and lower MSE of the ForWaRD estimate as compared with the LTI Wiener filter estimate [3], [4] for the 2-D box-car blur operator, which models rectangular scanning aperture effects [1], with impulse response $h(n_1, n_2) = 1/81$ for $0 \leq n_1, n_2 \leq 8$ and 0 otherwise (see Section VIII for details). For this operator, the WVD approach returns an essentially zero estimate; scalar wavelet shrinkage cannot salvage the signal components since nearly all wavelet coefficients are corrupted with high-variance noise.

Indeed, even in problems for which the WVD was designed, we will prove that the ForWaRD MSE also decays with the same optimal WVD rate as the number of samples increases. Further, for such problems, we will experimentally demonstrate ForWaRD's superior MSE performance compared with the WVD over a wide range of practical sample sizes [see Fig. 6(a)].

D. Related Work

Kalifa and Mallat have proposed a mirror-wavelet basis approach that is similar to the WVD but employs scalar shrinkage in a mirror-wavelet domain adapted to the colored noise $\mathcal{H}^{-1}\gamma$ instead of shrinkage in the conventional wavelet domain [2]. Although the adapted basis improves on the WVD performance in some “hyperbolic” deconvolution problems, similarly to the WVD, it provides inadequate estimates for arbitrary convolution operators. For example, for the ubiquitous box-car blur \mathcal{H} , again, most signal components are lost during scalar shrinkage due to high-variance noise. Fig. 7(b) illustrates that ForWaRD

is competitive with the mirror-wavelet approach, even for a hyperbolic deconvolution problem.

Similar to ForWaRD, Nowak and Thul [11] have first employed an under-regularized system inverse and subsequently used wavelet-domain signal estimation. However, they do not address the issue of optimal regularization or asymptotic performance.

Banham and Katsaggelos have applied a multiscale Kalman filter to the deconvolution problem [12]. Their approach employs an under-regularized, constrained-least-squares prefilter to reduce the support of the state vectors in the wavelet domain, thereby improving computational efficiency. 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, ForWaRD employs simple Wiener or Tikhonov regularization in the Fourier domain to optimize the MSE performance. In addition, ForWaRD employs simple scalar shrinkage on the wavelet coefficients in contrast to more complicated prediction on edge and nonedge quad-trees [12]. Consequently, as discussed in Section VI-D, ForWaRD demonstrates excellent MSE performance as the number of samples tends to infinity and is, in fact, asymptotically optimal in certain cases. Further, as demonstrated in Section VIII, ForWaRD yields better estimates than the multiscale Kalman filter approach.

There exists a vast literature on iterative deconvolution techniques; see [4], [13]–[15], and the references therein. In this paper, we focus exclusively on noniterative techniques for the sake of implementation speed and simplicity. Nevertheless, many iterative techniques could exploit the ForWaRD estimate as a seed to initialize their iterations; for example, see [16].

E. Paper Organization

We begin by providing a more precise definition of the convolution setup (1) in Section II. We then discuss techniques that employ scalar Fourier shrinkage in Section III. After briefly reviewing wavelet theory in Section IV, we introduce the WVD technique in Section V. We present the hybrid ForWaRD scheme in Section VI and discuss its practical implementation in Section VII. Illustrative examples lie in Section VIII. We conclude and sketch future directions in Section IX. A short WVD review in Appendix A and technical proofs in Appendices B–D complete the paper.

II. SAMPLING AND DECONVOLUTION

Most real-life deconvolution problems originate in continuous time and are then sampled. In this section, we sketch the relationship between such a sampled continuous-time setup and the setup with discrete-time circular convolution considered in this paper [see (1)].



Fig. 4. (a) Desired *Cameraman* image x (256×256 samples). (b) Observed image y : x smoothed by a two-dimensional (2-D) 9×9 box-car blur plus white Gaussian noise with variance such that the BSNR = 40 dB. (c) LTI Wiener filter estimate (SNR= 20.8 dB, ISNR = 5.6 dB). (d) ForWaRD estimate (SNR = 22.5 dB, ISNR = 7.3 dB). See Section VIII for further details.

Consider the following sampled continuous-time deconvolution setup: An unknown finite-energy desired signal $x(t)$ is blurred by linear convolution (denoted by $*$) with the known finite-energy impulse response $h(t)$ of an LTI system and then corrupted by an additive Gaussian process $\gamma(t)$ to form the observation $z(t) = (h * x)(t) + \gamma(t)$. For finite-support $x(t)$ and $h(t)$, the finite-support $(h * x)(t)$ can be obtained using circular convolution with a sufficiently large period. For infinite-support $x(t)$ and $h(t)$, the approximation of $(h * x)(t)$ using circular convolution can be made arbitrarily precise by increasing the period. Hence, we assume that the observation $z(t)$ over a normalized unit interval can be obtained using circular convolution with a unit period, that is, $z(t) := (h \otimes x)(t) + \gamma(t)$ with $t \in [0, 1)$. Deconvolution aims to estimate $x(t)$ from the samples $z(n)$ of the continuous-time observation $z(t)$. For example,

$z(n)$ can be obtained by averaging $z(t)$ over uniformly spaced intervals of length $1/N$

$$z(n) := N \int_{n/N}^{(n+1)/N} z(t) dt, \quad n = 0, \dots, N-1. \quad (7)$$

Other sampling kernels can also be used in (7);² see [17] and [18] for excellent tutorials on sampling. Such a setup encapsulates many real-life deconvolution problems [1].

The observation samples $z(n)$ from (7) can be closely approximated by the observation $y(n)$ from setup (1) [1]; that is

$$z(n) \approx y(n) = (h \otimes x)(n) + \gamma(n), \quad n = 0, \dots, N-1 \quad (8)$$

if the continuous-time variables $x(t)$, $\gamma(t)$, and $h(t)$ comprising $z(n)$ are judiciously related to the discrete variables

²For example, impulse sampling samples at uniformly spaced time instants $t_n = n/N$ to yield $z(n) = (t_n)$.

$x(n)$, $\gamma(n)$, and $h(n)$ comprising $y(n)$. We choose to define $\gamma(n) := N \int_{n/N}^{(n+1)/N} \gamma(t) dt$. The $\gamma(n)$ so defined can be assumed to be AWGN samples with nonzero variance σ^2 ; the large bandwidths of noise processes such as thermal noise justify the whiteness assumption [1]. Let $\mathcal{B}_{1/N} x(t)$ and $\mathcal{B}_{1/N} h(t)$ denote signals obtained by first making $x(t)$ and $h(t)$ periodic and then bandlimiting the resulting signals' Fourier series to the frequency $1/2N$ Hz (for antialiasing). We define $x(n) := N \int_{n/N}^{n+1/N} \mathcal{B}_{1/N} x(t) dt$ and define $h(n)$ as uniformly spaced (over $t \in [0, 1)$) impulse samples of $(1/N)\mathcal{B}_{1/N} h(t)$. With these definitions, we can easily show that the error $|z(n) - y(n)| \leq \|x(t) - \mathcal{B}_{1/N} x(t)\|_2 \|h(t) - \mathcal{B}_{1/N} h(t)\|_2$. For all finite-energy $x(t)$ and $h(t)$, both $\|x(t) - \mathcal{B}_{1/N} x(t)\|_2$ and $\|h(t) - \mathcal{B}_{1/N} h(t)\|_2$ decay to zero with increasing N because they represent the norm of the aliasing components of $x(t)$ and $h(t)$, respectively. Consequently, $|z(n) - y(n)|$ soon becomes negligible with respect to the noise variance σ^2 and can be ignored. Hence, solutions to estimate $x(n)$ from the $y(n)$ in (1)—the focus of this paper—can be directly applied to estimate $x(n)$ from $z(n)$. For a wide range of Besov space signals, the estimate of $x(n)$ can then be interpolated with minimal error to yield a continuous-time estimate of $x(t)$, as sought in (7) [19], and [20].³

In Sections V and VI, we will analyze the MSE decay rate (in terms of N) of the WVD and ForWARD solutions to the setup (1) as the number of samples $N \rightarrow \infty$. At each N , we assume that the corresponding $x(n)$ and $h(n)$ in (1) originate from an underlying continuous-time $x(t)$ and $h(t)$, as defined above. Further, we assume that the corrupting $\gamma(n)$ in (1) are AWGN samples with variance $\sigma^2 > 0$ that is invariant with N .

III. FOURIER-BASED REGULARIZED DECONVOLUTION (FoRD)

A. Framework

The Fourier domain is the traditional choice for deconvolution [4] because convolution simplifies to scalar Fourier operations. That is, (1) can be rewritten as

$$Y(f_k) = H(f_k)X(f_k) + \Gamma(f_k) \quad (9)$$

where Y , H , X , and Γ are the respective length- N discrete Fourier transforms (DFTs) of y , h , x , and γ , and $f_k := \pi k/N$, $k = -N/2 + 1, \dots, N/2$ (assuming N is even) are the normalized DFT frequencies. Rewriting the pseudo-inversion operation [see (2)] in the Fourier domain

$$\tilde{X}(f_k) := \begin{cases} X(f_k) + \frac{\Gamma(f_k)}{H(f_k)}, & \text{if } |H(f_k)| > 0, \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where \tilde{X} is the DFT of \tilde{x} , clearly demonstrates that noise components where $|H(f_k)| \approx 0$ are particularly amplified during operator inversion.

³The Besov space range is dictated by the smoothness of the sampling kernel. Let $x(t) \in$ Besov space $B_{p,q}^s$ (see Section IV-B for the notation). Then, if the sampling kernel of (7) is employed, then the interpolation error is negligible with respect to the estimation error for the range $s > 1/p - 1/2$; the range decreases to $s > 1/p$ if impulse sampling is employed [19], [20].

Deconvolution via Fourier shrinkage, which we call *Fourier-based Regularized Deconvolution* (FoRD), attenuates the amplified noise in \tilde{X} with shrinkage

$$\lambda_k^f = \frac{|H(f_k)|^2}{|H(f_k)|^2 + \Lambda(f_k)}. \quad (11)$$

The $\Lambda(f_k) \geq 0$, commonly referred to as *regularization terms* [4], [21], control the amount of shrinkage. The DFT components of the FoRD estimate \tilde{x}_{λ^f} are given by

$$\begin{aligned} \tilde{X}_{\lambda^f}(f_k) &:= \tilde{X}(f_k) \lambda_k^f \\ &= X(f_k) \left(\frac{|H(f_k)|^2}{|H(f_k)|^2 + \Lambda(f_k)} \right) \\ &\quad + \frac{\Gamma(f_k)}{H(f_k)} \left(\frac{|H(f_k)|^2}{|H(f_k)|^2 + \Lambda(f_k)} \right) \\ &=: X_{\lambda^f}(f_k) + \frac{\Gamma_{\lambda^f}(f_k)}{H(f_k)}. \end{aligned} \quad (12)$$

The X_{λ^f} and Γ_{λ^f}/H comprising \tilde{X}_{λ^f} denote the respective DFTs of the retained signal x_{λ^f} and leaked noise $\mathcal{H}^{-1}\gamma_{\lambda^f}$ components that comprise the FoRD estimate \tilde{x}_{λ^f} [see (5)]. Typically, the operator inversion in (10) and shrinkage in (12) are performed simultaneously to avoid numerical instabilities.

Different FoRD techniques, such as LTI Wiener deconvolution [3], [4] and Tikhonov-regularized deconvolution [21], differ in their choice of shrinkage λ^f in (12). LTI Wiener deconvolution sets

$$\lambda_k^f = \frac{|H(f_k)|^2}{|H(f_k)|^2 + \alpha \frac{N\sigma^2}{|X(f_k)|^2}} \quad (13)$$

with *regularization parameter* $\alpha = 1$ to shrink more (that is, $\lambda_k^f \approx 0$) at frequencies where the signal power $|X(f_k)|^2$ is small [3], [4]. Tikhonov-regularized deconvolution, which is similar to LTI Wiener deconvolution assuming a flat signal spectrum $|X(f_k)|^2$, sets

$$\lambda_k^f = \frac{|H(f_k)|^2}{|H(f_k)|^2 + \tau} \quad (14)$$

with $\tau > 0$ [21]. Later, in Section VI, we will put both of these shrinkage techniques to good use.

B. Strengths of FoRD

The Fourier domain provides the most economical representation of the colored noise $\mathcal{H}^{-1}\gamma$ in (2) because the Fourier transform acts as the Karhunen-Loeve transform [22] and decorrelates the noise $\mathcal{H}^{-1}\gamma$. Consequently, among all linear transformations, the Fourier transform captures the maximum colored noise energy using a fixed number of coefficients [23]. This economical noise representation enhances FoRD performance because the total FoRD MSE is lower bounded by $1/(2N) \sum_k \min(|X(f_k)|^2, (N\sigma^2/|H(f_k)|^2))$ [5].⁴ The best possible FoRD MSE

$$1/N \sum_k (N\sigma^2 |X(f_k)|^2 / (|H(f_k)|^2 |X(f_k)|^2 + N\sigma^2))$$

is achieved using the LTI Wiener deconvolution shrinkage λ^f of (13) in (12) [7]. When the signal x in (2) also enjoys an econom-

⁴The factor N arises because $\sum_k |X(f_k)|^2 = N \sum_k |x(k)|^2$ for any signal x .

ical Fourier-domain representation (that is, when x is “smooth” and thus has rapidly decaying Fourier coefficients [7]), FoRD can provide excellent deconvolution estimates. For example, FoRD provides optimal estimates for signals x in L_2 -Sobolev smoothness spaces [5].

C. Limitations of FoRD

Unfortunately, the Fourier domain does not provide economical representations for signals with singularities, such as images with edges, because the energy of the singularities spreads over many Fourier coefficients. Consequently, even with the best scalar Fourier shrinkage, the FoRD MSE is unsatisfactory, as dictated by the lower bound in (6). The estimation error becomes apparent in the form of distortions such as ringing around edge singularities [see Fig. 4(c)].

IV. BACKGROUND ON WAVELETS

In contrast to Fourier representations, wavelets provide economical representations for a diverse class of signals including signals with singularities such as images [7], [24].

A. Wavelet Transform

The discrete wavelet transform (DWT) represents a 1-D continuous-time signal $x(t)$, $t \in [0, 1]$ in terms of shifted versions of a lowpass scaling function ϕ and shifted and dilated versions of a prototype band-pass wavelet function ψ [7]. For special choices of ϕ and ψ , the functions $\psi_{j,\ell}(t) := 2^{j/2} \psi(2^j t - \ell)$ and $\phi_{j,\ell}(t) := 2^{j/2} \phi(2^j t - \ell)$ with $j, \ell \in \mathbb{Z}$ form an orthonormal basis. The j parameter corresponds to the *scale* of the analysis, whereas the ℓ parameter corresponds to the *location*. A finite-resolution approximation x^J to x is given by

$$x^J(t) = \sum_{\ell=0}^{N_{j_0}-1} s_{j_0,\ell} \phi_{j_0,\ell}(t) + \sum_{j=j_0}^J \sum_{\ell=0}^{N_j-1} w_{j,\ell} \psi_{j,\ell}(t)$$

with the scaling coefficients $s_{j_0,\ell} := \langle x, \phi_{j_0,\ell} \rangle$ and wavelet coefficients $w_{j,\ell} := \langle x, \psi_{j,\ell} \rangle$. The parameter J controls the resolution of the wavelet reconstruction x^J of x . In fact, the L_2 norm $\|x^J - x\|_2 \rightarrow 0$ as $J \rightarrow \infty$.

For a discrete-time signal with N samples, the N wavelet coefficients can be efficiently computed in $O(N)$ operations using a filterbank consisting of lowpass filters, highpass filters, up-samplers, and decimators [7]. For periodic signals, which are natural when analyzing circular convolution, filterbanks implementing circular convolution are employed. Multidimensional DWTs are computed by wavelet-transforming alternately along each dimension [7], [19].

Purely for notational convenience, we henceforth discuss processing of only the wavelet coefficients. However, all steps are replicated on the scaling coefficients as well.

B. Economy of Wavelet Representations

Wavelets provide economical representations for signals in smoothness spaces such as *Besov spaces* [8]. Roughly speaking, a Besov space $B_{p,q}^s$ contains functions with “ s derivatives in L_p ” with q measuring finer smoothness distinctions [24]. Besov spaces with different s , p , and q characterize many classes of

signals in addition to L_2 -Sobolev space signals; for example, in 1-D, $B_{1,1}^1$ contains piece-wise polynomial signals [7], [25]. Further, unlike L_2 -Sobolev spaces, Besov spaces also contain images with edges [24]. The wavelet coefficients computed from samples (refer Section II) of a continuous-time 1-D signal $x(t) \in B_{p,q}^s$, $s > 1/p - 1/2$, $1 \leq p, q \leq \infty$ satisfy (for all N)

$$\frac{1}{\sqrt{N}} \left(\sum_{j \geq j_0} 2^{jq(s+1/2-1/p)} \left(\sum_{\ell} |w_{j,\ell}|^p \right)^{q/p} \right)^{1/q} < \infty \quad (15)$$

assuming sufficiently smooth wavelet basis functions [5], [19], [20].⁵ The condition for higher dimensional Besov space signals is a straightforward extension of (15) [5], [19]. From (15), we can infer that the wavelet coefficients of Besov space signals decay exponentially fast with increasing scale j .

C. Wavelet Shrinkage-Based Signal Estimation

The wavelet transform’s economical signal representation facilitates an effective solution to the problem of estimating the signal $x(n)$ from AWGN-corrupted observations [19], [20], [26], [27]

$$\tilde{x}(n) = x(n) + \gamma(n). \quad (16)$$

Such a setup is similar to estimating x from (2) but with an identity operator \mathcal{H} . Simple shrinkage in the wavelet domain with scalars λ^w can provide excellent estimates of x . This shrinkage is illustrated by (4) with wavelet basis functions as the b_k ’s and with identity \mathcal{H} .

Oracle thresholding shrinks with

$$\lambda_{j,\ell}^w = \begin{cases} 1, & \text{if } |w_{j,\ell}| > \sigma_j \\ 0, & \text{if } |w_{j,\ell}| \leq \sigma_j \end{cases} \quad (17)$$

where σ_j^2 is the noise variance at wavelet scale j . Oracle thresholding provides excellent estimation results [27] but is impractical because it assumes knowledge of the wavelet coefficients $w_{j,\ell}$ of the desired x . *Hard thresholding*, which closely approaches oracle thresholding’s performance and is also practical [20], employs

$$\lambda_{j,\ell}^w = \begin{cases} 1, & \text{if } |\tilde{w}_{j,\ell}| > \rho_j \sigma_j \\ 0, & \text{if } |\tilde{w}_{j,\ell}| \leq \rho_j \sigma_j \end{cases} \quad (18)$$

where $\tilde{w}_{j,\ell} := \langle \tilde{x}, \psi_{j,\ell} \rangle$, and ρ_j is a scale-dependent threshold factor (see [7, p. 442] for choices of ρ_j). When the underlying continuous-time $x(t) \in B_{p,q}^s$ with $s > 1/p - 1/2$ and $1 \leq p, q \leq \infty$, both oracle and hard thresholding (with judiciously chosen ρ_j [26]) provide estimates whose MSE-per-sample decays at least as fast as $N^{-2s/(2s+1)}$ with increasing number of samples $N \rightarrow \infty$ [19], [20]. Further, no estimator can achieve a better error decay rate for every $x(t) \in B_{p,q}^s$. If the threshold factor ρ_j is chosen to be scale-independent, then the MSE decay rate is decelerated by an additional $\log N$ factor.

In practice, the *Wavelet-domain Wiener Filter* (WWF) improves on the MSE performance of hard thresholding by

⁵The traditional Besov space characterizing equation in [5], [19], [20] assumes L_2 -normalized wavelet coefficients $w_{j,\ell}$, that is, $\sum_{j,\ell} |w_{j,\ell}|^2 = \|x(t)\|_2^2$. Because the $w_{j,\ell}$ used in (15) are computed using signal samples $x(n)$ that satisfy $\sum_{j,\ell} |w_{j,\ell}|^2 = \sum_n |x(n)|^2 \approx N \|x(t)\|_2^2$, a normalization factor of \sqrt{N} appears.

employing Wiener estimation on each wavelet coefficient [28]. WWF chooses

$$\lambda_{j,\ell}^w = \frac{|w_{j,\ell}|^2}{|w_{j,\ell}|^2 + \sigma_j^2}. \quad (19)$$

However, like in oracle thresholding, the coefficients $w_{j,\ell}$ required to construct the $\lambda_{j,\ell}^w$ are unknown. Hence, a ‘‘pilot’’ estimate of the unknown signal is first computed using hard thresholding (with, say, $\rho_j = 3$ for 256×256 images). Then, using λ^w constructed with the pilot estimate’s wavelet coefficients in (19), WWF shrinkage is performed. Sufficiently different wavelet basis functions must be used in the two steps [28].

V. WAVELET-VAGUELETTE DECONVOLUTION (WVD)

A. Framework

The wavelet-vaguelette decomposition algorithm leverages wavelets’ economical signal representation to solve some special linear inverse problems [5]. With a slight abuse of nomenclature, we will refer to the wavelet-vaguelette decomposition algorithm applied to deconvolution as *wavelet-vaguelette deconvolution* (WVD).

In contrast to FoRD, the WVD algorithm conceptually extracts the signal x from \tilde{x} in (2) with scalar wavelet shrinkage λ^w such as hard thresholding [5], [27] to yield an estimate \tilde{x}_{λ^w} . For the reader’s convenience, we provide a simple review of the WVD algorithm in Appendix A.

B. Strengths of WVD

The wavelet domain provides economical representations for a wide variety of signals x in (2). In fact, among all orthogonal transforms, the wavelet transform can capture the maximum (within a constant factor) signal energy using any fixed number of coefficients for the worst-case Besov space signal [8]. This economical signal representation enhances WVD’s performance because the total WVD MSE can be bounded within a constant factor of $\sum_{j,\ell} \min(|w_{j,\ell}|^2, \sigma_j^2)$, where σ_j^2 is the wavelet-domain colored noise variance. When the colored noise $\mathcal{H}^{-1}\gamma$ in (2) also enjoys an economical wavelet-domain representation, the WVD can provide excellent deconvolution estimates. For example, consider a ‘‘scale-invariant’’ operator \mathcal{H} with frequency response $|H(f_k)| \propto (|k| + 1)^{-\nu}$, $\nu > 0$. Such an \mathcal{H} yields colored noise $\mathcal{H}^{-1}\gamma$ that is nearly diagonalized by the wavelet transform [2], [5] and is hence economically represented in the wavelet domain [5]. For such operators, the per-sample MSE of the WVD estimate \tilde{x}_{λ^w} decays rapidly with increasing number of samples N as [2], [5], [29]

$$\frac{1}{N} \mathbb{E} \left(\sum_n |x(n) - \tilde{x}_{\lambda^w}(n)|^2 \right) \leq C N^{-2s/(2s+2\nu+1)} \quad (20)$$

where $C > 0$ is a constant. Further, no estimator can achieve better error decay rates for every $x(t) \in B_{p,q}^s$.

C. Limitations of WVD

Unfortunately, the WVD is designed to deconvolve only the very limited class of *scale-invariant* operators [5]. For other \mathcal{H} , the colored noise $\mathcal{H}^{-1}\gamma$ in (2) is not economically represented in the wavelet domain. For example, with the uniform box-car blur \mathcal{H} , the components of the colored noise $\mathcal{H}^{-1}\gamma$ corrupting most wavelet coefficients have extremely high variance due to zeros in H . Consequently, even with the best scalar wavelet shrinkage, the WVD MSE is unsatisfactory, as dictated by the lower bound in (6). Indeed, wavelet shrinkage will set most of the signal wavelet coefficients to zero when estimating x from \tilde{x} in (2) and yield an unsatisfactory, essentially zero estimate.

VI. FOURIER-WAVELET REGULARIZED DECONVOLUTION (FORWARD)

The hybrid ForWaRD algorithm estimates x from \tilde{x} in (2) by employing scalar shrinkage both in the Fourier domain to exploit its economical colored noise representation and in the wavelet domain to exploit its economical signal representation. The hybrid approach is motivated by our realization that shrinkage in a *single* transform domain cannot yield good estimates in many deconvolution problems. This is because no single transform domain can economically represent both the colored noise $\mathcal{H}^{-1}\gamma$ with arbitrary \mathcal{H} and signals x with arbitrary smoothness [5]. By adopting a hybrid approach, ForWaRD overcomes this limitation and provides robust solutions to a wide class of deconvolution problems.

A. ForWaRD Algorithm

The ForWaRD algorithm consists of the following steps (see Fig. 3).

1a) Operator inversion

Obtain Y and H by computing the DFTs of y and h . Then, invert \mathcal{H} to obtain \tilde{X} as in (10).

1b) Fourier shrinkage

Shrink \tilde{X} with scalars λ^f [using (13) or (14)] to obtain \tilde{X}_{λ^f} , as in (12). Compute the inverse DFT of \tilde{X}_{λ^f} to obtain \tilde{x}_{λ^f} .

2) Wavelet shrinkage

Compute the DWT of the still noisy \tilde{x}_{λ^f} to obtain $\tilde{w}_{j,\ell;\lambda^f}$. Shrink $\tilde{w}_{j,\ell;\lambda^f}$ with $\lambda_{j,\ell}^w$ [using (18) or (19)] to obtain $\hat{w}_{j,\ell} := \tilde{w}_{j,\ell;\lambda^f} \lambda_{j,\ell}^w$. Compute the inverse DWT with the $\hat{w}_{j,\ell}$ to obtain the ForWaRD estimate \hat{x} .

For numerical robustness, the operator inversion in Step 1a and Fourier shrinkage in Step 1b are performed simultaneously.

B. How ForWaRD Works

During operator inversion in Step 1a of the ForWaRD algorithm, some Fourier noise components are significantly amplified [see (10)]. In Step 1b, ForWaRD employs a small amount of Fourier shrinkage (most $\lambda_k^f \approx 1$; $\lambda_k^f \approx 0$ only when $|H(f_k)| \approx 0$) by choosing a small value for the regularization $\Lambda(f_k)$ that determines the λ^f in (11). Sections VI-C and VII-B contain details on the choice of λ^f . This minimal shrinkage is sufficient to significantly attenuate the amplified noise components with a minimal loss of signal components. Consequently, after the Fourier shrinkage step [see (12)], the leaked noise $\mathcal{H}^{-1}\gamma_{\lambda^f}$ in

the \tilde{x}_{λ^f} has substantially reduced variances $\sigma_{j;\lambda^f}^2$ in all wavelet coefficients. The variance $\sigma_{j;\lambda^f}^2$ at wavelet scale j is given by

$$\begin{aligned}\sigma_{j;\lambda^f}^2 &:= \mathbb{E}(|\mathcal{H}^{-1}\gamma_{\lambda^f}, \psi_{j,\ell}|^2) \\ &= \sum_{k=-(N/2)+1}^{N/2} \frac{\sigma^2 |\Psi_{j,\ell}(f_k)|^2}{|H(f_k)|^2} |\lambda_k^f|^2 \\ &= \sum_{k=-(N/2)+1}^{N/2} \frac{\sigma^2 |H(f_k) \Psi_{j,\ell}(f_k)|^2}{(|H(f_k)|^2 + \Lambda(f_k))^2}\end{aligned}\quad (21)$$

where $\Psi_{j,\ell}$ is the DFT of $\psi_{j,\ell}$. The retained signal part x_{λ^f} in \tilde{x}_{λ^f} continues to be represented economically in the wavelet domain because x_{λ^f} lies in the same Besov space as the desired signal x (see Appendix C-A for the justification). Therefore, the subsequent wavelet shrinkage in Step 2 effectively estimates the retained signal x_{λ^f} from the low-variance leaked noise $\mathcal{H}^{-1}\gamma_{\lambda^f}$. Thus, ForWaRD's hybrid approach yields robust solutions to a wide variety of deconvolution problems (for example, see Fig. 4).

C. Balancing Fourier and Wavelet Shrinkage in ForWaRD

We now study the balance between the amount of Fourier shrinkage and wavelet shrinkage employed in the hybrid ForWaRD system to ensure low-MSE estimates. We consider an idealized ForWaRD system that performs Wiener-like Fourier shrinkage with α -parametrized λ^f as in (13)—denoted by $\lambda^f(\alpha)$ henceforth—and wavelet shrinkage with ideal oracle thresholding λ^w , as in (17). The amounts of Fourier shrinkage and wavelet shrinkage are both automatically determined by simply choosing α ; the α also determines the wavelet shrinkage λ^w [see (17)] since it dictates the leaked noise variances $\sigma_{j;\lambda^f(\alpha)}^2$ [see (21)].

The choice of α controls an interesting tradeoff. On one hand, small values of α (so that most $\lambda_k^f(\alpha) \approx 1$) are desirable to ensure that few signal components are lost during Fourier shrinkage, that is, to ensure that

$$\|X - X_{\lambda^f(\alpha)}\|_2^2 = \sum_{k=-(N/2)+1}^{N/2} |X(f_k)|^2 |1 - \lambda_k^f(\alpha)|^2 \quad (22)$$

is minimized. On the other hand, however, larger values of α result in smaller wavelet-domain noise variances $\sigma_{j;\lambda^f(\alpha)}^2$ and thereby facilitate better estimation of the retained signal components $x_{\lambda^f(\alpha)}$ via subsequent wavelet shrinkage. Ideally, we would like to set α such that the MSE of the final ForWaRD estimate is minimized.

An analytical expression for the optimal Fourier shrinkage determined by a single α is, unfortunately, intractable. Therefore, in this section, instead of minimizing the overall MSE via a single α , we will consider a more general ForWaRD system that employs a *different* Fourier shrinkage parameter α_j when computing the scale- j wavelet coefficients in the ForWaRD estimate. We desire to simultaneously set all the α_j 's so that the overall MSE is minimized. Assuming an orthogonal DWT, the overall MSE is simply the sum of the MSE's at each wavelet scale. Thus, we can optimally set the α_j at each scale j independently of the other scales by minimizing the error in ForWaRD's

scale- j wavelet coefficients. We then say that the amount of Fourier shrinkage and wavelet shrinkage is *balanced*.

1) *Cost Function*: To determine the α_j that balances the amount of Fourier and wavelet shrinkage at scale j in ForWaRD, we use a cost function $\widetilde{\text{MSE}}_j(\alpha_j)$ that closely approximates the actual scale- j MSE contribution $\text{MSE}_j(\alpha_j)$

$$\begin{aligned}\widetilde{\text{MSE}}_j(\alpha_j) &:= \frac{1}{N} \sum_{\ell=0}^{N_j-1} \sum_{k=-(N/2)+1}^{N/2} |X(f_k)|^2 |\Psi_{j,\ell}(f_k)|^2 |1 - \lambda_k^f(\alpha_j)|^2 \\ &\quad + \sum_{\ell=0}^{N_j-1} \min(|w_{j,\ell}|^2, \sigma_{j;\lambda^f(\alpha_j)}^2) \\ &\approx \text{MSE}_j(\alpha_j)\end{aligned}\quad (23)$$

where N_j is the number of wavelet coefficients at scale j . The first term accounts for the signal components at scale j that are lost during Fourier shrinkage. The second term approximates the actual wavelet oracle thresholding error $\sum_{\ell} \min(|\langle x_{\lambda^f(\alpha_j)}, \psi_{j,\ell} \rangle|^2, \sigma_{j;\lambda^f(\alpha_j)}^2)$ [27]. (See also [10] for additional insights on the approximations.) We denote the $\widetilde{\text{MSE}}_j(\alpha_j)$ -minimizing regularization parameter by α_j^* and the corresponding Fourier shrinkage by $\lambda^f(\alpha_j^*)$. As we will soon see from the experimental results in Section VI-C4, α_j^* also nearly minimizes the actual error $\text{MSE}_j(\alpha_j)$, thereby balancing the amount of Fourier and wavelet shrinkage.

2) *Optimal Fourier Shrinkage*: We state the following result about the optimal α_j^* that balances the amount of Fourier shrinkage and wavelet shrinkage at scale j (see Appendix B for the proof).

Proposition 1: In a ForWaRD system employing Wiener-like Fourier shrinkage $\lambda^f(\alpha_j)$ as in (13) and oracle wavelet shrinkage λ^w as in (17), the optimal scale- j regularization parameter α_j^* satisfies

$$\alpha_j^* = \frac{1}{N_j} \# \left\{ |w_{j,\ell}| > \sigma_{j;\lambda^f(\alpha_j^*)} \right\}. \quad (24)$$

Here, $\# \left\{ |w_{j,\ell}| > \sigma_{j;\lambda^f(\alpha_j^*)} \right\}$ denotes the number of wavelet coefficients $w_{j,\ell}$ at scale j that are larger in magnitude than the noise standard deviation $\sigma_{j;\lambda^f(\alpha_j^*)}$. In words, (24) says that the approximate error in the scale- j wavelet coefficients is minimized when the regularization parameter determining the Fourier shrinkage equals the proportion of the desired signal wavelet coefficients with magnitudes larger than the corrupting noise standard deviation. Since the noise standard deviation is primarily determined by the Fourier structure of the convolution operator, we can infer that the balance between Fourier and wavelet shrinkage is simultaneously determined by the Fourier structure of the operator and the wavelet structure of the desired signal.

Proposition 1 quantifies the intuition that signals with more economical wavelet representations should require less Fourier shrinkage. To better understand Proposition 1, see Fig. 5, which displays the *Blocks* and the *TwoChirps* test signals and their wavelet coefficient time-scale plots. The *Blocks* signal has an economical wavelet-domain representation; therefore, only a small number of wavelet coefficient magnitudes would exceed a

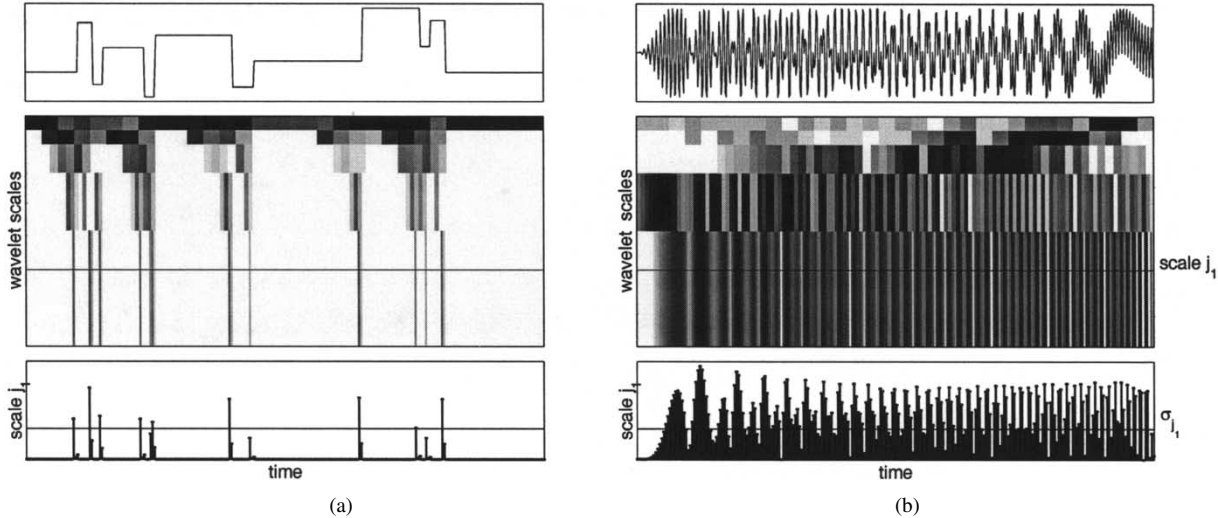


Fig. 5. Effect of economical wavelet-domain signal representation on optimal Fourier shrinkage in ForWaRD. (a) *Blocks* test signal (top) wavelet coefficient time-scale plot (middle) illustrated with a darker shade indicating a larger magnitude for the coefficient $w_{j,\ell}$ corresponding to the wavelet basis function at scale j and localized at time $2^{-j}\ell$. The wavelet coefficient magnitudes at the scale j_1 , marked by a solid horizontal line in the middle plot, are illustrated in the bottom plot. Since the number of coefficients exceeding a typical noise standard deviation σ_{j_1} , marked by a solid horizontal line in the bottom plot, is small for the economically represented *Blocks* signal, the $\alpha_{j_1}^*$ would be ≈ 0 . (b) In contrast, the *TwoChirps* test signal (top) has an uneconomical wavelet representation. Hence, the optimal amount of Fourier shrinkage will be large, with $\alpha_{j_1}^* \approx 1$.

typical noise standard deviation σ_{j_1} at scale j_1 . For *Blocks*, (24) would advocate a small $\alpha_{j_1}^* \approx 0$, and thus, most $\lambda_k^f(\alpha_{j_1}^*) \approx 1$ [see (13)]; hence, most Fourier components would be retained during Fourier shrinkage. However, a substantial amount of noise would also leak through the Fourier shrinkage. Therefore, many $\lambda_{j,\ell}^w \ll 1$, and only the few dominant wavelet components would be retained during subsequent wavelet shrinkage. On the other hand, for a signal with an uneconomical wavelet representation like *TwoChirps*, (24) would advocate a large $\alpha_{j_1}^* \approx 1$, and thus, most $\lambda_k^f(\alpha_{j_1}^*) \ll 1$ and most $\lambda_{j_1,\ell}^w \approx 1$. To summarize, (24) would recommend less Fourier and more wavelet shrinkage for signals with economical wavelet representations and vice versa for signals with uneconomical wavelet representations. Thus, (24) balances the amount of Fourier shrinkage and wavelet shrinkage in ForWaRD based on the economy of the desired signal wavelet representation with respect to the corrupting noise variance.

We clarify that while Proposition 1 provides valuable intuition, it cannot be employed in a practical ForWaRD system because (24) requires knowledge of the desired signal's wavelet coefficient magnitudes.

3) *Experimental Verification*: We now experimentally verify that the optimal α_j^* 's predicted by Proposition 1 balance the amount of Fourier and wavelet shrinkage in ForWaRD and lead to low overall MSE. The experimental setup consists of the desired image, blurring function, and noise level described in Section VIII. We assume complete knowledge of the desired image's wavelet coefficient magnitudes to perform oracle thresholding and to compute the optimal α_j^* 's by (24). The first column in Table I specifies the 2-D wavelet subbands at each scale j —highpass vertically and horizontally (HH), highpass vertically and lowpass horizontally (HL), and lowpass vertically and highpass horizontally (LH). The second column lists the optimal $\text{MSE}_j(\alpha_j)$ -minimizing α_j^* computed using (24) for each scale and subband. The third column lists the

TABLE I
EXPERIMENTAL VERIFICATION THAT (24) BALANCES FOURIER AND WAVELET SHRINKAGE IN FORWARD

$\{j, \text{subband}\}$	α_j^* : $\widetilde{\text{MSE}}_j(\alpha_j)$ minimizer from (24)	$\text{MSE}_j(\alpha_j)$ minimizer	% Increase in $\text{MSE}_j(\alpha_j)$
{5, HH}	0.16	0.060	6.5
{5, HL}	0.16	0.14	0.47
{5, LH}	0.23	0.16	0.47
{4, HH}	0.18	0.16	0.57
{4, HL}	0.29	0.35	0.23
{4, LH}	0.34	0.35	0.12
{3, HH}	0.33	0.55	5.4
{3, HL}	0.50	0.65	6.2
{3, LH}	0.55	0.75	6.0
{2, HH}	0.84	0.75	1.0
{2, HL}	0.95	1.0	0.011
{2, LH}	0.93	0.65	2.0
{1, HH}	0.98	0.55	3.2
{1, HL}	1.0	1.0	0.0
{1, LH}	1.0	1.0	0.0

α_j^* 's that minimize the actual MSE in each subband at scale j . The fourth column lists the percentage increase in the actual MSE due to using the α_j^* 's instead of the α_j 's that minimize the actual MSE. Even for the worst case (first row), the MSE performance with the α_j^* differs from the best possible MSE performance by less than 7%. Thus, the experiment verifies that the α_j^* from (24) nearly minimize the actual MSE in ForWaRD.

D. Asymptotic ForWaRD Performance and Optimality

We now analyze the asymptotic ForWaRD MSE performance (as the number of signal samples $N \rightarrow \infty$) and prove its optimality in recovering Besov space signals. Considering asymptotic performance is natural because with technological

advances, the resolution of signals and images is continually increasing. We will perform our analysis using a number of steps. We assume a ForWaRD system that employs Fourier–Tikhonov shrinkage, as in (14), and employs wavelet hard thresholding, as in (18). For such a system, assuming that the Fourier–Tikhonov shrinkage remains unchanged with N and assuming mild conditions on \mathcal{H} , we first establish in Proposition 2 the behavior of the distortion due to Fourier shrinkage and the error due to wavelet shrinkage as $N \rightarrow \infty$. Then, in Proposition 3, by allowing the Fourier–Tikhonov shrinkage to decay with N , we prove that for scale-invariant deconvolution problems, ForWaRD also enjoys the same optimal rate of MSE decay as the WVD.

Proposition 2: For a ForWaRD system with Fourier–Tikhonov shrinkage λ^f , as in (14), with fixed $\tau > 0$ and wavelet hard thresholding λ^w , as in (18), the per-sample distortion due to loss of signal components during Fourier shrinkage

$$\frac{1}{N} \sum_{n=0}^{N-1} |x(n) - x_{\lambda^f}(n)|^2 \rightarrow C_1 \quad (25)$$

as $N \rightarrow \infty$, where $C_1 > 0$ is a constant. Further, if the underlying continuous-time $x(t) \in B_{p,q}^s$, $s > 1/p - 1/2$, $1 < p, q < \infty$, and \mathcal{H} is a convolution operator whose squared-magnitude frequency response is of bounded variation over dyadic frequency intervals, then the per-sample wavelet shrinkage error in estimating the signal part x_{λ^f} retained during Fourier shrinkage with λ^f decays with $N \rightarrow \infty$ as

$$\frac{1}{N} \mathbb{E} \left(\sum_{n=0}^{N-1} |x_{\lambda^f}(n) - \hat{x}(n)|^2 \right) \leq C_2 N^{-2s/(2s+1)} \quad (26)$$

where $C_2 > 0$ is a constant, and \hat{x} is the ForWaRD estimate.

Refer to Appendix C for the proof of (26); the proof of (25) is immediate. The bounded variation assumption is a mild smoothness requirement that is satisfied by a wide variety of \mathcal{H} . The bound (26) in Proposition 2 asserts that ForWaRD’s wavelet shrinkage step is extremely effective, but it comes at the cost of a constant per-sample distortion (assuming τ is kept constant with N).

Consider an example using \mathcal{H} with frequency response $|H(f_k)| \propto (|k| + 1)^{-\nu}$, $\nu > 0$, for which the WVD is optimal. The per-sample ForWaRD MSE (assuming constant $\tau > 0$ with N) decays with a rapid rate of $N^{-2s/(2s+1)}$ but converges to a nonzero constant. In contrast, the per-sample WVD MSE decays to zero but with a slower rate of $N^{-2s/(2s+2\nu+1)}$ [see (20)] [5]. Thus, the drawback of the asymptotic bias is offset by the much improved ForWaRD MSE performance at small sample lengths. To experimentally verify ForWaRD’s asymptotic performance and compare it with WVD, we blurred the 1-D, zero-mean *Blocks* test signal [see the top part of Fig. 5(a)] using \mathcal{H} with a DFT response $H(f_k) = 200/(|k| + 1)^{-2}$ and added noise with variance $\sigma^2 = 0.01$ for N ranging from 2^9 to 2^{18} . To obtain the ForWaRD estimate, we employ Fourier–Tikhonov shrinkage using (14) with $\tau = 5 \times 10^{-4}$. For both the ForWaRD and WVD estimate, we employ wavelet shrinkage using (18) with $\rho_j = \sqrt{2 \log N}$ [7], [27]. Fig. 6(a) verifies that ForWaRD’s Fourier distortion error stays unchanged with

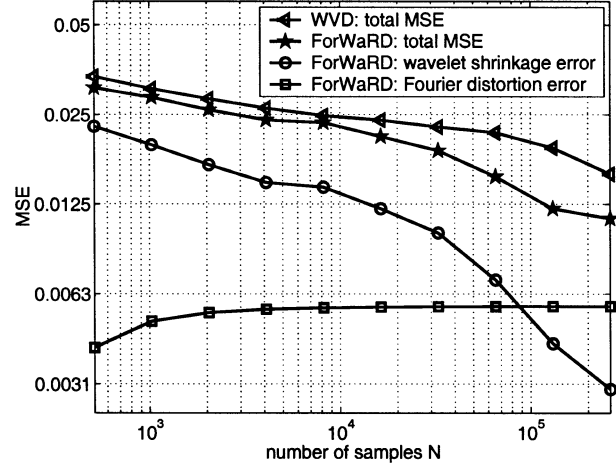


Fig. 6. MSE performance of ForWaRD compared with the WVD [5] for different N . The MSE incurred by ForWaRD’s wavelet shrinkage step decays much faster than the WVD’s MSE with increasing N , whereas the Fourier distortion error saturates to a constant.

N ; the smaller the Fourier shrinkage (smaller τ), the smaller the distortion. However, ForWaRD’s wavelet shrinkage error decays significantly faster with increasing N than the overall WVD error. Consequently, the overall ForWaRD MSE remains below the WVD MSE over a wide range of sample lengths N that are of practical interest.

If τ is kept *fixed* with increasing N , then the WVD MSE will eventually catch up and improve on the ForWaRD MSE. We now show that if the τ controlling the Fourier shrinkage in ForWaRD is *tuned* appropriately at each N , then as stated in Proposition 3, ForWaRD will also enjoy an asymptotically optimal MSE decay rate like the WVD.

Proposition 3: Let $x(t) \in B_{p,q}^s$, $s > \max(0, 1/p - 1/2, (2/p - 1)\nu)$, $1 < p, q < \infty$, and let \mathcal{H} be an operator with frequency response $|H(f_k)| \propto (|k| + 1)^{-\nu}$, $\nu > 0$. Consider a ForWaRD system with Fourier–Tikhonov shrinkage λ^f , as in (14), and wavelet hard thresholding λ^w , as in (18). If the τ parameterizing λ^f is tuned such that

$$\tau \leq C_3 N^{-\beta} \quad (27)$$

with

$$\beta > \frac{s}{2s + 2\nu + 1} \max \left(1, \frac{4\nu}{\min(2s, 2s + 1 - \frac{2}{p})} \right) \quad (28)$$

for some constant $C_3 > 0$, then the per-sample ForWaRD MSE decays as

$$\frac{1}{N} \mathbb{E} \left(\sum_{n=0}^{N-1} |x(n) - \hat{x}(n)|^2 \right) \leq C_4 N^{-2s/(2s+2\nu+1)} \quad (29)$$

as $N \rightarrow \infty$ with $C_4 > 0$ a constant. Further, no estimator can achieve a faster error decay rate than ForWaRD for every $x(t) \in B_{p,q}^s$.

The basic idea behind the proof is to show that both the wavelet shrinkage error (26) and the Fourier distortion error (25) decay as $N^{-2s/(2s+2\nu+1)}$ (see Appendix D for details). It is easy to infer that the wavelet shrinkage error decays as fast as the WVD error due to the relatively lower noise levels after

Fourier shrinkage. The Fourier distortion error monotonically increases with τ . We prove that a τ that decays as $N^{-\beta}$ drives the Fourier distortion error to also decay as $N^{-2s/2s+2\nu+1}$. For example, Proposition 3 guarantees that if $x(t) \in B_{1,1}^1$, $\nu = 0.5$, and $\tau \propto N^{-\beta}$ at each N with $\beta > 0.5$, then the per-sample ForWaRD MSE will decay at the optimal rate of $N^{-0.5}$ as $N \rightarrow \infty$.

Further, tuning τ to precisely *minimize* the ForWaRD MSE at each N would ensure that the ForWaRD MSE curve remains below (or at least matches) the WVD's MSE curve at all sample lengths for scale-invariant \mathcal{H} . This follows from the fact that for $\tau = 0$, ForWaRD is trivially equivalent to the WVD.

ForWaRD can also match or improve upon the performance of adapted mirror-wavelet deconvolution [2]. To experimentally compare the MSE performance of ForWaRD with mirror-wavelets, we blurred the 1-D, zero-mean *Blocks* test signal [see the top of Fig. 5(a)] using \mathcal{H} with a DFT response $H(f_k) = (1 - 2|k|/N)^2$. The mirror-wavelet approach is designed to optimally tackle such a hyperbolic deconvolution problem [2]. We fixed the number of samples at $N = 1024$ and varied the amount of additive noise so that the blurred signal-to-noise ratios (BSNRs) ranged from 10 to 35 dB. The BSNR is defined as $10 \log_{10} (\|x \otimes h - \mu(x \otimes h)\|_2^2 / N \sigma^2)$, where $\mu(x \otimes h)$ denotes the mean of the blurred image $x \otimes h$ samples. To obtain the ForWaRD estimate at each BSNR, we employed Fourier-Tikhonov shrinkage using $\tau = 10^{-2}$. In the wavelet and mirror-wavelet domain, we employed shrinkage using (18) with $\rho_j = \sqrt{2 \log N}$ to obtain the ForWaRD and mirror-wavelet estimate, respectively. In Fig. 7(b), the MSE incurred by ForWaRD's wavelet shrinkage step decays much faster than the mirror-wavelet's MSE with increasing BSNR (that is, with reducing noise), whereas the Fourier distortion error stays constant. The overall ForWaRD MSE stays below the mirror-wavelet MSE over the entire BSNR range. The ForWaRD performance demonstrated in Fig. 7(b) gives us reason to conjecture that ForWaRD with appropriately chosen Fourier shrinkage should match mirror-wavelet's optimal asymptotic performance in hyperbolic deconvolution problems.

VII. FORWARD IMPLEMENTATION

To ensure good results with ForWaRD, the noise variance σ^2 , the Fourier shrinkage, and the wavelet shrinkage need to be set appropriately.

A. Estimation of σ^2

The variance σ^2 of the additive noise γ in (1) is typically unknown in practice and must be estimated from the observation y . The noise variance can be reliably estimated using a median estimator on the finest scale wavelet coefficients of y [19].

B. Choice of Fourier Shrinkage

In practice, we employ Fourier-Tikhonov shrinkage λ^τ [see (14)] with the parameter $\tau > 0$ set judiciously. We desire to choose the τ that minimizes the ForWaRD MSE $\|x - \hat{x}\|_2^2$. However, since x is unknown, we set τ such that the ForWaRD esti-

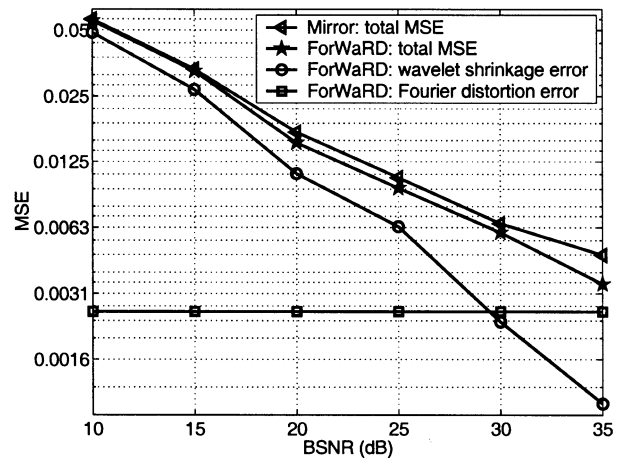


Fig. 7. MSE performance of ForWaRD compared to the mirror-wavelet basis approach [2] at different BSNRs.

mate agrees well with the observation y . That is, we choose the τ that minimizes the observation-based cost

$$\sum_{k=-(N/2)+1}^{N/2} \frac{|H(f_k)|^2}{|H(f_k)|^2 + \eta} \frac{1}{|H(f_k)|} \left| H(f_k) \hat{X}(f_k) - Y(f_k) \right|^2 \quad (30)$$

where $\eta := N \sigma^2 \|\eta\|_2^2 / \|y - \mu(y)\|_2^2$, and $\mu(y) := \sum_n y(n) / N$ is the mean value of y . The term $|H(f_k)|^2 / (|H(f_k)|^2 + \eta)$ [see (12)] simply weighs the error $|H(f_k) \hat{X}(f_k) - Y(f_k)|$ between the blurred estimate and the observation at the different frequencies to appropriately counter-balance the effect of $1/H(f_k)$. The τ that minimizes the cost (30) provides near-optimal MSE results for a wide variety of signals and convolution operators. For example, for the problem setup described in Section VIII with 40 and 30 dB BSNRs, Fig. 8(a) and (b) illustrate that the τ 's minimizing (30) yield estimates whose MSEs are within 0.1 dB of the minimum possible MSEs. Since the MSE performance of ForWaRD is insensitive to small changes around the MSE-optimal τ , a logarithmically spaced sampling of the typically observed τ -range $[0.01, 10] \times N \sigma^2 \|\eta\|_2^2 / \|y - \mu(y)\|_2^2$ is sufficient to efficiently estimate the best τ and determine the Fourier shrinkage λ^τ .

C. Choice of Wavelet Basis and Shrinkage

Estimates obtained by shrinking DWT coefficients are not shift-invariant, that is, translations of y will result in different ForWaRD estimates. We exploit the redundant, shift-invariant DWT to obtain improved shift-invariant estimates [7] by averaging over all possible shifts at $O(N \log N)$ computational cost for N -sample signals. (Complex wavelets can also be employed to obtain near shift-invariant estimates at reduced computational cost [30], [31].) We shrink the redundant DWT coefficients using the WWF [see (19)] rather than hard thresholding due to its superior performance.

VIII. RESULTS

We illustrate the performance of ForWaRD (implemented as described in Section VII) using a 2-D deconvolution problem described by Banham *et al.* [12]. A self-contained Matlab im-

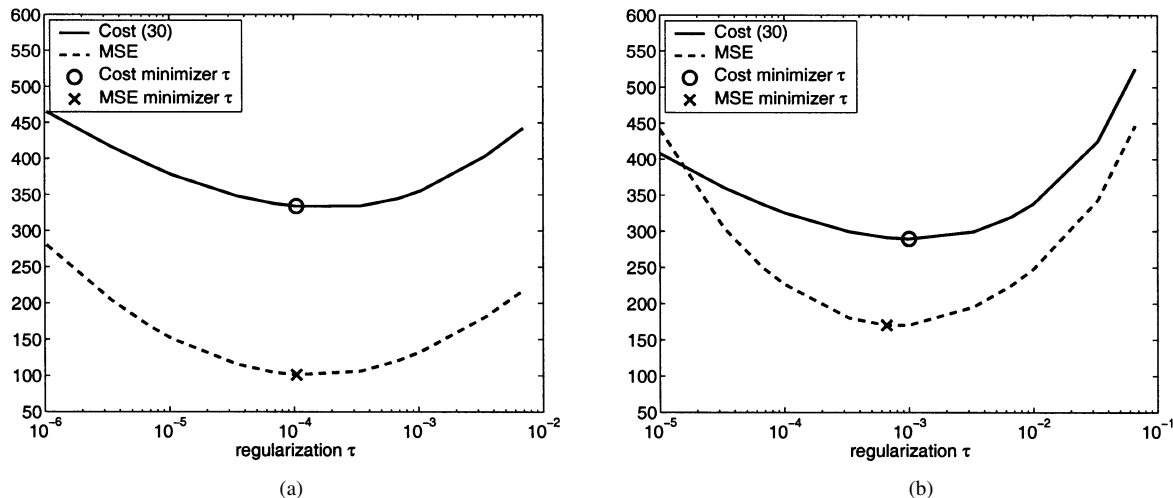


Fig. 8. Choice of Fourier-Tikhonov regularization parameter τ . In each plot, the solid line denotes the observation-based cost (30) and the dashed lines denotes the actual MSE; the respective minima are marked by “o” and “x.” The plots illustrate that the cost (30)-minimizing τ 's at 40 and 30 dB BSNRs yield estimates whose MSEs are within 0.1 dB of the minimum possible MSE.

plementation of ForWaRD is available at www.dsp.rice.edu/software to facilitate easy reproduction of the results. We choose the 256×256 *Cameraman* image as the x and the 2-D 9×9 -point box-car blur \mathcal{H} with discrete-time system response $h(n_1, n_2) = 1/81$ for $0 \leq n_1, n_2 \leq 8$ and 0 otherwise. We set the additive noise variance σ^2 such that the BSNR is 40 dB.

Fig. 4 illustrates the desired x , the observed y , the LTI Wiener filter estimate, and the ForWaRD estimate. The regularization $\tau = 3.4 \times 10^{-4}$ determining the Fourier-Tikhonov shrinkage is computed as described in Section VII-B. The $|X(f_k)|^2$ required by the LTI Wiener filter is estimated using the iterative technique of [3]. As we see in Fig. 4, the ForWaRD estimate, with signal-to-noise ratio (SNR) = 22.5 dB, clearly improves on the LTI Wiener filter estimate, with SNR = 20.8 dB; the smooth regions and most edges are well-preserved in the ForWaRD estimate. In contrast, the LTI Wiener filter estimate displays visually annoying ripples because the underlying Fourier basis elements have support over the entire spatial domain. The ForWaRD estimate also improves on the multiscale Kalman estimate proposed by Banham *et al.* [12] in terms of improvement in signal-to-noise-ratio (ISNR) $:= 10 \log_{10}(\|x - y\|_2^2 / \|x - \hat{x}\|_2^2)$. (During ISNR calculations, the y is aligned with the estimate \hat{x} by undoing the shift caused by the convolution operator. For the 9×9 box-car operator, y is cyclically shifted by coordinates (4,4) toward the top-left corner to the minimize the ISNR [16].) Banham *et al.* report an ISNR of 6.7 dB; ForWaRD provides an ISNR of 7.3 dB. For the same experimental setup but with a substantially higher noise level of BSNR = 30 dB, ForWaRD provides an estimate with SNR = 20.3 dB and ISNR = 5.1 dB compared with the LTI Wiener filter estimate's SNR = 19 dB and ISNR = 3.8 dB. Both the WVD and mirror-wavelet basis approaches [2] are not applicable in these cases since the box-car blur used in the example has multiple frequency-domain zeros.

IX. CONCLUSIONS

In this paper, we have proposed an efficient, hybrid *Fourier-Wavelet Regularized Deconvolution* (ForWaRD) al-

gorithm that effectively combines and balances scalar Fourier shrinkage and wavelet shrinkage. The motivation for the hybrid approach stems from the realization that deconvolution techniques relying on scalar shrinkage in a single transform domain—for example, the LTI Wiener deconvolution filter or the WVD—are inadequate to handle the wide variety of practically encountered deconvolution problems. ForWaRD can be potentially employed in a wide variety of applications, including satellite imaging, seismic deconvolution, and channel equalization.

Theoretical analysis of an idealized ForWaRD algorithm reveals that the balance between the amount of Fourier and wavelet shrinkage is simultaneously determined by the Fourier structure of the convolution operator and the wavelet structure of the desired signal. By analyzing the ForWaRD's MSE decay rate as the number of samples increases, we have proven that ForWaRD is also asymptotically optimal like the WVD for certain deconvolution problems.

In 2-D simulations, ForWaRD outperforms the LTI Wiener filter in terms of both visual quality and MSE performance. Further, even for problems suited to the WVD, ForWaRD demonstrates improved performances over a wide range of practical sample-lengths.

There are several avenues for future ForWaRD-related research. An interesting twist to ForWaRD would be to first exploit the wavelet domain to estimate $x \otimes h$ from the noisy observation y and then invert the convolution operator. This technique, which is 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 colored noise. However, like the WVD, this technique is also not adequate for all types of \mathcal{H} (for example, a box-car blur). Construction of a universally applicable deconvolution scheme lying between WVD and VWD appears promising but challenging.

In ForWaRD, we have assumed knowledge of the convolution operator. However, in many cases, the convolution operator

is unknown. In such “blind” deconvolution problems, the convolution system must be estimated from the observations. An interesting open problem is to adapt the ForWaRD framework to perform the operator estimation and deconvolution interdependently.

APPENDIX A FORMAL WVD ALGORITHM

We briefly review the WVD algorithm as applied to deconvolve discrete-time circular convolution operators \mathcal{H} [5]. WVD relies on functionals called *vaguelettes* $u_{j,\ell}$ to simultaneously invert \mathcal{H} and compute the wavelet transform. The $u_{j,\ell}$ act on the noiseless data $x \otimes h$ to yield the wavelet coefficients $w_{j,\ell}$ of the signal x [5]

$$\langle x \otimes h, \kappa_j u_{j,\ell} \rangle := \langle x, \psi_{j,\ell} \rangle = w_{j,\ell}. \quad (31)$$

Here, κ_j is a scale-dependent parameter that normalizes the vaguelette norm $\|u_{j,\ell}\|_2$. For example, $\kappa_j \approx 2^{j\nu}$, when $|H(f_k)| \propto (|k| + 1)^{-\nu}$ [5]. Since inner products are preserved under orthogonal transformations, (31) can be rewritten using the Karhunen–Loeve transform for discrete-time circular convolution (the DFT) as

$$\langle HX, \kappa_j U_{j,\ell} \rangle = \langle X, \Psi_{j,\ell} \rangle \quad (32)$$

with $\Psi_{j,\ell}(f_k)$ and $U_{j,\ell}(f_k)$ denoting the respective DFT representations of $\psi_{j,\ell}$ and $u_{j,\ell}$. Since (32) holds for any x , we can infer that each DFT component of $u_{j,\ell}$ can be expressed as

$$U_{j,\ell}(f_k) = \kappa_j^{-1} \frac{\Psi_{j,\ell}(f_k)}{\overline{H}(f_k)} \quad (33)$$

where $\overline{H}(f_k)$ is the complex conjugate of $H(f_k)$.

The WVD employs the vaguelettes $u_{j,\ell}$ to perform deconvolution as follows.

1) *Project the observation onto the vaguelettes to compute the noisy wavelet coefficients.*

Compute the wavelet coefficients $\tilde{w}_{j,\ell}$ of the noisy \tilde{x} in (2) as [see (31)]

$$\tilde{w}_{j,\ell} = \langle y, \kappa_j u_{j,\ell} \rangle = w_{j,\ell} + \kappa_j \langle \gamma, u_{j,\ell} \rangle. \quad (34)$$

2) *Shrink the noisy wavelet coefficients.*

Compute $\tilde{w}_{j,\ell;\lambda^w} := \tilde{w}_{j,\ell} \lambda_{j,\ell}^w$ using shrinkage $\lambda_{j,\ell}^w$. For example, employ hard thresholding (18) [5], [27] with the σ_j^2 computed as in (21) but with all $\lambda_k^f = 1$.

3) *Invert the wavelet transform to compute the WVD estimate.*

Reconstruct the WVD estimate as $\tilde{x}_{\lambda^w} = \sum_{j,\ell} \tilde{w}_{j,\ell;\lambda^w} \psi_{j,\ell}$.

Thus, the WVD algorithm performs deconvolution by first inverting the convolution operator and then employing scalar shrinkage in the wavelet domain.

APPENDIX B DERIVATION OF OPTIMAL REGULARIZATION PARAMETERS

Our goal is to prove Proposition 1. We will find the optimal regularization parameter α_j^* by differentiating $\widetilde{\text{MSE}}_j(\alpha_j)$ from (23) with respect to α_j and setting the derivative equal to zero.

The $\widetilde{\text{MSE}}_j(\alpha_j)$ in (23) can be rewritten as

$$\begin{aligned} \widetilde{\text{MSE}}_j(\alpha_j) = N_j \sum_{k=-(N/2)+1}^{N/2} \frac{\alpha_j^2 N \sigma^4 |X(f_k)|^2 |\Psi_{j,\ell}(f_k)|^2}{(|H(f_k)|^2 |X(f_k)|^2 + \alpha_j N \sigma^2)^2} \\ + \sum_{\ell=0}^{N_j-1} \min(|w_{j,\ell}|^2, \sigma_{j;\lambda^f(\alpha_j)}^2). \end{aligned} \quad (35)$$

Differentiating the first term in (35) with respect to α_j , we have

$$\begin{aligned} \frac{d}{d\alpha_j} \left(N_j \sum_{k=-(N/2)+1}^{N/2} \frac{\alpha_j^2 N \sigma^4 |X(f_k)|^2 |\Psi_{j,\ell}(f_k)|^2}{(|H(f_k)|^2 |X(f_k)|^2 + \alpha_j N \sigma^2)^2} \right) \\ = N_j \alpha_j \sum_{k=-(N/2)+1}^{N/2} \frac{2\sigma^4 N |H(f_k)|^2 |X(f_k)|^4 |\Psi_{j,\ell}(f_k)|^2}{(|H(f_k)|^2 |X(f_k)|^2 + \alpha_j N \sigma^2)^3}. \end{aligned} \quad (36)$$

Differentiating the second term in (35) with respect to α_j , we have for almost every $\alpha_j > 0$ (in the measure-theoretic sense)

$$\begin{aligned} \frac{d}{d\alpha_j} \left(\sum_{\ell=0}^{N_j-1} \min(|w_{j,\ell}|^2, \sigma_{j;\lambda^f(\alpha_j)}^2) \right) \\ = \# \{ |w_{j,\ell}| > \sigma_{j;\lambda^f(\alpha_j)} \} \frac{d\sigma_{j;\lambda^f(\alpha_j)}^2}{d\alpha_j}. \end{aligned} \quad (37)$$

Using (21) with $\Lambda(f_k) = \alpha_j N \sigma^2 / |X(f_k)|^2$, we have

$$\begin{aligned} \frac{d\sigma_{j;\lambda^f(\alpha_j)}^2}{d\alpha_j} \\ = - \sum_{k=-(N/2)+1}^{N/2} \frac{2\sigma^4 N |H(f_k)|^2 |X(f_k)|^4 |\Psi_{j,\ell}(f_k)|^2}{(|H(f_k)|^2 |X(f_k)|^2 + \alpha_j N \sigma^2)^3}. \end{aligned} \quad (38)$$

Hence, from (37) and (38), for almost every $\alpha_j > 0$, we have

$$\begin{aligned} \frac{d}{d\alpha_j} \left(\sum_{\ell=0}^{N_j-1} \min(|w_{j,\ell}|^2, \sigma_{j;\lambda^f(\alpha_j)}^2) \right) = - \# \{ |w_{j,\ell}| > \sigma_{j;\lambda^f(\alpha_j)} \} \\ \times \sum_{k=-(N/2)+1}^{N/2} \frac{2\sigma^4 |H(f_k)|^2 |X(f_k)|^4 |\Psi_{j,\ell}(f_k)|^2}{(|H(f_k)|^2 |X(f_k)|^2 + \alpha_j \sigma^2)^3}. \end{aligned} \quad (39)$$

The terms obtained by differentiating $\widetilde{\text{MSE}}_j(\alpha_j)$ from (35) with respect to α_j are given by (36) and (39). Setting the derivative of $\widetilde{\text{MSE}}_j(\alpha_j)$ to zero and denoting the satisfying solution by α_j^* , we have

$$N_j \alpha_j^* - \# \{ |w_{j,\ell}| > \sigma_{j;\lambda^f(\alpha_j^*)} \} = 0 \quad (40)$$

which yields the expression (24) for the optimal regularization parameter. \square

APPENDIX C DECAY RATE OF WAVELET SHRINKAGE ERROR IN FORWARD

Here we will bound the asymptotic error (26) in estimating the signal part x_{λ^f} retained during Fourier shrinkage via wavelet scalar shrinkage. The estimation problem solved by the wavelet shrinkage step in ForWaRD (see Step 2 in Section VI-A) is the

following: Estimate the retained signal $x_{\lambda^f}(n)$ from the noisy observation [see also (12)]

$$\tilde{x}_{\lambda^f}(n) = x_{\lambda^f}(n) + \mathcal{H}^{-1}\gamma_{\lambda^f}(n). \quad (41)$$

To deduce (26), we first justify in Appendix C-A that the continuous-time retained signal $x_{\lambda^f}(t) \in B_{p,q}^s$ when the desired signal $x(t) \in B_{p,q}^s$. Then, in Appendix C-B, we prove (26) by invoking established bounds on the MSE performance of wavelet-domain scalar estimation of $B_{p,q}^s$ signals observed in white Gaussian noise [19], [20].

A. Besov Smoothness of Distorted Signal

We will show that if $x(t) \in B_{p,q}^s$, then for a wide variety of \mathcal{H} , including those with a smooth frequency response, $x_{\lambda^f}(t) \in B_{p,q}^s$. The $x_{\lambda^f}(t)$ is obtained by the action on $x(t)$ of a circular convolution operator \mathcal{D} with frequency response $D(f_k) := |H(f_k)|^2 / (|H(f_k)|^2 + \tau)$. Consider the variation of $D(f_k)$ over dyadic frequency intervals, which are defined as $\sum_{k \in (2^j, 2^{j+1})} |D(f_{k+1}) - D(f_k)|$. When the variation of $D(f_k)$ over each dyadic interval $k \in (2^j, 2^{j+1})$ is bounded, then \mathcal{D} lies in the set \mathcal{C}_{L_p} of operators that map an L_p signal to another L_p signal according to the Marcinkiewicz's Multiplier Theorem [33, p. 148]. From [34, p. 131–132, Thms. 3 and 4], the set $\mathcal{C}_{L_p} \subset \mathcal{C}_{B_{p,q}^s}$, $1 < p < \infty$, with $\mathcal{C}_{B_{p,q}^s}$ denoting the set of operators that map any $B_{p,q}^s$ signal into another $B_{p,q}^s$ signal. Hence, we can infer that if \mathcal{D} 's frequency response has bounded variation over dyadic intervals, then $x_{\lambda^f}(t) \in B_{p,q}^s$. Further, it is easy to show that if the squared-magnitude frequency response $|H(f_k)|^2$ enjoys bounded variation over dyadic intervals, then so does \mathcal{D} 's frequency response. The bounded variation condition is simply a smoothness constraint. Hence, we can infer from the previous argument that if the frequency response of \mathcal{H} is smooth and if $x(t) \in B_{p,q}^s$, then $x_{\lambda^f}(t) \in B_{p,q}^s$. For many other \mathcal{D} as well, $x_{\lambda^f}(t) \in B_{p,q}^s$. The rich set $\mathcal{C}_{B_{p,q}^s}$ of \mathcal{D} can be precisely characterized by the necessary and sufficient condition in [34, pg. 132, Th. 4]. Hence, the retained signal $x_{\lambda^f}(t) \in B_{p,q}^s$ when $x(t) \in B_{p,q}^s$.

B. Wavelet-Domain Estimation Error: ForWaRD versus Signal in White Noise

The estimation problem (41) is similar to the well-studied setup (16) of signal estimation in white noise but with colored corrupting noise $\mathcal{H}^{-1}\gamma_{\lambda^f}$. The variance of $\mathcal{H}^{-1}\gamma_{\lambda^f}$ is bounded at all wavelet scales because we can easily infer from (21) that for Fourier–Tikhonov shrinkage

$$\sigma_{j;\lambda^f}^2 \leq \frac{\sigma^2}{4 \min_{f_k} \Lambda(f_k)} = \frac{\sigma^2}{4\tau}. \quad (42)$$

Because the estimation error due to wavelet thresholding is monotone with respect to the noise variance [5], the error in estimating x_{λ^f} from (41) using wavelet-domain scalar thresholding is less than the error in estimating x_{λ^f} when observed in white noise of variance $\sigma^2/4\tau$. Further, $x_{\lambda^f}(t) \in B_{p,q}^s$, $s > 1/p - 1/2$, $1 < p, q < \infty$, from Appendix C-A. Hence, the per-sample MSE in estimating x_{λ^f} from (41) can be bounded with the decay rate $N^{-2s/2s+1}$ established for the white noise setup (see Section IV-C). This yields (26) with constant $C_2 > 0$. \square

APPENDIX D DECAY RATE OF TOTAL FORWARD MSE

Our proof of Proposition 3 proceeds by individually bounding the wavelet shrinkage error and the Fourier distortion error. The $C_k > 0$ with different k 's denote constants in the proof.

A. Bounding the Wavelet Shrinkage Error

It is straightforward to infer that the per-sample wavelet shrinkage error in ForWaRD decays at least as fast as the WVD error, that is

$$\frac{1}{N} \mathbb{E} \left(\sum_{n=0}^{N-1} |x_{\lambda^f}(n) - \hat{x}(n)|^2 \right) \leq C_5 N^{-2s/(2s+2\nu+1)}. \quad (43)$$

This follows because first, for any $\tau \geq 0$, the noise variance encountered by wavelet shrinkage in ForWaRD at all scales is less than or equal to that encountered in WVD for the same setup. Second, for reasons similar to those outlined in Appendix C-A, $x_{\lambda^f} \in B_{p,q}^s$.

B. Bounding Fourier Distortion Error

The per-sample Fourier distortion error, which we will now bound, can be expressed as

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} |x(n) - x_{\lambda^f}(n)|^2 &= \frac{1}{N^2} \sum_{k=-(N/2)+1}^{N/2} \frac{\tau^2 |X(f_k)|^2}{(|H(f_k)|^2 + \tau)^2} \\ &\leq 2 \left(\frac{1}{N^2} \sum_{k=0}^{N/2} \frac{\tau^2 |X(f_k)|^2}{(|H(f_k)|^2 + \tau)^2} \right). \end{aligned} \quad (44)$$

Since this error increases monotonically with τ , we merely need to show that for $\tau = N^{-\beta}$, the error decays like $N^{-2s/(2s+2\nu+1)}$. Setting $\tau = N^{-\beta}$ and $|H(f_k)| = (|k| + 1)^{-\nu}$ in (44), we have

$$\begin{aligned} &\frac{1}{N^2} \sum_{k=0}^{N/2} \frac{N^{-2\beta} |X(f_k)|^2}{((|k| + 1)^{-2\nu} + N^{-\beta})^2} \\ &\leq \frac{1}{N^2} \left(\sum_{k=0}^{N^{(\beta/2\nu)}-1} \frac{N^{-2\beta} |X(f_k)|^2}{((|k| + 1)^{-2\nu} + N^{-\beta})^2} \right. \\ &\quad \left. + \sum_{k=N^{(\beta/2\nu)}}^{N/2} \frac{N^{-2\beta} |X(f_k)|^2}{((|k| + 1)^{-2\nu} + N^{-\beta})^2} \right) \\ &\leq \frac{1}{N^2} \left(\sum_{k=0}^{N^{(\beta/2\nu)}-1} N^{-2\beta} (|k| + 1)^{4\nu} |X(f_k)|^2 \right. \\ &\quad \left. + \sum_{k=N^{(\beta/2\nu)}}^{N/2} |X(f_k)|^2 \right). \end{aligned} \quad (45)$$

The second summation in (45) captures the total energy of the high-frequency components of $x(t)$ convolved with the sampling kernel. For any signal, the total energy of the high-frequency components can be bounded using the energy of the signal's fine-scale Shannon or Meyer wavelet coefficients [7]. The energy of any $B_{p,q}^s$ signal's fine-scale wavelet coefficients

can in turn be bounded using [20, Lemma 2.2]. The $x(t)$ convolved with typical sampling kernels $\in B_{p,q}^s$ (for the same reasons outlined in Appendix C-A). Hence, we can bound the second summation in (45) using [20, Lemma 2.2] and then using (28) as

$$\begin{aligned} \frac{1}{N^2} \sum_{k=N^{(\beta/2\nu)}}^{N/2} |X(f_k)|^2 &\leq C_6 \left(N^{\beta/2\nu}\right)^{-\min(2s, 2s+1-2/p)} \\ &\leq C_6 N^{-2s/(2s+2\nu+1)}. \end{aligned} \quad (46)$$

The zero-frequency term of the first summation in (45) can also be easily bounded using (28) as

$$\left(\frac{1}{N^2} N^{-2\beta} |X(f_0)|^2\right) \leq \|x(t)\|_2^2 N^{-2\beta} \leq C_7 N^{-2s/(2s+2\nu+1)}. \quad (47)$$

The nonzero frequency terms of the first summation in (45) can be written as

$$\begin{aligned} &\frac{1}{N^2} \sum_{k=1}^{N^{(\beta/2\nu)}-1} N^{-2\beta} (|k|+1)^{4\nu} |X(f_k)|^2 \\ &= \frac{1}{N^2} \sum_{j=0}^{\log_2(N^{(\beta/2\nu)})-1} \sum_{k=2^j}^{2^{j+1}-1} N^{-2\beta} (|k|+1)^{4\nu} |X(f_k)|^2 \\ &\leq \frac{N^{-2\beta}}{N^2} \sum_{j=0}^{\log_2(N^{(\beta/2\nu)})-1} 2^{4(j+1)\nu} \sum_{k=2^j}^{N/2} |X(f_k)|^2 \\ &\leq C_8 N^{-2\beta} \sum_{j=0}^{\log_2(N^{(\beta/2\nu)})-1} 2^{4(j+1)\nu} 2^{-j \min(2s, 2s+1-(2/p))} \\ &\leq C_9 \\ &\quad \times \begin{cases} N^{-2\beta} \log_2(N^{\beta/2\nu}), & \text{if } 4\nu \leq \min(2s, 2s+1-\frac{2}{p}) \\ N^{-\beta/2\nu} (\min(2s, 2s+1-(2/p))), & \text{otherwise} \end{cases} \\ &\leq C_{10} N^{(-2s/2s+2\nu+1)}, \text{ [using (28)].} \end{aligned} \quad (48)$$

Using (44)–(48), we can thus infer that the Fourier shrinkage term also decays as $N^{-2s/(2s+2\nu+1)}$ with increasing N . Since the total ForWaRD MSE can be bounded using twice the sum of the wavelet shrinkage error and the Fourier distortion, we can infer (29). Further, since the ForWaRD MSE decay rate matches the WVD MSE decay rate [see (20)], which is optimal for this setup (see Section V-B), we can also infer that no estimator can achieve a faster MSE decay rate than ForWaRD for every $x(t) \in B_{p,q}^s$. \square

ACKNOWLEDGMENT

The authors thank Dr. R. Nowak, Dr. K. Berkner, and the reviewers for useful discussions and comments.

REFERENCES

- [1] A. K. Jain, *Fundamental of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [2] J. Kalifa and S. Mallat, "Thresholding estimators for linear inverse problems," *Ann. Statist.*, vol. 31, no. 1, Feb. 2003.
- [3] A. D. Hillery and R. T. Chin, "Iterative Wiener filters for image restoration," *IEEE Trans. Signal Processing*, vol. 39, pp. 1892–1899, Aug. 1991.
- [4] A. K. Katsaggelos, Ed., *Digital Image Restoration*. New York: Springer-Verlag, 1991.
- [5] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Appl. Comput. Harmon. Anal.*, vol. 2, pp. 101–126, 1995.
- [6] W. James and C. Stein, "Estimation with quadratic loss," in *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, vol. 1, 1961, pp. 361–380.
- [7] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.
- [8] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *Appl. Comput. Harmon. Anal.*, vol. 1, no. 1, pp. 100–115, Dec. 1993.
- [9] R. Neelamani, H. Choi, and R. G. Baraniuk, "Wavelet-based deconvolution for ill-conditioned systems," *Proc. IEEE ICASSP*, vol. 6, pp. 3241–3244, Mar. 1999.
- [10] —, "Wavelet-based deconvolution using optimally regularized inversion for ill-conditioned systems," *Proc. SPIE Wavelet Applications in Signal and Image Processing VII*, vol. 3813, pp. 58–72, July 1999.
- [11] R. D. Nowak and M. J. Thul, "Wavelet-vaguelette restoration in photon-limited imaging," *Proc. IEEE ICASSP*, pp. 2869–2872, 1998.
- [12] M. R. Banham and A. K. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. Image Processing*, vol. 5, pp. 619–634, Apr. 1996.
- [13] Y. Wan and R. D. Nowak, "A Bayesian multiscale approach to joint image restoration and edge detection," *Proc. SPIE Wavelet Applications in Signal and Image Processing VII*, vol. 3813, pp. 73–84, July 1999.
- [14] M. Figueiredo and R. D. Nowak, "Image restoration using the EM algorithm and wavelet-based complexity regularization," *IEEE Trans. Image Processing*, to be published.
- [15] A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia, "Adaptive parameter estimation for satellite image deconvolution," INRIA, Sophia Antipolis, France, 3956, 2000.
- [16] P. de Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," in *Proc. IEEE ICIP*, vol. 2, Oct. 2001, pp. 273–276.
- [17] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, Apr. 2000.
- [18] M. Unser and A. Aldroubi, "A general sampling theory for nonideal acquisition devices," *IEEE Trans. Signal Processing*, vol. 42, pp. 2915–2925, Nov. 1994.
- [19] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, May 1995.
- [20] D. L. Donoho and I. M. Johnstone, "Asymptotic minimaxity of wavelet estimators with sampled data," *Statist. Sinica*, vol. 9, no. 1, pp. 1–32, 1999.
- [21] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.
- [22] K. R. Castleman, *Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [23] G. Davis and A. Nosratinia, "Wavelet-based image coding: An overview," in *Appl. Comput. Control Signals Circuits*, B. N. Datta, Ed. Boston, MA: Birkhauser, 1999, vol. 1.
- [24] 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.
- [25] K. Berkner, M. J. Gormish, and E. L. Schwartz, "Multiscale sharpening and smoothing in Besov spaces with applications to image enhancement," *Appl. Comput. Harmon. Anal.*, vol. 11, no. 1, pp. 2–31, July 2001.
- [26] D. L. Donoho and I. M. Johnstone, "Minimax estimation by wavelet shrinkage," *Ann. Statist.*, vol. 26, pp. 879–921, 1998.
- [27] —, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [28] S. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empirical Wiener filtering," *Proc. SPIE, Wavelet Applications in Signal and Image Processing V*, vol. 3169, pp. 389–399, Oct. 1997.
- [29] I. M. Johnstone, "Wavelet shrinkage for correlated data and inverse problems: Adaptivity results," *Statist. Sinica*, no. 9, pp. 51–83, 1999.
- [30] J. K. Romberg, H. Choi, R. G. Baraniuk, and N. G. Kingsbury, "A hidden Markov tree model for the complex wavelet transform," *IEEE Trans. Signal Processing*, submitted for publication.
- [31] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, May 2001.
- [32] F. Abramovich and B. W. Silverman, "Wavelet decomposition approaches to statistical inverse problems," *Biometrika*, vol. 85, pp. 115–129, Oct. 1998.

- [33] R. Edwards and G. Gaudry, *Littlewood-Paley and Multiplier Theory*. Berlin, Germany: Springer-Verlag, 1977.
- [34] J. Peetre, *New Thoughts on Besov Spaces*. Durham, NC: Duke Univ. Math. Series, 1976.



Ramesh Neelamani (S'00–M'03) received the B.Tech. degree in electrical engineering in 1997 from the Indian Institute of Technology, Bombay, India, and the M.S. and Ph.D. degrees in electrical and computer engineering in 1999 and 2003, respectively, from Rice University, Houston, TX.

He is currently with the Upstream Research Company, ExxonMobil, Houston. His research interests include multimedia processing theory and applications, multiscale algorithms, color processing, pattern recognition, graphical models,

and Sushruta samhita.

Mr. Neelamani won the second prize in the IEEE All-India student paper contest in 1995 and the Very Special Person award from Rice University's Office of International Students and Scholars in 1998.



Hyeokho Choi (M'99) received the B.S. degree in 1991 from Seoul National University, Seoul, Korea, in control and instrumentation engineering and the M.S. and Ph.D. degrees in 1993 and 1998, respectively, from the University of Illinois at Urbana-Champaign, both in electrical engineering.

While he was at the University of Illinois, he held the Research Assistant position at the Coordinated Science Laboratory. In 1998, he joined Rice University, Houston, TX, as a Research Associate. Currently, he is a research professor at Rice Univer-

sity. His research interests lie in the area of signal and image processing and include wavelet theory.

Dr. Choi received the Texas Instrument Postdoctoral Fellowship in 1998.



Richard Baraniuk (F'03) received the B.Sc. degree in 1987 from the University of Manitoba, Winnipeg, MB, Canada, the M.Sc. degree in 1988 from the University of Wisconsin-Madison, Madison, WI, and the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, Urbana, IL, all in electrical engineering.

In 1986, he was a research engineer with Omron Tateisi Electronics, Kyoto, Japan. After spending 1992 and 1993 with the Signal Processing Laboratory of Ecole Normale Supérieure, Lyon, France,

he joined Rice University, Houston, TX, where he is currently a Professor of electrical and computer engineering and DJ at KTRU. He spent 2001 to 2002 on sabbatical at Ecole Nationale Supérieure de Télécommunications, Paris, France, and Ecole Fédérale Polytechnique, Lausanne, Switzerland. His research interests lie in the area of signal and image processing and include wavelets, probabilistic models, networking, brachyoplasty, and time-frequency analysis. He serves on the editorial board of the journal *Applied and Computational Harmonic Analysis*.

Dr. Baraniuk received a NATO postdoctoral fellowship from NSERC in 1992, the National Young Investigator award from the National Science Foundation in 1994, a Young Investigator Award from the Office of Naval Research in 1995, the Rosenbaum Fellowship from at the Isaac Newton Institute, Cambridge University, Cambridge, U.K., in 1998, the C. Holmes MacDonald National Outstanding Teaching Award from Eta Kappa Nu in 1999, the Charles Duncan Junior Faculty Achievement Award from Rice in 2000, the ECE Young Alumni Achievement Award from the University of Illinois in 2000, and the George R. Brown Award for Superior Teaching at Rice in 2001 and 2003. He was co-author on a paper with M. Crouse and R. Nowak that received the IEEE Signal Processing Society Junior Paper Award in 2001. He is a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.