RICE UNIVERSITY

Wavelet-Based Deconvolution for Ill-Conditioned Systems

by

Ramesh Neelamani

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Master of Science

Approved, Thesis Committee:

Richard G. Baraniuk, Chair
Associate Professor of Electrical and
Computer Engineering

Don H. Johnson
Professor of Electrical and Computer
Engineering and Statistics

C. Sidney Burrus
Dean of Engineering,
Maxfield and Oshman Professor of
Electrical and Computer Engineering

Houston, Texas

May, 1999

Wavelet-Based Deconvolution for Ill-Conditioned Systems

Ramesh Neelamani

Abstract

This thesis proposes a new approach to wavelet-based image 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 metric, we strike an optimal balance between Fourier-domain regularization that is matched to the system and wavelet-domain regularization that is matched to the input signal. The resultant algorithm is fast, $O(N \log_2^2 N)$ where N denotes the number of samples, and is well-suited to signals and images with spatially-localized phenomena such as edges. In addition to enjoying asymptotically optimal rates of error decay for some systems, the algorithm also achieves excellent performance at fixed data lengths. In simulations with real data, the algorithm outperforms the conventional LTI Wiener filter and other wavelet-based deconvolution algorithms in terms of both visual quality and MSE performance.

To my parents Neelamani Ramasubramaniam and Sitalaxmi Neelamani,

and my sister Anusha Neelamani.

Acknowledgments

A big, big, big thanks to Hyeokho Choi for all those hours he spent sharing his insight and expertise with me; his help was invaluable. It was great fun to work on this project thanks to Justin, Vinay, Matt, Roger, Rohit, Rocco, and Jim Lewis, who never failed to provide lively company till early hours in the morning. I also extend my gratitude to Kathrin Berkner and Rob Nowak for the discussions I have shared with them.

I am truly fortunate to have RichB as my advisor. He has gone beyond being just a research advisor to being a mentor; I am truly at a loss of words to express my gratitude towards him.

I would also like to thank my other committee members, Dr. C. S. Burrus and Dr. D. H. Johnson. Their comments have gone a long way towards improving this thesis.

Good Indian food had a very important role to play in this project. It provided me with the energy I needed to plough along. So I have to extend my gratitude to Chis, Pots, Felix, and Whiney for providing me with Indian food that was seldom inedible!

I would like to acknowledge the support given by the National Science Foundation, grant MIP–9457438, DARPA/AFOSR, grant F49620–97–1–0513, Texas Instruments Leadership University Program, and the Rice Consortium on Computational Seismic Interpretation.

Contents

Illustrations

Chapter 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]. The importance of deconvolution can be grasped with a real-life example from image restoration. Satellite images obtained in practice are often blurred due to limitations such as aperture effects of the camera, camera motion and atmospheric turbulence. Deconvolution becomes necessary to get a crisp, deblurred image that is needed for viewing or further processing.

In its simplest form, the 1-d deconvolution problem runs as follows. The desired signal x is input to a known linear time-invariant (LTI) system $\mathcal{H}$ having impulse response h. Independent identically distributed (i.i.d.) samples of Gaussian noise γ with variance σ^2 corrupt the output samples of the system $\mathcal{H}$. The observations at discrete points t_n are given by

$$y(t_n) := (x \circledast h)(t_n) + \gamma(n) \qquad n = 1, \ldots, N. \tag{1.1}$$

For simplicity, we assume circular convolution, denoted by $\circledast$ in (1.1). Given y, we seek to estimate x. In the Discrete Fourier Transform (DFT) domain, we equivalently have

$$Y(f_n) = H(f_n)\, X(f_n) + \Gamma(f_n) \qquad n = 1, \ldots, N, \tag{1.2}$$

where Y, H, X and Γ are the length-N DFTs of the observation y, LTI system impulse response h, input x and noise γ respectively. The $f_n := \frac{2\pi}{N}(n-1)$ denote

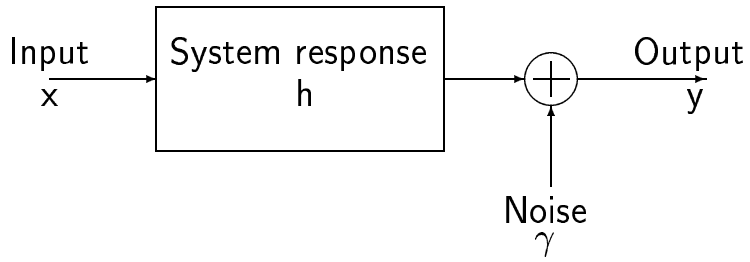


Figure 1.1 : *Convolution model setup.*

the normalized frequencies in the DFT domain. The problem formulation trivially extends to multi-dimensional data.

If the system frequency response $H(f_n)$ has no zeros, then an unbiased estimate of x can be obtained by inverting $\mathcal{H}$ as

$$\begin{aligned} \widehat{X}(f_n) &:= H^{-1}(f_n)Y(f_n) \\ &= X(f_n) + H^{-1}(f_n)\,\Gamma(f_n). \end{aligned} \tag{1.3}$$

However, if the system is ill-conditioned, i.e., $H(f_n)$ is small at any f_n, then enormous noise amplification results during inversion, resulting in an extremely noisy, useless estimate.

The problem of 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 distortion in the estimate [3]; regularization becomes essential in situations involving ill-conditioned systems.

Since the exact input is generally never recoverable, different measures can be used to evaluate the quality of an estimate; one commonly used measure of performance is the mean squared error (MSE) between the actual input x and the estimate. The LTI Wiener deconvolution filter is a commonly used example of regularized deconvolution.

It provides the MSE-optimal regularized LTI solution to the deconvolution problem. Formally, the estimation procedure used by the LTI Wiener deconvolution filter can be understood as follows: (1) Invert the convolution operator using $\frac{1}{H(f_n)}$ to obtain a noisy estimate $\widehat{X}(f_n)$. (2) Regularize by shrinking each frequency component of $\widehat{X}(f_n)$ proportionately according to the signal-to-noise-Ratio (SNR) at f_n [4]; shrink less/more when SNR is high/low.

The Fourier representation diagonalizes the convolution operator; hence the LTI Wiener deconvolution filter 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 LTI Wiener filter does not provide a good solution when the input signal contains spatially localized phenomena such as discontinuities, eg., edges (refer Figure 1.2(f)). The reason behind this drawback is the following: Though the Fourier domain exploits the structure of convolution operator $\mathcal{H}$ well, the domain is not well-suited to represent input signals with spatially localized features because the Fourier basis functions have support that extend over the entire spatial domain. So, LTI Wiener filter processing lacks spatial selectivity, consequently important spatial features such as edges get smeared.

Over the last decade, the wavelet transform have proven to be an invaluable tool for dealing with a wide class of signal including signals with spatially localized features. Wavelet representations use basis functions that are localized in time and frequency. Wavelets provide economical representations to a large class of signals such as signals belonging to Besov spaces [8]; wavelets capture most of the signal energy using a few large wavelet coefficients. The ability of wavelets to economically

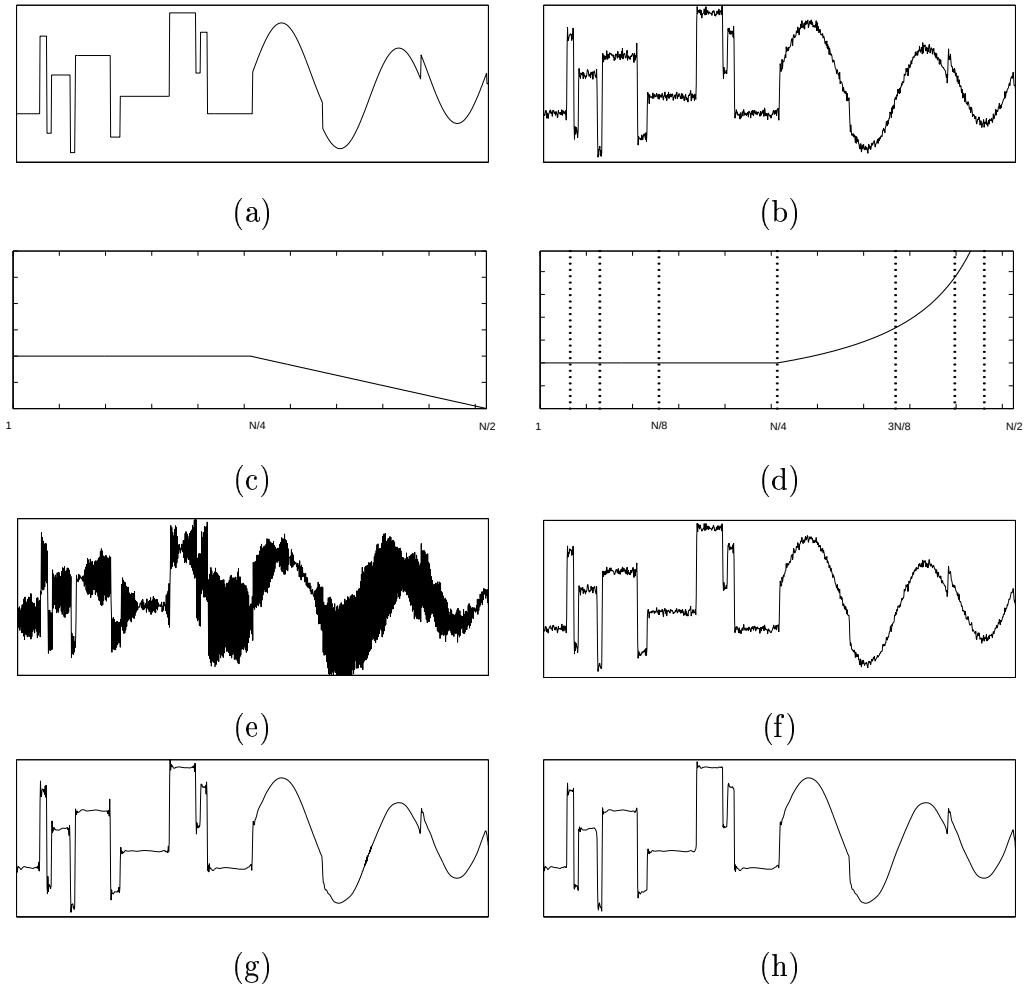


Figure 1.2 : *(a) Test signal x ($N = 1024$). It is a concatenation of Donoho's Blocks and Heavisine signals [5]. (b) Observed signal y (MSE=0.0060). It is obtained by blurring x with by H and adding noise. (c) System frequency response H. It contains a zero at index $\frac{N}{2}$ (index $\frac{N}{2}$ corresponds to the normalized DFT frequency $f = 0.5$). (d) System Inverse frequency response H^{-1}. It amplifies noise at high frequencies. The mirror wavelet frequency splits for estimate (g) are shown with dashed vertical lines. (e) Pure inverse estimate (MSE=1.1056). The pure inverse $\frac{1}{H}$ amplifies all the high frequency noise components to give an extremely noisy estimate. (f) Wiener filter estimate (MSE=0.0048). The Wiener filter is ineffective since the signal contains spatially localized phenomena such as edges; hence it provides an unsatisfactory estimate. (g) Mirror wavelet basis estimate [6, pp. 458–461] (MSE=0.0027). The frequency splits (shown using dotted lines in (d)) aim to isolate the frequency where H goes to zero. The estimate performs better than the WVD method of [7] and conventional Wiener filter. (h) WaRD estimate with regularization $\alpha^\star = 0.06$, (MSE=0.0017). The WaRD system tackles the noise amplification that stems from the action of H^{-1} (refer (d)) by employing partial Fourier-domain regularization followed by wavelet-domain estimation. It easily outperforms the mirror wavelet approach as well.*

represent a wide class of signals has been leveraged into powerful, spatially adaptive, signal-estimation algorithms that are based on simply shrinking the wavelet coefficients of the noisy signal. Further, such simple wavelet-based estimation schemes are MSE near-optimal for a wide class of signals; they outperform all linear estimation techniques [9].

The problem of deconvolution is equivalent to signal estimation in the presence of noise colored by $\mathcal{H}^{-1}$ (refer (1.3)). This motivates the following wavelet-based approach to the problem of deconvolution: (1) Invert the convolution operator using $\frac{1}{H(f_n)}$ to obtain a noisy estimate $\widehat{X}(f_n)$. (2) Regularize by shrinking each wavelet coefficient of the noisy estimate $\widehat{X}(f_n)$ using the noise variance at each wavelet scale.

Wavelet-based deconvolution techniques, hereby called WaD, aim to exploit the economical wavelet representation of signals to effectively identify and estimate the signal. In fact, for special classes of blurring operators such as the Radon transform, operator inversion followed by wavelet-domain shrinkage is near-optimal for a wide class of input signals [7, 10].

However, a WaD scheme cannot provide good estimates for all convolution systems $\mathcal{H}$. In fact, the wavelet-based scheme would provide a zero signal for an estimate in the presence of most systems that are non-invertible; eg., when the impulse response h is a box-car, a commonly used model in image blurring [2], the noisy estimate $\widehat{x}$ obtained after system inversion has infinite noise variance at all wavelet scales, rendering wavelet-domain estimation ineffective. In such a case, some amount of Fourier-domain regularization becomes imperative.

In this thesis, we propose an improved hybrid wavelet-based regularized deconvolution (WaRD) algorithm suitable for use with any ill-conditioned systems. The basic idea is simple: employ *both* Fourier-domain (Wiener-like) regularized inversion and wavelet-domain signal estimation. During this tandem processing, we exploit

Fourier-domain regularization to adapt to the convolution system and thereby control the noise but use it sparingly to keep the accompanying smearing distortions to the minimum required; bulk of the noise removal and signal estimation is achieved using wavelet shrinkage.

Using an MSE metric[1], we show that the optimal balance between local processing with wavelet basis and global processing with Fourier basis is heavily influenced by the distribution of input signal energy in the wavelet domain, and the convolution operator.

Interestingly, one extreme of the balance is to perform no Fourier-domain regularization; this is similar to the approaches of [6, 7]. In [7], Donoho introduced the Wavelet-Vaguellete Decomposition (WVD) technique to solve general linear inverse problems. For the special case of deconvolution, WVD is equivalent to WaD as described above. Donoho proved that, for a certain special class of operators that includes the Radon transform, WaD possesses asymptotically optimal rates of error decay as the number of observation samples increases. Since WaRD subsumes WaD, the proposed WaRD technique also possesses asymptotically optimal error decay rates for such special operators. In addition, WaRD also improves on the performance of WaD at any fixed samples by choosing the optimal amount of regularization. Further, unlike WaD, WaRD is applicable in the presence of any convolution operator.

Recently, Mallat proposed the use of a mirror-wavelet basis that adapts to the frequency response of convolution operator $\mathcal{H}$ to improve on the performance of WaD when the frequency response of $\mathcal{H}$ is as shown in Figure 1.2(b) [11]. Comparing Figures 1.2(g) and 1.2(h), we see that WaRD provided a better estimate than that

[1]Even though MSE metric does not capture visual appeal of images in general, we found that in case of wavelet-based deconvolution, optimizing over an MSE metric gave satisfactory results.

obtained by an adapted basis approach. Though an adapted basis provides better results than WaD (not shown), it is not effective for all types of ill-conditioned systems. eg., when $\mathcal{H}$ has a boxcar system response, adapting to the sinc frequency response of $\mathcal{H}$ using wavelets fails.

Nowak has employed an under-regularized filter in [12] during system inversion and subsequently used wavelet-domain signal estimation since the blurring operator in the problem was non-invertible. However he did not study the implications of using regularized inversion and the choice of the optimal amount of regularization.

After discussing regularization in more depth in Chapter 2, we briefly review wavelet transforms and its properties in Chapter 3. Previous wavelet-based deconvolution approaches are discussed in Chapter 4. We present our improved scheme in Chapter 5 and elaborate on its implementation in Chapter 6. Illustrative examples lie in Chapter 7. We conclude by summarizing our work and sketching out future directions in Chapter 8.

Chapter 2

Regularized Inversion of Ill-Conditioned Systems

Deconvolution involves estimating the input x from a blurred and noisy observation y (refer Eqn. (1.2)). Simply inverting the system fails to provide a satisfactory solution. The problems that arise can be understood by analyzing the Singular Value Decomposition (SVD) of the operator $\mathcal{H}$.

2.1 System Inversion using SVD

Let $h \circledast z$ denote the action of the convolution operator $\mathcal{H}$ on some signal z. This can be expressed in terms of the eigenvectors e_n and eigenvalues λ_n of the operator $\mathcal{H}$ as follows.

$$h \circledast z = \sum_n \lambda_n \langle z, e_n \rangle e_n, \tag{2.1}$$

where $\langle z, e_n \rangle$ denotes an inner product between z and e_n. e_n and λ_n are obtained by the SVD of the operator $\mathcal{H}$. The action of the inverse of $\mathcal{H}$ on any signal z, denoted by $h^{-1} \circledast z$, is given by

$$h^{-1} \circledast z = \sum_n \lambda_n^{-1} \langle z, e_n \rangle e_n. \tag{2.2}$$

Eqn. (2.2) reveals that the projection of z along eigenvectors that correspond to small eigenvalues of $\mathcal{H}$, get amplified during inversion. For any convolution operator, the sinusoidal DFT vectors form the eigenvectors and the DFT coefficients of the system impulse response h give the eigenvalues of the system $\mathcal{H}$. So, during inversion of the convolution operator (refer Eqn. (1.3)), the frequency components of noise corresponding to small Fourier coefficients of h are amplified during inversion.

An easy fix to the problem of noise amplification during inversion is to employ a windowed-SVD inversion [3].

$$\begin{aligned}\tilde{x} &= \tilde{h}^{-1} \circledast y \\ &= \sum_n r_n \lambda_n^{-1} \langle y, e_n \rangle e_n \\ &= \sum_n r_n \lambda_n^{-1} \langle h \circledast x + \gamma, e_n \rangle e_n \\ &= \sum_n r_n \langle x, e_n \rangle e_n + \sum_n r_n \lambda_n^{-1} \langle \gamma, e_n \rangle e_n, \end{aligned} \tag{2.3}$$

where r_n are the weights chosen to shrink the components of noise γ that get amplified during inversion of $\mathcal{H}$. This achieves noise reduction albeit at the cost of some bias in the estimate. The resultant operation is a common form of regularization that achieves quasi-inversion of $\mathcal{H}$.

2.2 LTI Wiener Deconvolution

For a convolution operator, the windowed-SVD inversion is equivalent to separately shrinking each Fourier component of the noisy observation obtained after pure inversion to attenuate the noise. The optimal weights that can be used to shrink each Fourier component during regularized inversion is a function of the SNR at each frequency. For WSS signals, the optimal weights $R(f_n)$ are given by

$$R(f_n) := \frac{|H(f_n)|^2 \, P_x(f_n)}{|H(f_n)|^2 \, P_x(f_n) \; + \; \sigma^2}, \tag{2.4}$$

where $P_x(f_n)$ is the power spectral density of the stochastic signal x and σ^2 is the variance of additive noise γ. The resulting LTI quasi-inverse operator is called the LTI Wiener deconvolution filter.

The action of the LTI Wiener deconvolution filter can be understood as being composed of the two steps[1](refer Figure 2.1).

[1]Both these steps are simultaneously employed in practice. So the LTI Wiener deconvolution

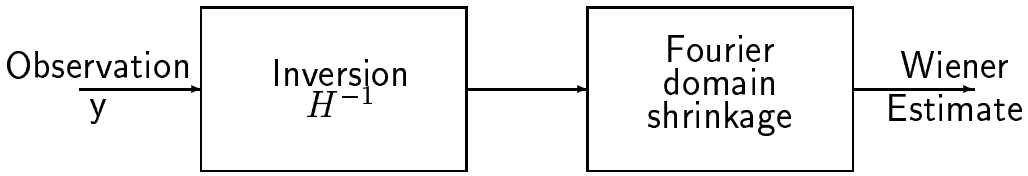


Figure 2.1 : *Block diagram of Wiener deconvolution filtering: Fourier domain shrinkage is used to estimate the signal in the presence of noise colored during inversion using* H^{-1}.

1. **Pure Inversion:** The noisy and blurred observation y is treated with H^{-1} to obtain a noisy, unbiased estimate $\hat{x}$ of the input signal as explained in Eqn. (1.3). This step necessarily amplifies the noise components at frequencies where $H(f_n)$ is small. Further processing is required to estimate the signal from this amplified, colored noise.

2. **Fourier-domain Signal Estimation:** Each frequency component of the noisy signal obtained after pure inversion is shrunk using weights $R(f_n)$. From expression (2.4), it is clear that less shrinkage is employed at components where the SNR is high. The LTI Wiener deconvolution estimate of the input signal is [4]

$$\widetilde{X}(f_n) := R(f_n)\left(\frac{1}{H(f_n)}\right)Y(f_n) \tag{2.5}$$

$$= R(f_n)X(f_n) + R(f_n)\left(\frac{1}{H(f_n)}\right)\Gamma(f_n). \tag{2.6}$$

For Gaussian WSS signals, the LTI Wiener deconvolution provides the globally MSE-optimal estimate of the input since the Fourier domain ideally represents *both*, the colored noise after inversion, and the signal of interest. The Fourier coefficients of the signal and the noise colored by H^{-1} are independent making individual, scalar

filter is applicable even when the system $\mathcal{H}$ is non-invertible.

estimation in the Fourier domain optimal. However, when the input signal or noise is not Gaussian, Wiener filtering no longer remains optimal.

When the input signal x is deterministic, the optimal weights are obtained by substituting $P_x(f_n) = |X(f_n)|^2$ in Eqn. (2.4)[2]. In the deterministic case, the conditions when the Wiener estimate (or in general the windowed-SVD) provides the optimal solution get more involved. We just mention that smoothness constraints need to be imposed on the signal and the convolution operator so the windowed-SVD is optimal. We refer the reader to [7, 13] for further details. We have assumed a deterministic signal for the rest of the thesis.

2.3 Drawbacks of the Wiener Deconvolution Filter

The LTI Wiener filter does not provide a good estimate when the input signal comprises of spatially localized phenomena such as edges. Though the Fourier domain is still the ideal domain to represent the noise colored by the inversion of $\mathcal{H}$, the domain is not well-suited to represent the signal of interest x. The Fourier basis functions that are exploited by the Wiener filter have support that extend over the entire domain. So, scalar operations on the Fourier coefficients lack any spatial localization thereby resulting in all spatial components of the input signal being processed uniformly.

For example, images can been modeled as piece-wise smooth functions, an important class of spatially varying signals [14]. For discontinuous signals such as piece-wise polynomial signals, the energy of the signal in the Fourier domain decays very slowly, $O(\frac{1}{N^2})$, as the number of samples N of the underlying continuous process increases. As a result, for a low-pass $\mathcal{H}$, the error per sample decays very slowly, $O(\frac{1}{N^{2\nu+2}})$ where ν characterizes the low-pass filter, as increasingly dense samples of the underlying

[2]Note that the quasi-inverse is now signal-dependent and hence non-linear in the strict sense.

continuous time observation are obtained (refer Appendix B).

The slow error decay results because the variance of noise colored by H^{-1} is large at high frequencies leading to excessive coefficient shrinkage (refer expression (2.4)) during Wiener estimation. As a result, the high frequency components required to reproduce sharp edges in the estimate are lost. The significant signal energy loss at high frequencies reflects in the form of an unsatisfactory estimate.

2.4 Alternative Solutions

Consider the following modified weights that can be used for shrinking the Fourier coefficients of the noisy signal obtained after inversion of $\mathcal{H}$.

$$R_\alpha(f_n) := \frac{|H(f_n)|^2\,|X(f_n)|^2}{|H(f_n)|^2\,|X(f_n)|^2 \;+\; \alpha\,\sigma^2}, \tag{2.7}$$

where α is the called the regularization parameter [15]; the Wiener deconvolution filter uses $\alpha = 1$. For values of $\alpha < 1$, $R_\alpha(f_n)$ shrink the Fourier coefficients of the noisy signal $\widehat{X}(f_n)$ obtained after pure inversion (refer Eqn. (1.3)) less than the weights used by the Wiener filter. As a result, they cause less distortion and consequently result in sharper edges in the estimate. However, small α values back-fire in the form of higher noise levels in the estimate. On the other hand, for values of $\alpha > 1$, $R_\alpha(f_n)$ efficiently suppresses noise but at the cost of excessive distortion and smearing of edges. Thus, due to the inherent bias-variance tradeoff, simply changing the amount of Fourier-domain shrinkage does not provide a satisfactory solution to the problem of preserving spatially varying phenomena.

Better deconvolution techniques need to take the spatial vagaries of the signal into account. One such technique is to employ the best linear estimator, the time-varying or matrix version of the Wiener inverse [16]. However, the time-varying Wiener filter is impractical because it requires the input signal cross-correlation matrix which in turn

needs precise knowledge of the spatially varying phenomena of the signal. Further, the processing is computationally intensive, $O(N^3)$ where N is the number of samples, because the estimator possesses no special structure.

Deconvolution is equivalent to the problem of estimating the signal in convolutionally colored noise. In Utopia, we would come up with a transform that captures the signal energy and colored noise energy in separate, distinguishable coefficients. However, this is impossible for most interesting problems. Instead, we must compromise and find a domain where most of the signal energy is concentrated in a few identifiable, relatively noise-free coefficients that we would retain. It is also important that the domain be able to adapt to the unknown spatial characteristics of the signal.

Chapter 3

Background on Wavelets and Signal Estimation

3.1 Wavelet Transform

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 [6, 17]. For special choices of ϕ and ψ, the functions

$$\psi_{j,k}(t) \;:=\; 2^{j/2}\,\psi\big(2^j t - k\big)\,, \tag{3.1}$$

$$\phi_{j,k}(t) \;:=\; 2^{j}\,\phi\big(2^j t - k\big)\,,\;\; j,k \in \mathbb{Z}, \tag{3.2}$$

form an orthonormal basis, and we have the representation [6, 17]

$$x(t) \;=\; \sum_k u_{j_0,k}\,\phi_{j_0,k}(t) \;+\; \sum_{j=j_0}^{J}\sum_k w_{j,k}\,\psi_{j,k}(t), \tag{3.3}$$

with $u_{j,k} := \int x(t)\,\phi^*_{j,k}(t)\,dt$ and $w_{j,k} := \int x(t)\,\psi^*_{j,k}(t)\,dt$. The parameter J controls the resolution of the wavelet representation.

For a discrete-time signal, the N wavelet coefficients $\{u_{j_0,k}, w_{j,k}\}$ of $x(t_n)$ can be easily computed using a filter bank consisting of low-pass filters, high-pass filters, and decimators [17]. Due to the special filter band structure the forward and inverse wavelet transform can be computed in $O(N)$ operations, where N is the length of the signal. We have used the periodic DWT, which uses circular convolutions in the filter bank, in this thesis. For brevity, we will collectively refer to the set of scaling and wavelet coefficients as $\{\theta_{j,k}\} := \{u_{j_0,k}, w_{j,k}\}$. Multidimensional DWTs are easily obtained by alternately wavelet-transforming along each dimension [6, 17].

3.1.1 Multiresolution and time-frequency localization of wavelets

Wavelet provide a multiresolution representation of a signal, i.e., the wavelet coefficients capture the signal features at different resolution levels. In the wavelet representation notation used, j indexes the *scale* or the resolution of analysis - larger j corresponds to higher resolution of analysis. $j = j_0$ indicates the coarse scale or lowest resolution of analysis and $j = J$ indicates the fine scale or highest resolution of analysis. In the wavelet representation, 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. The relation of the DWT coefficients to tilings in a time-frequency plane is graphically depicted in Figure 3.1(a). In contrast, Figure 3.1(b) depicts the time-frequency tiling of the Fourier transform.

3.1.2 Wavelets as Unconditional Bases

Wavelets enjoy the following important property: they provide an unconditional basis for a signal classes such as Besov classes, which determines membership to any space in the class based on the signal smoothness; signal with different smoothness belong to different parameterized spaces in the Besov class [8]. For example, important spaces such as Sobolev spaces and L_p spaces belong to the Besov class. In essence, the unconditional basis property of wavelets means that signals belonging to such classes are characterized only by the amplitudes of the wavelet coefficients [7, 8]. In contrast, the Fourier basis do not provide an unconditional basis for such a wide class of signals.

The implications of this abstract notion of unconditional basis are extremely ap-

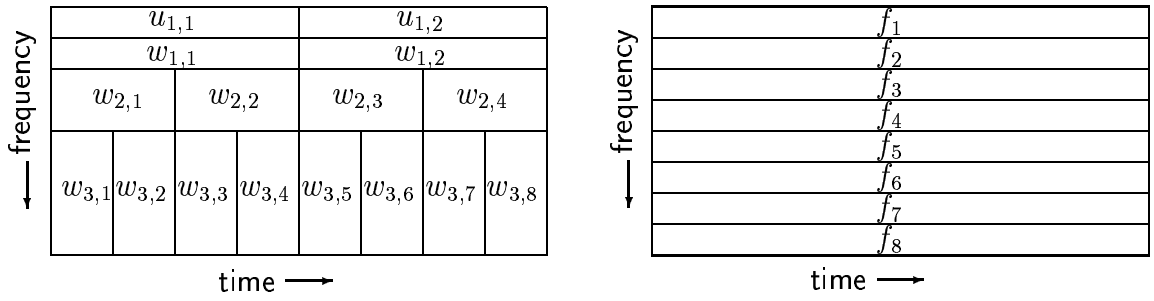


Figure 3.1 : *(a) Time-frequency tiling for DWT: The DWT basis function corresponding to the fine scale wavelet coefficients* $w_{3,1}, w_{3,2}, \ldots w_{3,8}$ *are well-localized in time but have large bandwidths. They capture the high-frequency characteristics of a signal at different times. The DWT basis functions corresponding to the coarser scales wavelet coefficients* $w_{2,1}, \ldots w_{2,4}$ *and* $w_{1,1}$*and*$w_{1,2}$*, capture the intermediate frequency characteristics of a signal. The scaling coefficients* $u_{1,1}$*and*$u_{1,2}$ *capture the lowest frequency characteristics of the signal at different times. (b) Time-frequency tiling for DFT: In contrast to the DWT basis functions, the DFT basis functions have support over the entire time domain but are perfectly localized in frequency. The DFT coefficients capture the overall signal energy at a particular frequency* f_n*.*

pealing. In [18], Donoho shows that unconditional basis are desirable because they typically express the signal very economically by capturing most of the signal energy in a few coefficients. The wavelet coefficients of signals in a Besov class decay exponentially with scale in the wavelet domain!

In addition to having direct implications on data compression, economical representations are also desirable for signal estimation using non-linear approximation [19, 20]. In fact, unconditional basis are optimal for signal estimation and data compression [18]. Further, the unconditional basis property also ensures that even simple scalar operations on each wavelet coefficient is sufficient to ensure near-optimal performances.

3.2 Signal Estimation by Wavelet Shrinkage

Recently, wavelets have provided effective solutions to the problem of signal estimation in the presence of noise [5, 9]. Consider the problem of signal estimation in the presence of white noise. Many real world signals have economical wavelet-domain representations where a few large wavelet coefficients capture most of the signal energy [6, 14]. However, since a wavelet transform is orthonormal and linear, the energy of white noise remains scattered over all 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.

Since wavelets provide good economical representations in the presence of spatial varying features of the signal, signal estimation in the wavelet domain preserves such features more effectively than conventional LTI techniques such as estimation methods based on windowed-SVD. In fact, wavelet-domain signal estimation techniques are near-optimal for a wide class of signals including piece-wise smooth signals and signals in the Besov class[5, 18].

Such 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 the signal class [10]

Chapter 4

Wavelet-based Deconvolution (WaD)

4.1 The WaD Algorithm

Since Eqn. (1.3) and Eqn. (1.1) are equivalent, deconvolution is equivalent to estimating the input signal in the presence of noise colored by H^{-1}. The near-optimality of wavelet-based techniques for a wide variety cases encountered in signal estimation combined with the appealing ability of wavelets to adapt to unknown spatial features of a signal has inspired the use of wavelets, particularly when the signal contains spatially localized features, to solve the problem of deconvolution in the following way (refer Figure 4.1).

1. **Pure Inversion:** Similar to Wiener deconvolution, the first step involves obtain a noisy, unbiased estimate $\widehat{x}$ of the input signal as explained in Eqn. (1.3).

2. **Wavelet-based Signal Estimation:** In contrast to Fourier-domain shrinkage employed by the Wiener deconvolution filter, WaD estimates the signal from

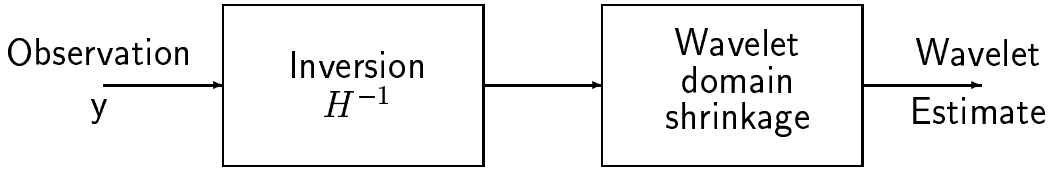


Figure 4.1 : *Block diagram of Wavelet-based deconvolution (WaD): Wavelet-domain shrinkage is used to estimate the signal in the presence of noise colored during inversion using* H^{-1}.

the noisy $\widehat{x}$ obtained after pure inversion by shrinking each wavelet coefficient of the noisy signal[1]. The variance of noise corrupting all wavelet coefficients in a particular scale is the same, but the variance varies with different scales [10]. So scale-dependent shrinkage is employed to estimate the signal wavelet coefficients. Inverse DWT is then used to obtain the wavelet-based estimate from the estimated signal wavelet coefficients.

4.2 Optimality of WaD

This basic idea was first exploited by Donoho who studied the applications of wavelets to linear inverse problems and showed that a wavelet-based deconvolution approach is near-optimal to recover a wide class of signals (eg., Besov classes) when the linear operator $\mathcal{H}$ satisfies

$$h(t) \circledast y(at) = a^{\beta}\, h(at) \circledast y(t), \tag{4.1}$$

for some exponent β. Such an operator $\mathcal{H}$ is said to be scale invariant or dilation homogeneous. The Radon transform is an important example of dilation homogeneous operators [7, 21]. To the best of our knowledge, the optimality of wavelet-based deconvolution for general dilation-inhomogeneous operators is still unknown. Using wavelet-based techniques, Nowak [12] has observed impressive results for some common LTI operators as well. Mallat [11] has also advocates a similar philosophy and claimed state-of-the-art performances in satellite image recovery.

We again consider the example of piece-wise polynomial functions to gain insight into the advantages of such an approach. Wavelets represent an length-N piece-wise polynomial signal very economically using just $O(\log_2 N)$ non-zero wavelet

[1]Note that different shrinkage techniques can be used on the noisy wavelet coefficients but the philosophy remains the same: Use wavelet-domain estimation instead of Fourier-domain estimation

coefficients[2]. As a result, assuming a invertible low-pass $\mathcal{H}$, the error per sample decays very rapidly, $O\left(\frac{\log_2 N}{N^{\frac{1}{2\nu+1}}}\right)$, as increasingly dense samples of the underlying continuous time observations are obtained (refer Appendix C). The error decay rate of $O\left(\frac{\log_2 N}{N^{\frac{1}{2\nu+1}}}\right)$ is significantly faster than $O\left(\frac{1}{N^{\frac{1}{2\nu+2}}}\right)$ that is achieved by Fourier-domain shrinkage.

WaD achieves faster error decay rates because the energy of the input signal gets concentrated in a few large coefficients in the wavelet domain. If the noise variance in these large signal coefficients is not excessive then identifying and retaining these signal coefficients, and setting the other coefficients to zero using shrinkage gives an excellent estimate. Because the retained signal coefficients capture most of the spatial features such as edges in the signal, the final estimate retains its sharp edges as well as noise-free smooth regions.

4.3 Drawbacks of WaD

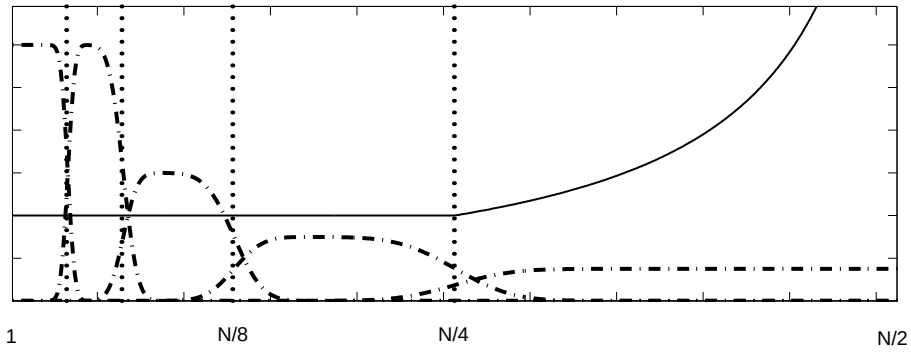
However, such a wavelet-based approach has its limitations. These can be understood by focusing on the variance of the wavelet coefficients of noise colored by H^{-1}. The noise variance σ_j^2 at scale j can be approximately by[3]

$$\sigma_j^2 \approx \frac{1}{2^{j-1}} \sum_{n=2^{j-1}}^{n=2^j-1} \frac{1}{|H(f_n)|^2} \sigma^2, \tag{4.2}$$

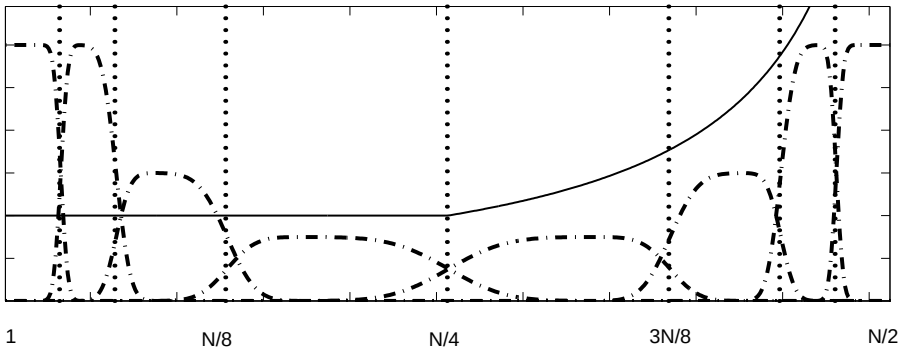
where $H(f_n)$ are the discrete Fourier coefficients of the blurring operator $\mathcal{H}$, and σ^2 is the variance of the AWGN γ (refer Figure 4.2).

[2]Consider a wavelet system with basis functions that possess number of zero-moments that is greater than or equal to the degree of all the polynomial pieces. If the support of any wavelet basis function lies within any polynomial piece, then the corresponding wavelet coefficient is zero.

[3]This approximation is valid only when the zeros in the frequency response H are of sufficiently low order.



(a)



(b)

Figure 4.2 : *(a) The solid line is the frequency response of the inverse of a typical lowpass system. The dot-and-dashed lines show the frequency bands corresponding to wavelet basis functions at different scales. The wavelet basis functions are nearly band-limited and almost constant within dyadic frequency bands (e.g.* $\frac{N}{4}$ *to* $\frac{N}{2}$*). These dyadic frequency bands are demarcated in the figure using dotted vertical lines. If blurring system frequency response does not have a high-order zero at frequency* $\frac{N}{2}$*, the variance of noise colored by* H^{-1} *(solid line) at different scales is primarily determined by the frequency response of the inverse within the corresponding dyadic frequency band. (b) The dot-and-dashed lines show the frequency responses of Mallat's mirror wavelet basis. The mirror wavelet basis functions adopt a frequency split that aims to isolate the singularities in the inverse system frequency response and thereby reduce the noise variance in most of the mirror wavelet basis subbands. However, if the singularity is not located at a dyadic frequency point (*$\frac{N}{2}, \frac{N}{4}, \frac{N}{8}, \ldots$*) then the non-zero frequency overlap leads to infinite noise variance at all scales in any adapted wavelet basis system. The same problem also crops up when the blurring system response has a high order zero. So an adapted wavelet basis system is not effective for all convolution operators.*

From the expression (4.2), it is clear that even if $H(f_n)$ is small at any isolated frequency f_n, the variance of the colored noise in the entire corresponding wavelet scale explodes. Consequently, the task of extracting the signal coefficient from this scale using any form of scalar shrinking in the wavelet domain is rendered ineffective. As shown in Figure 4.2, wavelet basis functions are not exactly band-limited; the DFT coefficients of the operator outside the band 2^{j-1} to $2^j - 1$ also contribute to the noise variance but to a lesser extent. However, if the the operator contains singularities that are placed at non-dyadic frequencies, then the noise variance is infinite at all the wavelet scales. In such a case, WaD provides a zero estimate.

4.4 Best-Basis Solution Improves Performance

Mallat advocates adapting the wavelet basis to the frequency response of the inverse of $\mathcal{H}$ to improve on the performance of WaD [6, 11]. For a convolution system with frequency response shown in Figure 1.2(c), the adaptation is achieved by using a mirror-wavelet basis that possess a time-frequency tiling structure different from that possessed by conventional wavelets (refer Figure 4.2). The tiling structure used by the mirror wavelet basis functions aims to isolate the frequency where the convolution operator $H(f)$ approaches zero because the variance of the noise at any scale is primarily influenced by the singularities of $H^{-1}(f)$ that lie in the frequency band corresponding to that wavelet scale. An estimate obtained by using such an approach is shown in Figure 1.2(f).

However, it is not always possible to come up with 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, any adapted wavelet basis scheme cannot provide a satisfactory solution when the blurring operator has a boxcar system impulse

response; the frequency response of the $\mathcal{H}$ is the sinc function. In such a case, the variance of noise colored by H^{-1} is high in all the wavelet scales, rendering signal estimation in adapted basis domain ineffective.

Recollect that during inversion using H^{-1}, only the noise components corresponding to the small Fourier coefficients of the convolution operator get amplified. So it is natural to exploit the structure of the operator using the Fourier domain to restrict excessive noise amplification in addition to using the frugal wavelet representation of the signal. This motivates our hybrid approach to deconvolution.

Chapter 5

Wavelet-based Regularized Deconvolution (WaRD)

In WaD, estimating the signal coefficients in the wavelet domain is reliable only if *the colored noise does not corrupt the signal wavelet coefficients heavily.* The noise variance in the wavelet domain can be significantly reduced by attenuating the amplified Fourier components of noise. This attenuation can be easily achieved by using a small amount of Fourier-domain regularization during inversion. So, we propose Wavelet-based Regularized Deconvolution (WaRD) that simultaneously exploit the spatial adaptivity of wavelets and diagonalized Fourier domain representation of the convolution operator to solve the deconvolution problem.

5.1 The WaRD Algorithm

The WaRD algorithm consists of the following steps.

1. **Regularized Fourier-domain Inversion:** Instead of using a pure inverse, we use a LTI Fourier-domain regularized filter. As shown in Figure 5.1, this filter partially attenuates noise by mildly shrinking the noisy coefficients obtained after inversion of $\mathcal{H}$ in the Fourier-domain using weights $R_\alpha(f_n)$ (refer Eqn. (2.7)). The signal obtained after quasi-inversion is

$$\widetilde{X}_\alpha(f_n) = \left(\frac{1}{H(f_n)}\right) R_\alpha(f_n)\, Y(f_n) \tag{5.1}$$

$$= R_\alpha(f_n)\, X(f_n) + \left(\frac{1}{H(f_n)}\right) R_\alpha(f_n)\, \Gamma(f_n) \tag{5.2}$$

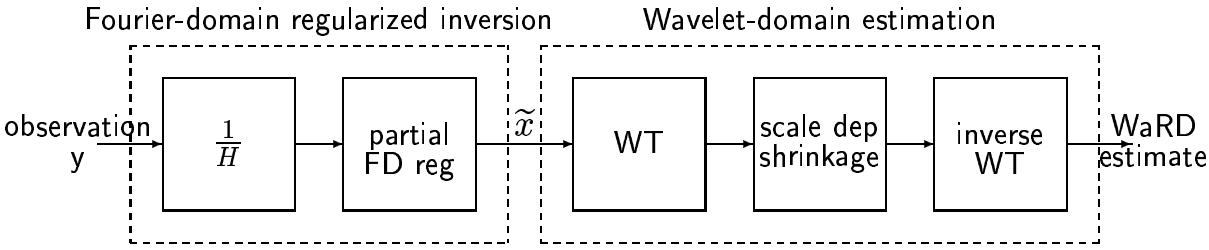


Figure 5.1 : *Block diagram of Wavelet-based regularized deconvolution (WaRD): The first block, Fourier domain regularized inversion, performs inversion followed by partial Fourier-domain regularization to achieve partial noise attenuation. Wavelet-domain estimation that is based on shrinkage of wavelet coefficients of $\widetilde{x}$ subsequently estimates the signal in the presence of noise colored during regularized inversion.*

where the parameter α that controls the amount of Fourier-domain shrinkage employed during quasi-inversion, is typically small ($\alpha < 1$). The choice of α is discussed in the later sections.

2. **Wavelet-domain Signal Estimation:** Since the amount of regularization used in the first stage is small, the estimate $\widetilde{X}_\alpha$ obtained after regularized inversion still contains some residual noise. We obtain the signal estimate by shrinking wavelet coefficients of $\widetilde{X}_\alpha$ at each scale according to the noise variance at that scale. We then employ the inverse DWT on the estimated signal wavelet coefficients to obtain the WaRD estimate of the input signal (refer Figure 5.1). We emphasize that wavelet-domain signal estimation remains effective since most of the signal energy is still captured by just a few large wavelet coefficients since the amount of distortion introduced by $R_\alpha(f_n)$ is minimal and the noise in these wavelet coefficients is not excessive, thanks to Fourier-domain regularization used during inversion.

5.2 Tradeoff: Distortion vs. Noise Attenuation

Noise reduction using Fourier-domain regularization comes at the cost of signal distortion and hence needs to be controlled. In other words, this raises the question: how to pick the right value for the regularization parameter α? The tradeoff is clear: On one hand, since regularization smears non-stationary signal features like edges and ridges, we would prefer α as small as possible. On the other hand, large α prevents excessive noise amplification during inversion which aids the wavelet-domain signal estimation. Thus α controls the bias-variance tradeoff in the WaRD system.

We seek to determine the *optimal* regularization parameter for the WaRD system by minimizing the overall MSE. This optimal amount of regularization would balance the amount of noise reduction carried out in the Fourier domain and the wavelet domain. The overall MSE is well-approximated by a proposed cost function that includes the distortion error due to the Fourier-domain regularized inversion stage and error incurred during wavelet-domain signal estimation stage.

5.3 The Cost Function

We derive the cost function that is approximately equal to the MSE of the WaRD estimate. The cost function assumes that ideal thresholding $T(.)$ is employed during signal estimation in the wavelet domain. Ideal 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.

$$T\left(\widetilde{\theta}^{\alpha}_{j,k}\right) = \begin{cases} \widetilde{\theta}^{\alpha}_{j,k}, & \text{if } |\theta^{\alpha}_{j,k}|^2 > \sigma_j^2(\alpha) \\ 0, & \text{if } |\theta^{\alpha}_{j,k}|^2 \leq \sigma_j^2(\alpha) \end{cases}, \tag{5.3}$$

where $\widetilde{\theta}^{\alpha}_{j,k}$ are the wavelet coefficients of the noisy signal $\widetilde{X}_{\alpha}(f_n)$ obtained after partially regularized inversion in (5.2), $\theta^{\alpha}_{j,k}$ are the wavelet coefficients of the the distorted

but noiseless input signal $R_\alpha(f_n)\,X(f_n)$, and $\sigma_j^2(\alpha)$ is the variance of noise $\frac{R_\alpha(f_n)}{H(f_n)}\Gamma(f_n)$ at wavelet scale j. Ideal thresholding assumes that the signals under consideration are known.

The deconvolution MSE for the WaRD system approximately consists of the signal distortion due to Fourier-domain regularization and the error due ideal thresholding that is used to achieve wavelet-domain noise shrinkage. So we propose a cost function $\widetilde{\mathrm{MSE}}(\alpha)$ given by

$$\widetilde{\mathrm{MSE}}(\alpha) \;:=\; \sum_{n=1}^{n=N}[1-R_\alpha(f_n)]^2\,|X(f_n)|^2 \;+\; \sum_{j,k}\min\big(|\theta_{j,k}^\alpha|^2,\,\sigma_j^2(\alpha)\big)\,. \qquad (5.4)$$

Here $\theta_{j,k}^\alpha$ denotes the wavelet coefficients of the input signal distorted by Fourier domain regularization, $R_\alpha(f_n)\,X(f_n)$, and $\sigma_j^2(\alpha)$ is the variance of colored noise $\frac{R_\alpha(f_n)}{H(f_n)}\Gamma(f_n)$ at scale j. The first term in $\widetilde{\mathrm{MSE}}(\alpha)$ is an estimate of the distortion in the input signal due to regularized inversion [15]. This distortion error is an increasing function of α. The second term is an estimate of the error due to ideal wavelet-domain thresholding [22]. This thresholding error is a decreasing function of α since both, the signal energy and the noise variance, reduce with increase in α. The optimal regularization parameter, denoted by $\alpha^\star$, corresponds to the minimum of $\widetilde{\mathrm{MSE}}(\alpha)$.

5.4 Accuracy of the cost function

We now show that the proposed cost function $\widetilde{\mathrm{MSE}}(\alpha)$ approximates the actual MSE of a WaRD system well. The cost function assumes that the total error is composed of independent contributions from the distortion error incurred during the Fourier-domain regularization stage and the error due to subsequent ideal wavelet-domain thresholding. We analyze two following two cases and their effects on the cost function $\widetilde{\mathrm{MSE}}(\alpha)$.

$|\theta^{\alpha}_{j,k}|^2 > \sigma^2_j(\alpha)$: When a wavelet coefficient of $R_\alpha(f_n)\,X(f_n)$ has greater energy than the noise variance $\sigma^2_j(\alpha)$, ideal thresholding retains the corresponding noisy wavelet coefficient intact (refer (5.3)). In this case, the error contribution due to ideal thresholding error term in $\widetilde{\mathrm{MSE}}(\alpha)$ is $\sigma^2_j(\alpha)$, which is exactly equal to the actual MSE.

$|\theta^{\alpha}_{j,k}|^2 \leq \sigma^2_j(\alpha)$: However, when $|\theta^{\alpha}_{j,k}|^2 \leq \sigma^2_j(\alpha)$, ideal thresholding sets the corresponding wavelet coefficient to zero. The actual error is the energy of the wavelet coefficient of the original signal x, $|\theta_{j,k}|^2$. However, the contribution of the proposed cost function $\widetilde{\mathrm{MSE}}(\alpha)$ is $|\theta_{j,k}-\theta^{\alpha}_{j,k}|^2+|\theta^{\alpha}_{j,k}|^2$. This contribution is approximately equal to $|\theta_{j,k}|^2$ when the distortion $|\theta_{j,k}-\theta^{\alpha}_{j,k}|^2$ is small in comparison to $|\theta^{\alpha}_{j,k}|^2$. So, for small values of α, which typically is the case, the distortion is comparatively small and hence the cost function is fairly accurate. The proposed cost function is always within a factor of the error 2 within the actual error for any positive value of α. However, we believe that these extreme cases are not generally encountered in practice.

So the proposed cost function is a good approximation of the actual error incurred (assuming ideal thresholding); hence $\widetilde{\mathrm{MSE}}(\alpha)$ can be minimized to find the regularization parameter that balances Fourier-domain regularized inversion and effective wavelet-domain signal estimation.

5.5 Optimal α for each Scale

An interesting generalization to having a single Fourier-domain regularized inverse filter for all subbands is to have Fourier-domain regularized inverse filters with a different regularization parameter for every wavelet scale (refer Figure 5.2). In this case, we need to determine the optimal regularization parameter for each wavelet

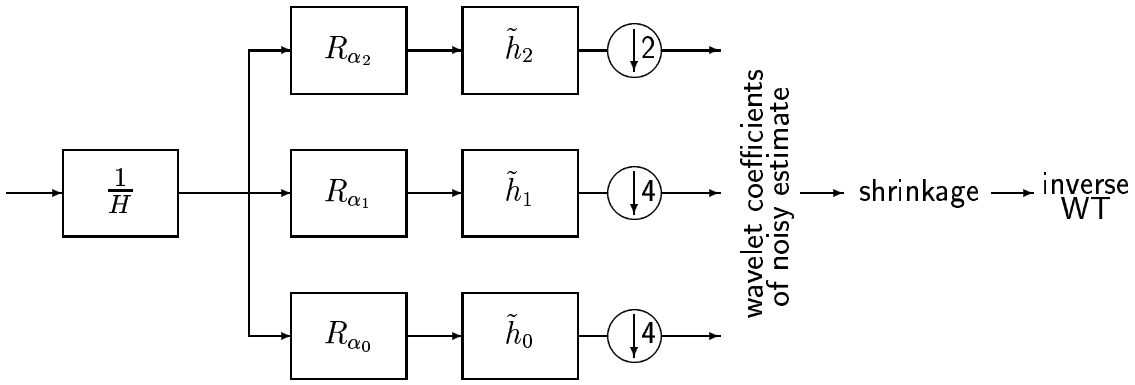


Figure 5.2 : *Different regularization parameters for each wavelet scale: In the figure, the filters $\widetilde{h}_0$, $\widetilde{h}_1$ and $\widetilde{h}_2$ are the equivalent, cumulative filters at each scale in a 2-level decomposition DWT filter-bank. Such an equivalent filter-bank representation can be obtained by transforming the filters in the filter bank using the noble identities [23, pp. 119] and then grouping the synthesis filters and the decimators together. Different regularization parameters, α_0, α_1, and α_2 are used at each scales as shown.*

subband. Such a generalization helps making the cost function separable with respect to the regularization parameter of each scale thereby simplifying the analysis. It turns out (refer appendix A) that the optimal regularization parameter $\alpha_p^\star$ for the scale p satisfies

$$\alpha_p^\star \approx \frac{1}{2^p} \# \left(|\theta_{p,k}^{\alpha_p^\star}| > \sigma_p^2(\alpha_p^\star) \right), \tag{5.5}$$

where $\theta_{p,k}^{\alpha_p}$ denotes the wavelet coefficients of the distorted input signal $R_\alpha(f_n)\, X(f_n)$ at scale p, and $\sigma_p^2(\alpha_p)$ is the variance of noise $\left(\frac{R_\alpha(f_n)}{H(f_n)}\right) \Gamma(f_n)$ at scale p. $\# \left(|\theta_{p,k}^{\alpha_p}| > \sigma_p^2(\alpha_p) \right)$ denotes the number of wavelet coefficients $\theta_{p,k}^{\alpha_p}$ that have energy greater than the noise variance $\sigma_p^2(\alpha_p)$.

Expression (5.5) reveals that the selection of the optimal regularization parameter is heavily influenced by the distribution of the wavelet coefficients relative to the variance of the colored noise. Expression (5.5) is a *recursive* expression since both

sides of the equation depend on $\alpha_p^\star$.

Consider a signal such that the signal energy in some wavelet scale p is concentrated in very small proportion of the wavelet coefficients at that scale, i.e. few wavelet coefficients are large while most of the other wavelet coefficients are nearly zero. This economical distribution does not change significantly when the signal is distorted by using a small amounts of regularization because the derivative of the total distortion introduced at zero regularization is 0 (put $\alpha_p = 0$ in Eqn. (A.5)). On the other hand, from Eqn. (A.16), one can infer that the rate of decrease of the noise variance in the wavelet scale is maximum at $\alpha_p = 0$. So, the variance of the colored noise in this scale decreases rapidly at small values of α_p. The optimal regularization parameter condition suggests that the noise variance be reduced using regularization so that a sufficient number of wavelet coefficients of the signal stand out above the noise variance.

We emphasize that the optimal regularization parameter $\alpha_p^\star$ is never zero. If $\alpha_p = 0$ satisfies the optimality condition, it would imply that the noise variance is higher than the energy of all the wavelet coefficients. Hence using $\alpha_p = 0$ would result in all wavelet coefficients at scale p being shrunk to zero during wavelet-domain estimation stage! For most signals, a significant proportion of the wavelet coefficients are very small; so $\alpha_p = 1$ will not satisfy (5.5). Hence $\alpha_p = 1$ also is not an optimal choice for the regularization parameter.

Several values of α_p could potentially satisfy the optimality criterion. The choice of the optimal regularization parameter in such a case is not clear. However, this is not a significant issue in practice since choosing any α_p sufficiently greater than zero ($\alpha_p \approx 0.2$) was found to provide estimates comparable to that obtained by choosing the optimal α.

Thus, the expression for the optimal regularization parameter quantifies the no-

tion of balancing Fourier-domain inversion and wavelet-domain estimation within a wavelet-based deconvolution approach.

5.6 Optimality of WaRD

We emphasize that the existing wavelet-based deconvolution algorithms of [6, 7, 24] are special cases of WaRD[1] with $\alpha = 0$. By construction, WaRD includes the value $\alpha = 0$ in the search-space for the optimal $\alpha^\star$ that it uses. Hence WaRD enjoys all the good properties of Donoho's wavelet-based deconvolution such as optimal rate of error decay for dilation-homogeneous operators.

The optimal $\alpha^\star$ is never exactly zero at a finite resolution N (though it can tend to zero with increasing N). Hence, WaRD will actually outperform wavelet-based deconvolution methods described in [6, 7, 24] in MSE-terms at a given resolution.

Techniques based on conventional wavelet-based deconvolution [7, 11] described in Section 4 are in general not applicable when $\mathcal{H}$ is not invertible. However, thanks to optimal amount of regularization WaRD uses, WaRD give excellent estimates even when $\mathcal{H}$ is not invertible.

[1] For comparisons with [6], the use of a similar adapted wavelet basis would be required

Chapter 6

WaRD implementation

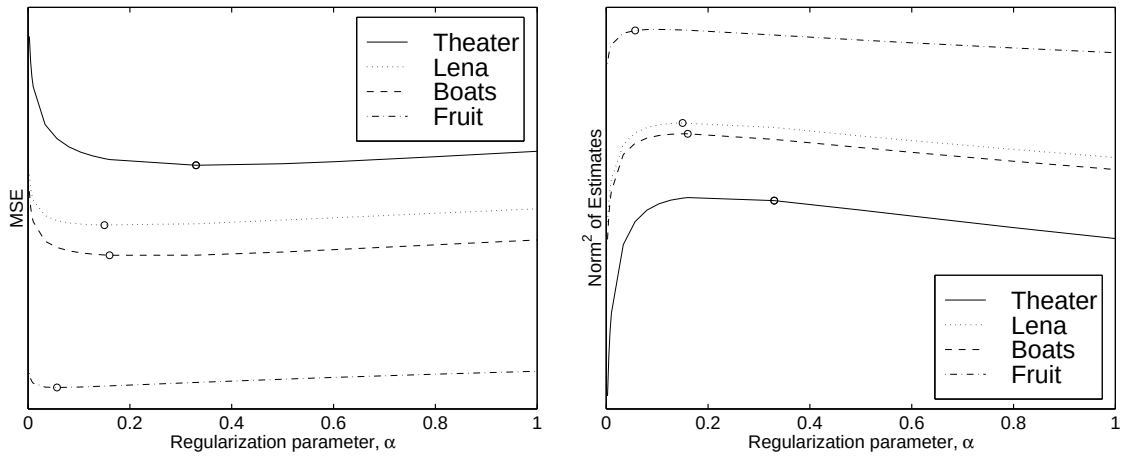
Several issues crop up during the implementation of the WaRD algorithm in practice.

6.1 Choice of σ^2 and $|X(f)|^2$

The WaRD algorithm as outlined in Section 5 assumes knowledge of the variance σ^2 of additive noise γ and the Fourier spectrum $|X(f)|^2$ of the input signal. However, these are typically unknown in practice and hence need to be estimated. The variance of γ can be reliably estimated from the observation y by using a median estimator in the finest wavelet scale[5]. The choice of $|X(f)|^2$ is however more tricky. Any crude estimate of the spectrum of $|X(f)|^2$ typically has attenuated high-frequency components. Employing such an estimate during regularized inversion step in WaRD would eliminate all the high frequency components in the final estimate as well. So we believe that the use of prior information on $|X(f)|^2$ instead of an estimate is appropriate. For the purpose of understanding the WaRD algorithm and gauging its potential, we have assumed knowledge of σ^2 and $|X(f)|^2$ in our experiments.

6.2 Empirical Estimation of α

Condition (5.5) provides the necessary condition for the regularization parameters used during regularized inversion to optimally balance Fourier-domain regularization and wavelet-domain estimation. However, this criterion cannot be used in practice because the expression assumes knowledge of the distorted wavelet coefficients of



(a) (b)

Figure 6.1 : *(a) The MSE of WaRD estimates is plotted against the regularization parameter α for different images. Ideally, we would like to choose the regularization parameter that results in the lowest MSE (denoted by the small circles). Note that the MSE does not change significantly when any α slightly higher the optimal value is chosen. (b) The norm of the WaRD estimate obtained at different regularization parameters is plotted against α. Note that the α that achieves the lowest MSE is located near the knee point of the curve. So in practice, near-optimal performance can be achieved by choosing any α near the knee point of the norm curve.*

the *unknown original signal*. Further, the derivation for the optimal regularization parameter is derived assuming ideal thresholding which cannot be implemented in practice. In practice, the optimal regularization parameter is also influenced by the wavelet-domain estimation algorithm used in the WaRD system.

This raises the necessity to resort to empirical techniques to find the regularization parameters. In the course of our simulations, we observed that the improvement gained by using different empirical regularization parameters for each wavelet scale is almost negligible. Hence we focus on estimating an optimal regularization parameter that is common for all wavelet scales.

Figure 6.1(a) shows the plot of the MSE of WaRD system against Fourier-domain regularization parameter α for different images. We observe that the typical MSE

decreases very rapidly when α is small but changes gradually thereafter. Further we can also infer that choosing any α sufficiently large gives near-optimal MSE performances. The robustness of the WaRD system performance to small changes in the regularization parameter about the optimal value makes the estimation of a near-optimal regularization parameter easy. We exploit this robustness to come up with a simple scheme to estimate the near-optimal regularization parameter.

Figure 6.1(b) shows the typical plot of the l_2 norm of the WaRD estimate against Fourier-domain regularization parameter α for the same images. We observe that similar to the behavior of the MSE, the WaRD solution norm also changes very rapidly when α is small but the change is gradual thereafter. The similarity in the trend observed in Figure 6.1(a) and (b) is expected because the MSE, WaRD estimate norm and the actual signal norm are related to each other via the triangle inequality. Since the the error of the estimate is essentially unchanged when the norm of the solution remains practically constant, we can empirically determine the regularization parameter by choosing an α near the knee point in the plot of the WaRD solution norm against α. Such an empirically determined α gives near-optimal results.

Our experience showed that due to the robustness of the WaRD system to the choice of the α, even such a search can be avoided. Simply choosing $\alpha \approx 0.2$ gave satisfactory results for a wide class of images.

6.3 Choice of Wavelet-domain Estimation Scheme

A variety of wavelet-domain estimation techniques can be used in practice. The choice of a good wavelet-domain estimation scheme influences the final performance significantly. Commonly studied wavelet-domain shrinkage schemes include hard-thresholding [25], soft-thresholding [26] and a variety of other similar techniques

based on non-linear wavelet domain shrinkage [27, 28]. Recently, the wavelet-domain Wiener shrink estimation algorithm was proposed by Ghael et. al. [29] and analyzed further by Choi [30]. It was observed that this technique outperforms the conventional wavelet-domain estimation schemes that employ hard and soft thresholding in the MSE sense. Estimation using a wavelet-domain Wiener shrinkage consists of the following steps: First obtain a rough estimate of the input signal using conventional wavelet-domain techniques. Use this estimate to obtain a final refined estimate by employing a Wiener filter on each *wavelet* coefficient.

Significant improvements can also be obtained by using a redundant, shift-invariant DWT. Wavelet-domain estimation schemes that use the DWT are not shift-invariant, shifts of y will result in different estimates. The redundant, shift-invariant DWT yeilding significantly improved estimation [31, 32] by simply averaging over all possible shifts of the observation y at no significant increase in the overall computational cost. So we employ a redundant, shift-invariant DWT with a wavelet-domain Wiener shrinkage to estimate the input signal.

Chapter 7

Results

We illustrate the performance of the WaRD algorithm using simulations in 1-d and 2-d.

We first compare the WaRD with other methods for the 1-d deconvolution problem presented in the Introduction. In order to observe the behavior of each method for both smooth and edgy regions, we take for x a concatenation of Donoho's *Blocks* and *Heavisine* signals [5] (normalized to be zero mean and unit energy). We employ Daubechies length-4 wavelets throughout. Figure 1.2(a) depicts the signal. For the system, we take the example of [6, pp. 459]

$$H(f) = \begin{cases} 1, & |f| \in [0, 0.25] \\ 2 - 4\,|f|, & |f| \in (0.25, 0.5] \end{cases} \tag{7.1}$$

with DFT frequencies f normalized to $(-0.5, 0.5]$ (see Figure 1.2(b)). The corrupting noise variance was $\sigma^2 = 4 \times 10^{-6}$. Figure 1.2(c) plots the blurred, noisy signal. The inverse filter frequency response H^{-1} amplifies all high frequency noise components (see Figure 1.2(d)).

Use of a pure inverse $\frac{1}{H}$ amplifies all the high frequency noise components thereby giving an extremely noisy estimate (see Figure 1.2(e)) of the input signal. The Wiener filter estimate (Figure 1.2(f)) was implemented using $P_x(f) = |X(f)|^2$ in (2.4). The Wiener filter bases its deconvolution on the signal-to-noise ratio at each frequency. The Fourier-domain that is exploited by the Wiener filter is not well-suited to represent the signal containing spatially localized features such as edges. Consequently,

the Wiener deconvolution is ineffective.

WaD also fails to provide a satisfactory solution in this case. Due to the null in the frequency response of the system at $H(0.5) = 0$, all the wavelet coefficients at the finest scale are corrupted by infinite variance noise after inversion using H^{-1}; hence these coefficients are lost during wavelet-domain estimation.

The mirror wavelet basis deconvolution method of [6, 11] adapts a wavelet packet basis to the colored noise in the inverted data $\widehat{X}(f) = H^{-1}(f)\,Y(f)$. The frequency splits of the mirror wavelet basis for H are shown with dashed lines in Figure 1.2(d). The adapted basis aims to isolate the singularity at $f = 0.5$. The signal estimation step was implemented by hard-thresholding the coefficients of a shift-invariant mirror wavelet basis. This algorithm outperforms the methods of [7, 24] and the standard Wiener filter in this case (see Figure 1.2(g))[1].

Figure 1.2(h) plots the WaRD estimate obtained using $\alpha^\star = 0.06$. Wavelet-domain Wiener shrinkage was applied to the coefficients of a shift-invariant DWT for the wavelet-domain signal estimation stage. The crude estimate required by the wavelet-domain Wiener shrinkage was obtained by using shift-invariant hard thresholding using Daubechies length-6 wavelets. WaRD outperforms the other algorithms in terms of both estimate's visual quality and MSE.

Next, we consider image restoration using WaRD. The input x is the 256×256 Lena image (normalized to zero mean and unit energy) and the discrete-time system response h is a 2-d, 4×4-point smoother $[1 \;\; 1 \;\; 1 \;\; 1]^T [1 \;\; 1 \;\; 1 \;\; 1]$. Such a response is commonly used as a model for blurring due to a square scanning aperture such as in a

[1]Even though $H(f)$ is not invertible, the wavelet packet deconvolution method does not fail, because the location and the order of the zero of H are such that only a few, high-frequency signal wavelet coefficients are obliterated by the infinite noise amplification at $f = 0.5$. This method will fail when $H(f)$ has zeros of arbitrary location and order, however.

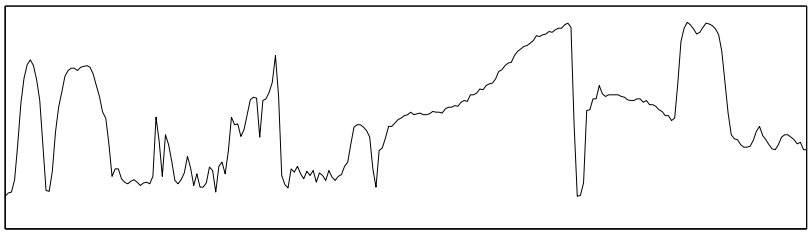
CCD camera [2]. The noise variance was set to $\sigma^2 = 9.6 \times 10^{-7}$. Figure 7.1 illustrates the desired x, the observed y, the Wiener filter estimate $\widehat{x}_1$, and the WaRD estimate for $\alpha^\star = 0.2$. The value for α was chosen to be close to the knee point in the graph shown in Figure 6.1(b). The methods of [6, 7] are not applicable in this situation, due to the many zeros in $H(f_x, f_y)$. (The Wiener filter outperformed Mallat's adapted wavelet basis in this case.) The WaRD estimate is clearly better that the Wiener estimate in overall visual quality and MSE. However, we point out that the lines on the hat are almost lost in the WaRD estimate in comparision to the Wiener estimate.

The difference in the quality of the estimates obtained using the Wiener filter and WaRD can be grasped by viewing the cross-sections through the images. Figure 7.2(a), (b), (c), and (d) shows the cross-sections through row 220 of the original image, the observed image, the Wiener estimate and the WaRD estimate respectively. The Wiener estimate cross-section shown in Figure 7.2(d) illustrates the failure of the Wiener technique to adapt to the smooth regions and the edges in the image simultaneously. This lack of spatial-localization reflects as ripples in the Wiener estimate. In contrast, Figure 7.2(d) clearly illustrates the spatial-adaptivity of WaRD. We observe the smooth regions and the edges are simultaneously preserved.

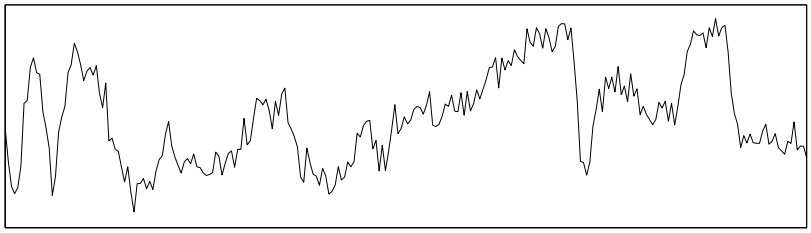
(a) (b)

(c) (d)

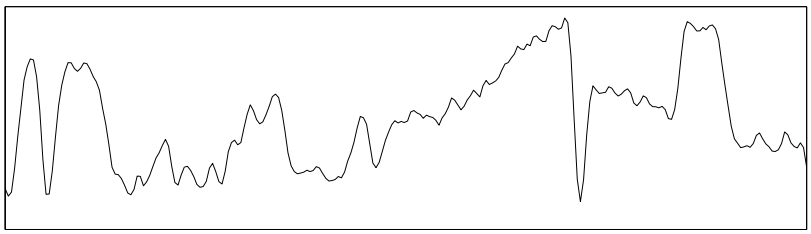
Figure 7.1 : *(a) Lena image x (256×256). (b) Observed image y (MSE=0.2473). (c) Wiener filter estimate (MSE=0.0498). The ripples in the image result because the Fourier basis used by the Wiener filter have support over the entire spatial domain. (d) WaRD with $\alpha = 0.2$ (MSE=0.0427). In contrast 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 the lines on the hat are lost during wavelet domain estimation.*



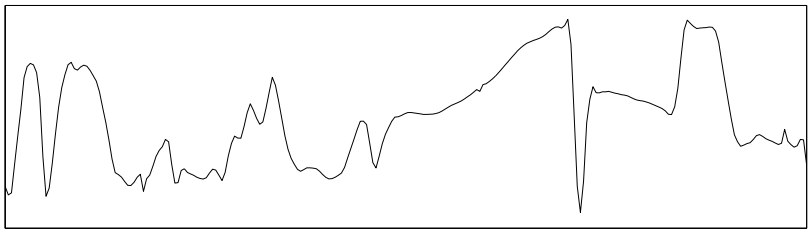
(a)



(b)



(c)



(d)

Figure 7.2 : *(a) Cross-section of original image x (row 220 from Figure 7.1a). The image cross-section contains 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 technique. The Wiener deconvolution estimate exhibits ringing artifacts. (d) Cross-section of the hybrid WaRD estimate. Controlled Fourier-domain regularization ensures that residual noise can be effectively tackled by subsequent spatially-adaptive wavelet domain estimation. WaRD preserves smooth regions and edges simultaneously even when the blurring system $\mathcal{H}$ is non-invertible.*

Chapter 8

Conclusions

In this thesis, we have proposed an efficient multi-scale deconvolution algorithm WaRD that optimally combines Fourier-domain regularized inversion and wavelet-domain signal estimation. The WaRD can be potentially applied in a wide variety of applications such as satellite imagery, seismic deconvolution and channel equalization to obtain enhanced performances.

For spatially varying signals, the WaRD outperforms the LTI Wiener filter and other wavelet-based deconvolution algorithms in terms of both visual quality and MSE performance. Since WaRD subsumes WaD, WaRD also enjoys 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 WaD at any fixed resolution. Furthermore, WaRD continues to provide a good estimate of the original signal even in the presence of any ill-conditioned system. The computational complexity of the WaRD algorithm is just $O(N \log_2^2 N)$ where 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 are greater than the variance noise colored by regularized inversion. From the expression for the optimal regularization parameter it follows that for finite data samples, employing inversion without 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 sub-optimal. In essence, the optimal regularization parameter is simultaneously determined by the frugality of the wavelet representation of the input signal *and* the Fourier-domain structure of the convolution operator.

However, fortunately, the final performance is observed to be quite insensitive to changes in the value of the regularization parameter around the optimal value. So, a near-optimal regularization parameter can be obtained from the norm of the WaRD solution for different regularization parameters. As a guide, in simulations spanning many real-world images and convolution systems, $\alpha^\star$ was almost always lay in the range $[0.2, 0.3]$. In the 1-d example above, the optimal α turned out to be small ($\alpha^\star = 0.06$) because the test signal was represented very economically in the wavelet domain. However, choosing a $\alpha^\star$ in the range $[0.2, 0.3]$ gave near-optimal results as well.

There are several avenues for future WaRD related research:

1. We have focused on scalar processing during wavelet-domain estimation. However, there exist dependencies between the wavelet coefficients that can be exploited to improve the performance of wavelet-based deconvolution systems. We are currently working on combining WaRD concepts with hidden Markov model (HMM) tree based wavelet-estimation [33] to obtain a better deconvolution system. We believe that exploiting such inter-dependencies in the wavelet domian will help preserve edges and other spatially localized phenomena (such as the lines on *lena*'s hat better.

2. One interesting twist to the approach to wavelet-based deconvolution is 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 Vaguellete-Wavelet Decomposition (VWD) has been studied by Silverman and Abramovich recently

[34]. The salient point of such a technique is that the wavelet-domain estimation has to deal with white noise instead of the complicated colored noise. However, in such a case the reconstruction basis is no longer a signal-adapted wavelet-basis, but a hybrid basis that is not spatially localized. Further, this technique is again not applicable when the $\mathcal{H}$ is not invertible. We find the construction of a universally, applicable, hybrid deconvolution scheme that lies in between WVD and VWD promising and challenging.

Appendix A

Deriving the Optimal Regularization Parameter α

The MSE of the estimate obtained by using WaRD is approximated by the sum of the distortion incurred during Fourier-domain regularized inversion and the error due to ideal hard thresholding of the wavelet coefficients.

$$\begin{aligned}\widetilde{\mathrm{MSE}}(\alpha) &= \sum_{n=1}^{n=N}[1-R_\alpha(f_n)]^2\,|X(f_n)|^2 \;+\; \sum_{j,k}\min\left(|\theta_{j,k}^\alpha|^2,\,\sigma_j^2(\alpha)\right), \qquad \text{(A.1)}\\ &= \sum_{n=1}^{N}\frac{\alpha^2\sigma^4|X(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha\sigma^2)^2} + \left(\sum_{j=j_0}^{J}\sum_{k=1}^{2^j}\min(|\theta_{j,k}^\alpha|^2,\sigma_j^2(\alpha))\right),\end{aligned}$$

where $\{\theta_{j,k}^\alpha\}$ are the wavelet coefficients of the distorted but noise-free signal $R_\alpha(f_n)X(f_n)$ and $\sigma_j^2(\alpha)$ is the variance of noise $\left(\frac{R_\alpha(f_n)}{H(f_n)}\right)\Gamma(f_n)$ at wavelet scale j (refer Eqn. (5.2). $j=j_0$ denote the coarsest scale coefficients and $j=J$ denote the finest resolution coefficients.

By grouping the distortion terms according to the frequency subbands of a wavelet system (refer Figure 4.2), we have

$$\sum_{n=1}^{N}\frac{\alpha^2\sigma^4|X(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha\sigma^2)^2} = \sum_{j=j_0}^{J}\sum_{n=2^{j-1}}^{2^j-1}\frac{2\alpha^2\sigma^4|X(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2+\alpha\sigma^2)^2} \qquad \text{(A.2)}$$

Therefore,

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

A.1 Different α for Each Wavelet Scale

$\widetilde{\mathrm{MSE}}(\alpha)$ in A.3 is derived assuming that the the same regularization parameter α for each wavelet scale. An interesting generalization is to have different regularization parameters α_j for each wavelet scale j (refer Figure 5.2). Assuming that the wavelets at different scales have no overlapping frequencies, the cost function can be written as a function of the different regularization parameters as

$$\begin{aligned}\widetilde{\mathrm{MSE}}(\alpha_{\tilde{j}_0}, \alpha_{j_0}, \ldots, \alpha_{J-1}) \approx & \sum_{j=j_0}^{J} \left(\sum_{n=2^{j-1}}^{2^j-1} \frac{2\alpha_j^2 \sigma^4 |X(f_n)|^2}{(|H(f_n)|^2|X(f_n)|^2 + \alpha_j\sigma^2)^2} \right) \\ & + \sum_{j=j_0}^{J} \left(\sum_{k=1}^{2^j} \min(|\theta_{j,k}^{\alpha_j}|^2, \sigma_j^2(\alpha_j)) \right) \end{aligned} \tag{A.3}$$

Note that the cost function $\widetilde{\mathrm{MSE}}(\alpha)$ is separable with respect to the different regularization parameters α_j. We attempt to find the optimal regularization parameters at each scale by partially differentiating the cost function with respect to α_j and setting the derivative to zero.

A.2 Differentiating Distortion Terms

We denote the total distortion error by

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

Taking the partial derivative of $\mathrm{D}(\alpha_{j_0}, \ldots, \alpha_J)$ with respect to α_p we have

$$\frac{\partial}{\partial \alpha_p} \left(\mathrm{D}(\alpha_{j_0}, \ldots, \alpha_J) \right) = \sum_{n=2^{p-1}}^{2^p-1} \frac{4\alpha_p \sigma^4 |H(f_n)|^2 |X(f_n)|^4}{(|H(f_n)|^2|X(f_n)|^2 + \alpha_p\sigma^2)^3} \tag{A.5}$$

Note that the partial derivative of the distortion is zero at $\alpha_p = 0$. For the sake of convenience, we define

$$C(\alpha_p) = \sum_{n=2^{p-1}}^{2^p-1} \frac{4\sigma^4 |H(f_n)|^2 |X(f_n)|^4}{(|H(f_n)|^2|X(f_n)|^2 + \alpha_p\sigma^2)^3}. \tag{A.6}$$

$$\text{Therefore,} \qquad \frac{\partial}{\partial \alpha_p}\left(\mathrm{D}(\alpha_{j_0},\ldots,\alpha_J)\right) = \alpha_p\, C(\alpha_p) \tag{A.7}$$

A.3 Differentiating Ideal Thresholding Terms

The error due to ideal thresholding at each wavelet scale p is given by

$$\begin{aligned}
\sum_{k=1}^{2^p} \min(|\theta_{p,k}^{\alpha_p}|^2, \sigma_p^2(\alpha_p)) &= \sum_{|\theta_{p,k}^{\alpha_p}|<\sigma_p(\alpha_p)} (w_{\alpha_p}^{\alpha_p})^2 + \sum_{|\theta_{p,k}^{\alpha_p}|\geq\sigma_p(\alpha_p)} \sigma_p^2, \\
&= 2^p \int_0^{\sigma_p(\alpha_p)} f(w_p^{\alpha_p})\,(w_p^{\alpha_p})^2\, dw_p^{\alpha_p} \qquad \text{(A.8)} \\
&\quad + 2^p \left(1 - \int_0^{\sigma_p(\alpha_p)} f(w_p^{\alpha_p})\, dw_p^{\alpha_p}\right) \sigma_p^2(\alpha_p), \qquad \text{(A.9)}
\end{aligned}$$

where

$$f(w_p^{\alpha_p}) := \frac{1}{2^p} \sum_{k=1}^{2^p} \delta(w_p^{\alpha_p} - |w_{(p,k)}^{\alpha_p}|) \tag{A.10}$$

acts like the probability density of the magnitude of the wavelet coefficients of the distorted but noise-free signal at scale p. $f(w_p^{\alpha_p})$ is not smooth but can be approximated arbitrarily closely by some smooth density $\tilde{f}(w_p)$. Using this smooth density $\tilde{f}(w_p^{\alpha_p})$ instead of $f(w_p^{\alpha_p})$ in expression A.9,

$$\begin{aligned}
\sum_{k=1}^{2^p} \min(|\theta_{p,k}^{\alpha_p}|^2, \sigma_p^2(\alpha_p)) &\approx 2^p \left(\int_0^{\sigma_p(\alpha_p)} \tilde{f}(w_p^{\alpha_p})\,(w_p^{\alpha_p})^2\, dw_p^{\alpha_p}\right) \\
&\quad + 2^p \left(1 - \int_0^{\sigma_p(\alpha_p)} \tilde{f}(w_p^{\alpha_p})\, dw_p^{\alpha_p}\right) \sigma_p^2(\alpha_p), \qquad \text{(A.11)}
\end{aligned}$$

Let the total thresholding error be denoted by

$$\mathrm{T}(\alpha_{j_0},\ldots,\alpha_J) := \sum_{j=j_0}^{J} \sum_{k=1}^{2^j} \min(|\theta_{j,k}^{\alpha_j}|^2, \sigma_j^2(\alpha_j)) \tag{A.12}$$

Taking the partial derivative of $\mathrm{T}(\alpha_{j_0},\ldots,\alpha_J)$ with respect to α_p we have

$$\frac{\partial}{\partial \alpha_p}\left(\mathrm{T}(\alpha_{j_0},\ldots,\alpha_J)\right) = \frac{\partial}{\partial \alpha_p}\left(\sum_{k=1}^{2^p} \min(|\theta_{p,k}^{\alpha_p}|^2, \sigma_p^2(\alpha_p))\right)$$

$$\approx 2^p \frac{\partial}{\partial \alpha_p} \left[\left(\int_0^{\sigma_p(\alpha_p)} \tilde{f}(w_p^{\alpha_p}) \, (w_p^{\alpha_p})^2 \, dw_p^{\alpha_p} \right) + \left(1 - \int_0^{\sigma_p(\alpha_p)} \tilde{f}(w_p^{\alpha_p}) \, dw_p^{\alpha_p} \right) \sigma_p^2(\alpha_p) \right] \tag{A.13}$$

We recollect that the purpose of Fourier-domain regularization is to shrink noise components which explode during inversion of ill-conditioned $\mathcal{H}$; the residual noise is tackled using shrinkage in the wavelet-domain. Since excessive Fourier-domain regularized inversion distorts the signal and smears the spatially varying features such as edges, we aim to keep the amount of Fourier-domain regularization minimal.

Fortunately, the rogue noise components at frequencies where $|H(f_n)| \approx 0$, which prevent effective wavelet-domain signal estimation, are greatly attenuated even for small values of α_p (refer Eqn. (2.7). Once these rogue noise components are attenuated, we expect wavelet-domain signal estimation to be effective. Hence, it is justified to expect a optimal regularization parameter that achieve drastic noise attenuation but minimal distortion. So, we also expect that any small change in the α_p about its optimal value would cause much larger changes in amount of noise attenuated than the amount of signal distorted. So we neglect any changes in distribution of the wavelet coefficients with respect to α_p in comparison to changes in the variance the noise $\sigma_p^2(\alpha_p)$. This assumption simplifies Eqn. (A.13) to

$$\begin{aligned} & 2^p \frac{\partial}{\partial \sigma_p} \left[\left(\int_0^{\sigma_p(\alpha_p)} \tilde{f}(w_p^{\alpha_p}) \, (w_p^{\alpha_p})^2 \, dw_p^{\alpha_p} \right) + \left(1 - \int_0^{\sigma_p(\alpha_p)} \tilde{f}(w_p^{\alpha_p}) \, dw_p^{\alpha_p} \right) \sigma_p^2(\alpha_p) \right] \frac{d\sigma_p(\alpha_p)}{d\alpha_p} \\ = \; & 2^p \left[2\sigma_p(\alpha_p) \left(1 - \int_0^{\sigma_p(\alpha_p)} \tilde{f}(w_p^{\alpha_p}) \, dw_p^{\alpha_p} \right) \right] \frac{d\sigma_p(\alpha_p)}{d\alpha_p} \\ \approx \; & \#\left(|\theta_{p,k}^{\alpha_p}| > \sigma_p^2(\alpha_p) \right) \frac{d\sigma_p^2(\alpha_p)}{d\alpha_p}, \end{aligned} \tag{A.14}$$

where $\#\left(|\theta_{p,k}^{\alpha_p}|^2 > \sigma_p^2(\alpha_p) \right)$ is the number of wavelet coefficients at scale p that possess energy greater than the noise variance $\sigma_p^2(\alpha_p)$ in scale p.

The noise variance at scale p is equal to

$$\begin{aligned}\sigma_p^2(\alpha_p) &= \sum_{n=1}^{N} \frac{|\hat{\psi}_p(f_n)|^2 |H(f_n)|^2 |X(f_n)|^4}{(|H(f_n)|^2|X(f_n)|^2 + \alpha_p\sigma^2)^2}\sigma^2 \\ &\approx 2\sum_{n=2^{p-1}}^{2^p-1} \frac{(2^{-\frac{p}{2}})^2\, |H(f_n)|^2 |X(f_n)|^4}{(|H(f_n)|^2|X(f_n)|^2 + \alpha_p\sigma^2)^2}\sigma^2, \qquad \text{(A.15)}\end{aligned}$$

where the approximation in Eqn. (A.15) arises from assuming that the wavelets at scale p are perfectly bandlimited. The derivative of $\sigma_p^2(\alpha_p)$ with respect to α_p is given by

$$\begin{aligned}\frac{d\sigma_p^2(\alpha_p)}{d\alpha_p} &\approx \frac{d}{d\alpha_p}\left[2\sum_{n=2^{p-1}}^{2^p-1} \frac{(2^{-\frac{p}{2}})^2\, |H(f_n)|^2 |X(f_n)|^4}{(|H(f_n)|^2|X(f_n)|^2 + \alpha_p\sigma^2)^2}\sigma^2\right] \\ &= 2\sum_{n=2^{p-1}}^{2^p-1} \frac{-2\sigma^4 2^{-p}\, |H(f_n)|^2 |X(f_n)|^4}{(|H(f_n)|^2|X(f_n)|^2 + \alpha_p\sigma^2)^3} \\ &= -\sum_{n=2^{p-1}}^{2^p-1} \frac{4\sigma^4 2^{-p}\, |H(f_n)|^2 |X(f_n)|^4}{(|H(f_n)|^2|X(f_n)|^2 + \alpha_p\sigma^2)^3} \\ &= -2^{-p} C(\alpha_p) \qquad \text{(A.16)}\end{aligned}$$

Combining Eqn. (A.14) and Eqn. (A.16) we have

$$\frac{\partial}{\partial\alpha_p}\left(\mathrm{T}(\alpha_{j_0},\ldots,\alpha_J)\right) = -\frac{1}{2^p}\#\left(|\theta_{p,k}^{\alpha_p}| > \sigma_p^2(\alpha_p)\right)\left(C(\alpha_p)\right) \qquad \text{(A.17)}$$

The partial derivative of $\widetilde{\mathrm{MSE}}(\alpha_{\tilde{j}_0}, \alpha_{j_0},\ldots,\alpha_{J-1})$ with respect to α_p is

$$\frac{\partial}{\partial\alpha_p}\left(\widetilde{\mathrm{MSE}}(\alpha_{\tilde{j}_0}, \alpha_{j_0},\ldots,\alpha_{J-1})\right) \approx \frac{\partial}{\partial\alpha_p}\left(\mathrm{D}(\alpha_{j_0},\ldots,\alpha_J) + \mathrm{T}(\alpha_{j_0},\ldots,\alpha_J)\right) \qquad \text{(A.18)}$$

Using (A.7) and (A.17),

$$\begin{aligned}\frac{\partial}{\partial\alpha_p}\left(\widetilde{\mathrm{MSE}}(\alpha_{\tilde{j}_0}, \alpha_{j_0},\ldots,\alpha_{J-1})\right) &\approx \alpha_p\, C(\alpha_p) - \frac{1}{2^p}\#\left(|\theta_{p,k}^{\alpha_p}| > \sigma_p^2(\alpha_p)\right) C(\alpha_p) \\ &= \left(\alpha_p - \frac{1}{2^p}\#\left(|\theta_{p,k}^{\alpha_p}| > \sigma_p^2(\alpha_p)\right)\right) C(\alpha_p) \qquad \text{(A.19)}\end{aligned}$$

Setting Eqn. (A.19) to zero, we get the optimal regularization parameter

$$\alpha_p^\star \approx \frac{1}{2^p}\#\left(|\theta_{p,k}^{\alpha_p^\star}| > \sigma_p^2(\alpha_p^\star)\right) \qquad \text{(A.20)}$$

Appendix B

Error Decay Rates for LTI filters

Let the input x be a piece-wise polynomial signal as shown in Figure B.1. Since x is discontinuous, its energy decays as $O(\frac{1}{f})$ in the frequency domain. Consider N equally spaced samples of the signal. In the Fourier domain, the energy change in the first N Fourier coefficients due to aliasing is less than a multiplicative factor of two, hence we neglect the effects of aliasing and approximate the the magnitude of the DFT coefficients of the sampled signal x as

$$|X(f_n)| \approx \frac{\sqrt{N}}{n} \qquad n = 1, \ldots, \frac{N}{2} \tag{B.1}$$

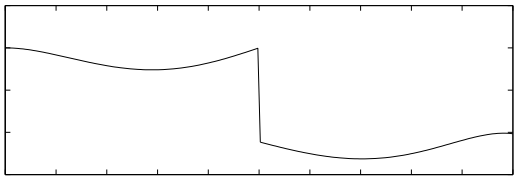


Figure B.1 : *The signal shown in the figure is piece-wise polynomial. The polynomial "piece" on the left of the discontinuity is of degree 4 and the polyomial "piece" on the right is of degree 8. The Fourier coefficients of such a signal decay very slowly due to the presence of the discontinuity in the signal. However, wavelets with 8 vanishing moments represent such a piece-wise polynomial signal very economically using just* $O(\log_2 N)$ *non-zero coefficients.*

where $n = 1 \ldots \frac{N}{2}$ indexes the discrete Fourier coefficients of the positive frequencies. The total energy $E_x s$ of the signal x is approximately given by

$$\begin{aligned} E_s &\approx N \sum_{n=1}^{\frac{N}{2}} \frac{1}{n^2} \\ &\approx N \left(\frac{\pi^2}{6} - \frac{2}{N} \right) \qquad \text{(for large } N\text{)} \\ &\approx N \frac{\pi^2}{6}. \end{aligned} \tag{B.2}$$

The variance of additive white Gaussian noise is σ^2 at each time sample, so the variance $\Gamma(f_n)$ at all discrete Fourier components is also σ^2. The total noise energy is E_γ is $N\,\sigma^2$. Note that the ratio of the average signal energy per sample and the average noise energy per sample are kept constant with increasing resolution N.

Consider any low-pass convolution operator $\mathcal{H}$ that decays polynomially in the Fourier domain (eg. Radon transform). The Fourier coefficients of such a convolution operator can be written as

$$|H(f_n)| \approx \frac{C}{n^\nu} \qquad n = 1 \ldots \tfrac{N}{2}, \text{for some } \nu > 0 \tag{B.3}$$

Since the constant C turns out to be inconsequential asymptotically, we choose to ignore it.

$$\frac{1}{|H(f_n)|} \approx n^\nu \qquad n = 1 \ldots \tfrac{N}{2} \tag{B.4}$$

The variance of noise colored by the inversion of $\mathcal{H}$ at each frequency f_n is given by $\sigma^2\, n^{2\,\nu}$, for $n = 1 \ldots \frac{N}{2}$. The best LTI estimate of the input signal is given by

$$\begin{aligned} \widetilde{X}_{LTI}(f_n) &= \frac{1}{H(f_n)} \frac{|H(f_n)|^2\, X(f_n)|^2}{|H(f_n)|^2\, |X(f_n)|^2 + \sigma^2} Y(f_n) \\ &= \frac{X(f_n)|^2}{|X(f_n)|^2 + \sigma^2\, n^{2\,\nu}} \left(X(f_n) + \frac{\Gamma(f_n)}{H(f_n)} \right) \end{aligned} \tag{B.5}$$

The average MSE of such an optimal LTI estimate is

$$\frac{1}{N}\,\mathcal{E}\left\{\sum_{n=1}^{N}|\widetilde{X}_{LTI}-X|^2\right\}=\frac{1}{N}\sum_{n=1}^{\frac{N}{2}}\left(\frac{4\,|X(f_n)|^2\,\sigma^2\,n^{2\,\nu}}{|X(f_n)|^2+\sigma^2\,n^{2\,\nu}}\right), \tag{B.6}$$

where $\mathcal{E}\{.\}$ denotes expectation with respect to the noise statistics. It is easily shown that that

$$2\,\min\left(|X(f_n)|^2,\sigma^2\,n^{2\,\nu}\right)\leq\left(\frac{4\,|X(f_n)|^2\,\sigma^2\,n^{2\,\nu}}{|X(f_n)|^2+\sigma^2\,n^{2\,\nu}}\right)\leq 4\,\min\left(|X(f_n)|^2,\sigma^2\,n^{2\,\nu}\right) \tag{B.7}$$

Hence,

$$\frac{2}{N}\sum_{n=1}^{N/2}\min\left(|X(f_n)|^2,\sigma^2\,n^{2\,\nu}\right)\leq\frac{1}{N}\,\mathcal{E}\left\{|\widetilde{X}_{LTI}-X|^2\right\}\leq\frac{4}{N}\sum_{n=1}^{N/2}\min\left(|X(f_n)|^2,\sigma^2\,n^{2\,\nu}\right) \tag{B.8}$$

The colored noise variance at frequency f_n increases monotonically with increasing n. In contrast, the signal energy decreases monotonically with n. Let f_q denote the DFT frequency where the signal coefficient energy is approximately equal to the colored noise variance. The index q is given by

$$\begin{aligned} q &\approx \left(N\,\sigma^2\right)^{\frac{1}{2\,\nu+2}},\, q\in\{1,\ldots,\frac{N}{2}\}. \\ |X(f_n)|^2 &< \sigma^2\,n^{2\,\nu} \qquad \text{for all } n>q \\ \text{and } |X(f_n)|^2 &> \sigma^2\,n^{2\,\nu} \qquad \text{for all } n<q \end{aligned} \tag{B.9}$$

$$\begin{aligned} \text{Hence, } \frac{1}{N}\sum_{n=1}^{N}\min\left(|X(f_n)|^2,\sigma^2\,n^{2\,\nu}\right) &= \frac{2}{N}\left(\sum_{n=1}^{q}\sigma^2\,n^{2\,\nu}\right)+\frac{2}{N}\sum_{n=q+1}^{\frac{N}{2}}|X(f_n)|^2 \\ &\geq \frac{2q}{N}\,\sigma^2+\frac{2}{N}\sum_{n=q+1}^{N/2}\frac{N}{n^2} \\ &\approx \frac{2q}{N}\,\sigma^2+2\left(\frac{\pi^2}{6}-\frac{2}{N}\right)-2\left(\frac{\pi^2}{6}-\frac{1}{q}\right) \\ &\approx \frac{\left(\sigma^2\right)^{\frac{2\,\nu+3}{2\,\nu+2}}}{N^{\frac{2\,\nu+1}{2\,\nu+2}}}+\left(\frac{2}{N\,\sigma^2}\right)^{\frac{1}{2\,\nu+2}}-\frac{4}{N} \\ &= \mathrm{O}\left(\frac{1}{N^{\left(\frac{1}{2\,\nu+2}\right)}}\right) \end{aligned} \tag{B.10}$$

Thus, the rate at which the average error per sample decays for the best LTI estimator is $\mathrm{O}\left(N^{\frac{-1}{2\nu+2}}\right)$.

Appendix C

Error Decay for Wavelet-based Techniques

Again, we consider a piece-wise polynomial function input signal (refer Figure B.1). Consider a wavelet system such that the number of zero moments of the wavelet basis functions is greater than or equal to the maximal degree of of the polynomial pieces in the input signal. Such a wavelet system represent piece-wise polynomial functions very economically. All the wavelet coefficients corresponding to the wavelet basis functions that are fully contained in the smooth polynomial region of the signal are zero [17]. The wavelet coefficients are non-zero only if supports of the corresponding wavelet basis function contains any point of discontinuity. Since there are $\log_2 N$ scales in wavelet decomposition, the number of non-zero wavelet coefficients is $O(\log_2 N)$. For the sake of simplicity, we assume that there are exactly $\log_2 N$ non-zero coefficients. The energy of the non-zero wavelet coefficients whose basis support contains the discontinuous event decay as 2^{-p} with increasing wavelet scales p [14]. So the wavelet coefficient at scale p is given by

$$|w_{p,k}|^2 \approx 2^{-p} |w_{0,k}|^2 \qquad \text{for } p = 1, \ldots . \log_2 N \tag{C.1}$$

The total signal energy is given by $\left(2 - \frac{1}{N}\right) |w_{0,k}|^2$. With denser sampling, the energy of the signal increases proportional to N. So the energy of the coarsest scale $|w_{0,k}|^2$ wavelet coefficient increases as $N\,C$, where C is some non-zero constant.

$$|w_{p,k}|^2 \approx 2^{-p} N\,C \qquad \text{for } p = 1, \ldots . \log_2 N \tag{C.2}$$

When the wavelets have sufficient zero moments, they are well-localized in fre-

quency and nearly band-limited. Then, the variance of the colored noise at wavelet scale p after inversion with H^{-1} is approximately given by

$$\begin{aligned}
\sigma_p^2 &= \sum_{n=1}^{N} \frac{|\hat{\psi}_p(f_n)|^2}{|H(f_n)|^2}\sigma^2 \\
&\approx \frac{1}{2^{p-1}} \sum_{n=2^{p-1}}^{2^p-1} n^{2\nu}\sigma^2 \\
&\approx \frac{\sigma^2}{2^{p-1}} \left[(2^p-1)^{2\nu+1} - \left(2^{p-1}\right)^{2\nu+1}\right] \\
&\approx \sigma^2\, 2^{2p\nu+1}
\end{aligned} \tag{C.3}$$

Note that the signal energy per sample and the noise energy per sample stays constant. We consider the error that results from employing ideal scale-dependent thresholding of the wavelet coefficients, where ideal thresholding is defined in 5.3. The average error per sample for an ideal thresholding estimator is given by $\frac{1}{N}\sum_{j,k}\min(|w_{j,k}|^2, \sigma_j^2)$. The colored noise variance monotonically increases with increasing scale p. In contrast, the signal energy decays with scale p. Let $\tilde{q}$ denote the scale where the signal wavelet coefficient energy is approximately equal to the colored noise variance. Then,

$$\tilde{q} \approx \log_2 \left(\frac{C\,N}{2\,\sigma^2}\right)^{\frac{1}{2\nu+1}} \tag{C.4}$$

$$\begin{aligned}
\frac{1}{N}\sum_{j,k}\min(|w_{j,k}|^2, \sigma_j^2) &= \frac{1}{N}\sum_{j=1}^{\log_2 N}\min(|w_{j,k}|^2, \sigma_j^2) \\
&= \frac{1}{N}\left(\sum_{j=1}^{\tilde{q}}\sigma_j^2 + \sum_{j=\tilde{q}+1}^{\log_2 N}|w_{j,k}|^2\right) \\
&= O\left(\frac{1}{N^{\left(\frac{1}{2\nu+1}\right)}}\right)
\end{aligned} \tag{C.5}$$

In practice, it is possible to estimate the signal within a $\log_2 N$ factor of the ideal thresholding error using wavelet-domain shrinkage [10]. So in practice, it is possible to achieve an error decay rate of $\mathrm{O}\left(\frac{\log_2 N}{N^{\left(\frac{1}{2\nu+1}\right)}}\right)$ using wavelet-based deconvolution

which is clearly faster that the rate achieved by LTI estimators asymptotically with increasing N.

Bibliography

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

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

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

[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] 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.

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

[10] 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.

[11] J. Kalifa, S. Mallat, and B. Rougé, "Image deconvolution in mirror wavelet bases," in *Proc. IEEE Int. Conf. Image Processing — ICIP '98*, (Chicago), pp. 565–569, Oct. 1998.

[12] 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.

[13] I. M. Johnstone and B. W. Silverman, "Discretization effects in statistical inverse problems," *J. Complexity*, no. 7, pp. 1–34, 1991.

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

[15] 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.

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

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

[18] 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.

[19] 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.

[20] 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.

[21] N.-Y. Lee and B. J. Lucier, "Inverting the Radon transform in the presence of noise," *In preparation.*

[22] 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.

[23] P. P. Vaidyanathan, *Multirate Systems and Filter Banks.* Englewood Cliffs, NJ: Prentice Hall, 1992.

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

[25] D. L. Donoho and I. M. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases," tech. rep., Stanford University, September 1994.

[26] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, May 1995.

[27] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[28] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Processing*, vol. 7, pp. 319–355, July 1998.

[29] 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.

[30] 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.

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

[32] H. Guo, "Denoising using redundant wavelet transforms," tech. rep., Rice Univ., Dept. Elec. and Comp. Eng., Dec. 1994.

[33] 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).

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