MULTIPLE BASIS WAVELET DENOISING USING BESOV PROJECTIONS

Hyeokho Choi and Richard Baraniuk

Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251–1892, USA

ABSTRACT

Wavelet-based image denoising algorithm depends upon the energy compaction property of wavelet transforms. However, for many real-world images, we cannot expect good energy compaction in a single wavelet domain, because most real-world images consist of components of a variety of smoothness. We can relieve this problem by using multiple wavelet bases to match different characteristics of images. In this paper, we propose a novel image denoising algorithm that uses multiple wavelet bases. By establishing a new relationship between the deterministic Besov space theory and the wavelet-domain statistical models, we generalize the Besov theory for finite sampled data. After defining convex sets in Besov spaces that contain the true image, we obtain an estimate of the true image by the method of projection onto convex sets. The algorithm outperforms existing multiple wavelet basis denoising algorithms; in particular, it shows excellent performance at low signal-to-noise ratios.

1. INTRODUCTION

Image restoration from noisy data is an important problem encountered in many image processing applications. Recently, nonlinear, non-parametric estimation methods based on linear, unitary transforms have figured prominently in denoising real-world signals and images. Wavelets are known to provide a sparse representation for a wide class of real-world signals, enabling simple threshold-based denoising methods.

When the signal is corrupted by additive white Gaussian noise, wavelet-domain thresholding provides excellent performance for noise removal. Donoho and Johnstone [1] showed that with a properly chosen threshold, wavelet thresholding is asymptotically optimal for removing noise from signals and images in Besov spaces [2].

Recently, there have been many attempts to improve the denoising performance at small sample sizes by using statistical inference methods based on wavelet statistical models and Bayesian estimation [3, 4]. It has been shown that optimal Bayesian estimation is equivalent to wavelet-domain thresholding under certain wavelet-domain statistical models [3]. These approaches have led to the many different thresholding schemes that outperform the SureShrink algorithm [1] by using different choices of thresholds and sometimes scale-adaptive or spatially-adaptive thresholding.

This work was supported by the National Science Foundation, grant no. MIP–9457438, DARPA, grant no. DARPA/AFOSR F49620–97–1–0513, ONR, grant no. N00014–99-1–0813, and Texas Instruments.

Email: choi@ece.rice.edu, richb@rice.edu

Web: www.dsp.rice.edu

Because wavelet-domain denoising algorithms rely on the sparse representation of signals in the wavelet domain, we must choose the wavelet basis to match the signal to be denoised. There have been several attempts to adaptively select the basis functions to match the given signal using either dynamic pruning of a wavelet packet tree to minimize the entropy [5, Sec. 9.3] or basis matching pursuit [5, Sec. 9.4]. However, the performance of these adaptive basis selection algorithms is limited, because the library of basis functions is usually not enough. When different portions of the signal have different properties, a library of basis functions with similar shapes is not sufficient for a compact representation.

For improved denoising performance, the use of multiple wavelet bases is desirable because each portion of the signal can be represented compactly in the wavelet domain that matches its properties. Recently, there have been several attempts to use more than one wavelet basis for signal denoising. In [6], a wavelet-domain empirical Wiener filter was used for denoising, and the filter was designed using a signal estimate in second wavelet domain. It was later shown that the two wavelet bases assist each other to alleviate the over-smoothing effect of the usual wavelet-domain thresholding [7]. In [8], convex projections onto the confidence tubes defined around the given noisy signal were used to estimate the signal. Using multiple confidence tubes defined in multiple wavelet domains, the algorithm demonstrated a good denoising performance.

Although the use of multiple wavelet bases for signal denoising is intuitively appealing, it is not clear how to utilize the multiple bases in a structured fashion. The set of multiple orthogonal wavelet basis functions is a tight frame [5], and the usual frame inversion method yields a signal estimate that is the average of signal estimates in each domain. However, if one or more of the wavelet domains does not represent the signal compactly, then simple averaging does not necessarily yield an improved result. Furthermore, existing algorithms using multiple domains are very difficult to analyze, and they are not guaranteed to provide improved performance.

Besov spaces have emerged as promising new tools for wavelet-domain signal processing. Because many images encountered in the real world ("photograph-like" images) consist of textured regions separated by edges, they possess particular smoothness characteristics. Besov spaces characterize these smoothness properties. Besov spaces have been utilized to assess the performance of signal estimation [1] or compression algorithms [2]. Wavelets provide unconditional bases for the Besov spaces, and many real-world signals are known to belong to Besov spaces with certain parameters

[2]. Recently, a direct connection between wavelet thresholding and Besov spaces was revealed through the formulation of signal estimation as a variational regularization problem in Besov space [9]. The success of this method indicates a close relationship between wavelet shrinkage algorithms [1] and Besov spaces.

In this paper, we propose a novel signal denoising scheme that uses multiple wavelet bases. By defining convex sets in a Besov space based on multiple wavelet bases, we define a set of functions that contains the true signal. By projecting the noisy signal onto this set, we obtain an excellent signal estimate. We call the convex sets "Besov balls" because of the close relationship between the convex sets and Besov spaces. We define multiple convex sets using multiple wavelet bases and obtain the signal estimate by alternately projecting onto the Besov balls. The new algorithm outperforms existing methods in mean-squared-error (MSE) and in particular shows excellent performance when the signal-to-noise ratio (SNR) is low.

2. WAVELETS AND BESOV SPACE

The discrete wavelet transform (DWT) represents a one-dimensional (1-d) signal $z(t)$ in terms of shifted versions of a lowpass scaling function $\phi(t)$ and shifted and dilated versions of a prototype bandpass wavelet function $\psi(t)$ [5]. For special choices of $\phi(t)$ and $\psi(t)$, the functions $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ and $\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k)$, with $j, k \in \mathbb{Z}$ form an orthonormal basis, and we have the representation [5]

$$z = \sum_{j=j_0}^{\infty} \sum_{k} w_{j,k}\, \psi_{j,k} + \sum_{k} u_{j_0,k}\, \phi_{j_0,k}, \qquad (1)$$

with $u_{j,k} \equiv \int z(t)\, \phi^*_{j,k}(t)\, dt$ and $w_{j,k} \equiv \int z(t)\, \psi^*_{j,k}(t)\, dt$.

The *wavelet coefficient* $w_{j,k}$ measures the signal content around time $2^{-j}k$ and frequency $2^j f_0$. The scaling coefficient $u_{j,k}$ measures the local mean around time $2^{-j}k$. The DWT (1) employs scaling coefficients only at scale j_0; wavelet coefficients at scales $j > j_0$ add higher resolution details to the signal.

We can easily construct two-dimensional (2-d) wavelets from the 1-d ψ and ϕ by setting for $\mathbf{x} \equiv (x, y) \in \mathbb{R}^2$, $\psi^{\mathrm{HL}}(x, y) = \psi(x)\phi(y), \psi^{\mathrm{LH}}(x, y) = \phi(x)\psi(y), \psi^{\mathrm{HH}}(x, y) = \psi(x)\psi(y)$, and $\phi(x, y) = \phi(x)\phi(y)$. If we let $\Psi \equiv \{\psi^{\mathrm{HL}}, \psi^{\mathrm{LH}}, \psi^{\mathrm{HH}}\}$, then the set of functions $\{\psi_{j,\mathbf{k}} \equiv 2^j\psi(2^j\mathbf{x} - \mathbf{k})\}_{\psi\in\Psi, j\in\mathbf{Z}, \mathbf{k}\in\mathbf{Z}^2}$ and $\{\phi_{j,\mathbf{k}} \equiv 2^j\phi(2^j\mathbf{x} - \mathbf{k})\}_{j\in\mathbf{Z}, \mathbf{k}\in\mathbf{Z}^2}$ forms an orthonormal basis for $L_2(\mathbb{R}^2)$. That is, for every $z \in L_2(\mathbb{R}^2)$, we have

$$z = \sum_{j>j_0, \mathbf{k}\in\mathbf{Z}^2, \psi\in\Psi} w_{j,\mathbf{k},\psi}\, \psi_{j,\mathbf{k}} + \sum_{\mathbf{k}\in\mathbf{Z}^2} u_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}, \qquad (2)$$

with $w_{j,\mathbf{k},\psi} \equiv \int z(\mathbf{x})\psi_{j,\mathbf{k}}(\mathbf{x})d\mathbf{x}$ and $u_{j_0,\mathbf{k}} \equiv \int z(\mathbf{x})\phi_{j_0,\mathbf{k}}(\mathbf{x})d\mathbf{x}$.

For the discrete processing of images, the original image is given as the discrete samples of the continuous image $z(\mathbf{x})$. With proper prefiltering, we can approximate the discrete samples as the scaling coefficients of $z(\mathbf{x})$ at a certain scale J, that is, the sampled image $z(\mathbf{k}) = u_{J,\mathbf{k}}$. Equivalently, we can build a continuous-time image corresponding to $z(\mathbf{k})$ as

$$\widetilde{z} = \sum_{k\in\mathbf{Z}^2} u_{J,\mathbf{k}}\phi_{J,\mathbf{k}}, \qquad (3)$$

or, using wavelet coefficients

$$\widetilde{z} = \sum_{j_0<j<J, \mathbf{k}\in\mathbf{Z}^2, \psi\in\Psi} w_{j,\mathbf{k},\psi}\psi_{j,\mathbf{k}} + \sum_{k\in\mathbf{Z}^2} u_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}, \qquad (4)$$

where the coefficients $w_{j,\mathbf{k},\psi}$ and $u_{j_0,\mathbf{k}}$ are easily computed using the 2-d discrete-time wavelet filters and decimators operating on the samples $z(\mathbf{k})$ [5].

Wavelets provide a simple characterization for a wide variety of function *smoothness spaces* [2]. The norms of these spaces measure signal smoothness: smaller norms imply smoother functions. The scale of *Besov spaces* $B^\alpha_q(L_p)$, $0 < \alpha < \infty$, $0 < p \le \infty$, $0 < q \le \infty$, are particularly useful, for they contain many life-like signals. For analyzing ϕ and ψ possessing $r > \alpha$ vanishing moments [5], the Besov norm $\|z\|_{B^\alpha_q(L_p)}$ can be defined as a sequence norm on the wavelet coefficients of z [2]

$$\|z\|_{B^\alpha_q(L_p)} \equiv \|u_{j_0,\mathbf{k}}\|_p + \left(\sum_{j\ge j_0}\left(\sum_{\mathbf{k},\psi\in\Psi} 2^{j(\alpha p+p-2)}|w_{j,\mathbf{k},\psi}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}. \qquad (5)$$

The three hyperparameters have natural interpretations: a p-norm of the wavelet coefficients is taken within each scale j, a weighted q-norm is taken across scale, and the smoothness parameter α controls the rate of decay of the $w_{j,\mathbf{k},\psi}$ across scale (frequency). Roughly speaking, the parameter α corresponds to the number of well-behaved derivatives of z in L_p; hence, the larger the α, the smoother the functions in $B^\alpha_q(L_p)$ [2]. Do not be misled by the terminology "smoothness space" — for $\alpha < 1$, Besov spaces contain discontinuous functions.

In the applications of signal and image processing, we have two particular cases of interest. The first case is when $p = q$. Then the Besov norm reduces to

$$\|z\|_{B^\alpha_p(L_p)} \equiv \|u_{j_0,\mathbf{k}}\|_p + \left(\sum_{j\ge j_0,\mathbf{k},\psi\in\Psi} 2^{j(\alpha p+p-2)}|w_{j,\mathbf{k},\psi}|^p\right)^{\frac{1}{p}}, \qquad (6)$$

which is a simple weighted p-norm of the wavelet coefficients. A simple but useful set of the Besov spaces are the *Sobolev spaces*, obtained as $W^\alpha(L_2) \equiv B^\alpha_2(L_2)$, which measure smoothness of order α in L_2. Another Besov space of interest to image processing is $B^1_\infty(L_1)$, because $B^1_1(L_1) \subset \mathrm{BV} \subset B^1_\infty(L_1)$ and typical real-world images belong to BV [9]. The corresponding norm is

$$\|z\|_{B^1_\infty(L_1)} \equiv \|u_{j_0,\mathbf{k}}\|_p + \sup_{j\ge j_0}\sum_{\mathbf{k},\psi\in\Psi}|w_{j,\mathbf{k},\psi}|. \qquad (7)$$

3. INDEPENDENT GENERALIZED GAUSSIAN WAVELET-DOMAIN MODELS

Besov spaces provide a *deterministic* model of real-world images, in contrast to a *statistical* model [3, 4]. However, these ideas are intimately intertwined, as we now show.

After reviewing a simple statistical wavelet-domain image model in this section, we make connections with Besov spaces in Section 4.

The simplest wavelet statistical models are obtained by assuming that the wavelet coefficients are independent of each other. This assumption is motivated by the approximate decorrelation property of wavelet transforms [5]. Under the independence assumption, modeling reduces to specifying the marginal distribution of each wavelet coefficients.

The popular Gaussian distribution is not appropriate for modeling the peaky, heavy-tailed marginal distribution of wavelet coefficients. After Mallat et al. [10] discovered that the statistics of wavelet coefficients of real-world images are well approximated with the generalized Gaussian distribution (GGD), there have been attempts to model the wavelet coefficients statistically using the GGD priors in the context of Bayesian inference [3, 4].

The zero-mean GGD function $\mathrm{GGD}_\nu(0,\sigma^2)$ with shape parameter ν and variance σ^2 is defined as

$$f(x) = \frac{\nu\eta(\nu)}{2\Gamma(1/\nu)}\frac{1}{\sigma}\exp\left\{-[\eta(\nu)|x|/\sigma]^\nu\right\}, \tag{8}$$

with $\eta(\nu) = \sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}}$. The GGD model includes the Gaussian and Laplacian distribution as special cases, using $\nu = 2$ and $\nu = 1$, respectively.

Under an independent GGD wavelet model, we posit that each wavelet coefficient is generated independently according to a zero-mean GGD. For the tractability of the model, all wavelet coefficients at each scale are assumed to be independent and identically distributed (iid-in-scale). That is, under the independent GGD model, we have $w_{j,\mathbf{k},\psi} \overset{\text{iid}}{\sim} \mathrm{GGD}_{\nu_j}(0,\sigma_j^2)$. Due to the iid-in-scale assumption, the shape parameter ν_j and the variance σ_j^2 do not depend on the spatial location $\mathbf{k}$.

The shape parameter ν_j of the GGD function at scale j characterizes the peakiness and the heavy tail of the actual distribution of the wavelet coefficients. We may specify the shape parameters without reference to the given data, because most real-world images tend to have similar energy compaction in the wavelet domain [3]. For practical applications, further simplification of the model comes from assuming that the shape parameter ν_j is same over all scales.

The variance σ_j^2 represents the signal energy at scale j. It can be empirically estimated based on the given data by estimating the variance of wavelet coefficients at each scale [3], or it can be specified to decay exponentially based on the $1/f$ spectral behavior of real-world images.

In summary, an independent GGD wavelet model $\mathbf{\Theta}$ with exponentially decaying variance ($1/f$ model) takes the form

$$\mathbf{\Theta} : w_{j,\mathbf{k},\psi} \overset{\text{iid}}{\sim} \mathrm{GGD}_\nu(0,\sigma_j^2) \text{ with } \sigma_j = 2^{-j\beta}\sigma_0. \tag{9}$$

Because the decay of wavelet coefficients across scale determines the smoothness of the underlying signal, the realizations of the independent GGD model in (9) with exponentially decaying variance can be shown to belong to certain scales of Besov space [11].

4. STATISTICAL INTERPRETATION OF BESOV SPACES

Although the GGD model in (9) is seemingly unrelated to the theory of Besov spaces, there exists an interpretation of the Besov norm in terms of the likelihood function under the model in (9). This relation enables us to generalize the Besov space theory to the set of finite sampled images, yielding the notion of the "Besov ball" that is applicable to practical image processing problems.

4.1. Normalized likelihood function

Given the pdf $f(x|\mathbf{\Theta})$ of a random variable $\mathbf{x}$ under a model $\mathbf{\Theta}$ and realizations x_1 and x_2, we can compute the likelihoods of the data $f(x_1|\mathbf{\Theta})$ and $f(x_2|\mathbf{\Theta})$. The likelihoods indicate how "likely" the given data are under the given pdf; for instance, $f(x_1|\mathbf{\Theta}) > f(x_2|\mathbf{\Theta})$ means that x_1 is more likely than x_2 under the model.

However, the likelihood $f(x_1|\mathbf{\Theta})$ is meaningful only when it is compared with another likelihood such as $f(x_2|\mathbf{\Theta})$, and we cannot quantify how likely x_1 is merely based on the value $f(x_1|\mathbf{\Theta})$. To be able to tell how likely an observation is under a given pdf model, we need to normalize the likelihood appropriately. A natural way of accomplishing the normalization is to compare the likelihood with the maximum likelihood achievable. We define the *normalized likelihood function* $f^N(x|\mathbf{\Theta})$ by

$$f^N(x|\mathbf{\Theta}) = \frac{f(x|\mathbf{\Theta})}{\sup_x f(x|\mathbf{\Theta})}, \tag{10}$$

with the assumption that $0 < \sup_x f(x|\mathbf{\Theta}) < \infty$. Then, $f^N(x|\mathbf{\Theta}) \in [0,1]$, and we can say that an observation x is "likely" if $f^N(x|\mathbf{\Theta})$ is close to 1 and not likely if it is close to 0. We can easily generalize the concept of the normalized likelihood to finite random vectors using their joint pdfs.

For discrete random processes, the joint pdf of an infinite random vector is not defined, and we cannot define the normalized likelihood as in (10). However, we can generalize the normalized likelihood using the limit of the normalized likelihood when the limit exists.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be an infinite sequence of random variables. Then for the vector of the first n random variables $\boldsymbol{x}_n = \{\mathbf{x}_k\}_{k=1}^n$, we can define the normalized likelihood $f_n^N(\boldsymbol{x}_n)$ using (10). Then, if the limit $\lim_{n\to\infty} f_n^N$ exists, we define the normalized log likelihood of the infinite sequence to be the limit.

For the independent wavelet-domain model considered in Section 3, we can take the limit as we move to the fine scales and define the normalized likelihood as

$$f^N(\boldsymbol{w}) = \lim_{J\to\infty} \frac{\prod_{j=0}^J \prod_{\mathbf{k},\psi} f_{j,\mathbf{k},\psi}(w_{j,\mathbf{k},\psi})}{\sup \prod_{j=0}^J \prod_{\mathbf{k},\psi} f_{j,\mathbf{k},\psi}(w_{j,\mathbf{k},\psi})}, \tag{11}$$

when $w_{j,\mathbf{k},\psi} \sim f_{j,\mathbf{k},\psi}(w_{j,\mathbf{k},\psi})$. For the independent GGD model with exponentially decaying variance defined in (9), the normalized likelihood is well defined because the supremum is finite and the limit exists.

4.2. Besov spaces and the GGD model

Consider an independent GGD model $\mathbf{\Theta}_p^\alpha$ for the wavelet coefficients with $w_{j,\mathbf{k},\psi}$ modeled as a zero-mean GGD model

as in (9) with $\nu = p$ and $\beta = s + 2 - 2/p$. In this case, it is easy to see that[1]

$$-\log f^N(\boldsymbol{w}|\boldsymbol{\Theta}_p^\alpha) = C\|z\|^p_{B_p^\alpha(L_p)}, \tag{12}$$

the homogeneous Besov norm of the image z. When $p = 2$, we obtain an iid Gaussian model for the wavelet coefficients and the corresponding normalized negative log likelihood is equivalent to the Sobolev norm of the function.

In terms of normalized likelihood function for the iid GGD model $\boldsymbol{\Theta}_p^\alpha$, the Besov space $B_p^\alpha(L_p)$ can be equivalently defined as the set $\{\boldsymbol{w} : f^N(\boldsymbol{w}|\boldsymbol{\Theta}_p^\alpha) \neq 0\}$. Thus, the signals in the Besov space $B_p^\alpha(L_p)$ are the "likely" signals under the statistical model $\boldsymbol{\Theta}_p^\alpha$.

The characterization of Besov space in terms of the likelihood function can be generalized to other scales of Besov space (with $p \neq q$) that are of interest in image processing. For $B_\infty^1(L_1)$, for instance, we are interested in the scale-by-scale normalized likelihood under the GGD model with $\nu = 1$ and $\beta = 0$. It is easily shown that $B_\infty^1(L_1) = \{\boldsymbol{w} : \inf_j g_j^N(\boldsymbol{w}) > 0\}$, where $g_j^N(\boldsymbol{w}) = \prod_{\mathbf{k},\psi} \exp(|w_{j,\mathbf{k},\psi}|/\sigma_0)$ is the normalized log likelihood of the wavelet coefficients at scale j. That is, rather than the joint likelihood of all the wavelet coefficients, $B_\infty^1(L_1)$ is concerned with the scale of least likelihood.

4.3. Besov balls for finite data

In practice, the available data are a finite number of samples of the image, and we cannot have wavelet coefficients beyond a certain fine scale. Suppose we have wavelet coefficients up to scale J, and that the coefficients with scale $j > J$ are unknown. Let $\boldsymbol{w}_J$ be the vector of available wavelet coefficients. Then, the definition of Besov space must be modified appropriately. The set of "likely" signals can be modified as $\{\boldsymbol{w}_J : f_J^N(\boldsymbol{w}_J|\boldsymbol{\Theta}_p^\alpha) \geq \epsilon\}$, where $f_J^N(\boldsymbol{w}_J|\boldsymbol{\Theta}_p^\alpha)$ is the normalized likelihood of the (finite) vector $\boldsymbol{w}_J$ and ϵ is a positive constant controlling the "likelihood" of signals. We can choose ϵ so that the set contains the images of interest. This is a generalization of the Besov space for finite data, which has not been investigated to date.

In general, the definition of the Besov spaces is concerned with the asymptotic decay of the signal energy across scales, and it is impossible to apply the same theory to finite data (all finite truncated wavelet transforms always have finite Besov norms). The above generalization of the Besov space for finite data is thus interesting in its own right.

Under the independent GGD model, the negative log likelihood function is the truncated form of the Besov norm defined in (6), and the set of signals is a "ball" in Besov space defined as $\{x : \|x\|_B \leq R\}$, where R is the radius of the ball and $\|\cdot\|_B$ is the truncated Besov norm for finite samples. Although we considered the homogeneous Besov spaces $B_p^\alpha(L_p)$ above, the Besov balls can be defined for other scales of Besov spaces in the same way. If $p, q \geq 1$, Besov balls are convex sets in L_p.

5. BESOV BALL DENOISING ALGORITHM

The notion of a Besov ball for finite sampled data leads us to an intuitively appealing signal denoising algorithm. Given an image z, we model the noisy observation as $z + n$, with n an additive Gaussian white noise with variance σ^2. The estimation problem is to restore the original image z from the noisy observation $z + n$. We can project the noisy observation $z + n$ onto a Besov ball defined as above to obtain an estimate of the original signal that is close to the given observation and is "likely" in terms of the underlying wavelet statistical model.

Suppose that we have M different wavelet bases, $W_1, \ldots, W_M$, and that the number of vanishing moments of each wavelet basis is greater than the Besov smoothness parameter α of the Besov space under consideration. For each wavelet basis W_i, we define the Besov ball $B_i(r_i)$ with radius r_i as $B_i(r_i) = \{z : \|z\|_B \leq r_i\}$ where the norm is the truncated Besov norm defined using wavelet basis W_i. With properly chosen r_i, the original signal z is in $B_i(r_i)$ for all i, and hence we have $z \in \cap_i B_i(r_i)$. To obtain an estimate of z, we project the noisy signal $z+n$ onto $\cap_i B_i(r_i)$ in the l_2 sense. We can find the projection using the method of projection onto convex sets (POCS) [13] starting from $z + n$. For an $N \times N$ image, each Besov ball $B_i(r_i)$ is a convex set in $\mathbb{R}^{N^2}$.

As we change the radius r_i, the size of the Besov ball changes. In $\mathbb{R}^{N^2}$, the change of basis from one orthonormal wavelet domain to another orthonormal wavelet domain corresponds to a rotation of the coordinate axes, and hence each Besov ball $B_i(r_i)$ is a rotation of other Besov balls defined using other wavelet transforms. Multiple wavelet bases define Besov balls with different rotations and possibly with different sizes, but all contain the original image z. As we increase the number of wavelet domains, the size of the intersection $\cap_i B_i(r_i)$ never increases, and it defines an ever smaller set of feasible signals. By projecting the noisy signal onto this small convex set, we can improve over a single-basis denoising algorithm.

The proper choice of the radii of the Besov balls is essential for the success of the algorithm. This is analogous to choosing the threshold in wavelet thresholding. We should choose each r_i so that the Besov ball $B_i(r_i)$ contains the true noise-free image z. At the same time, r_i should be chosen as small as possible for accurate denoising. Ideally, r_i should be chosen as $r_i = \|z\|_B$, where the norm is the truncated Besov norm defined using basis W_i. However, because z is unknown, we must estimate $\|z\|_B$.

One method of estimating the Besov norm of z is to compute the Besov norm of an estimate of z obtained using some other denoising method. A natural choice is the Besov norm of the denoised signal using Donoho and Johnstone's universal thresholding [1]. However, because the signal is over-smoothed by this universal thresholding, we need to choose somewhat larger r_i than the Besov norm of the thresholded signal. For real images, we found out that 1.5 to 2 times the Besov norm of the SureShrink signal estimate provides a good estimate of the actual Besov norm.

The choice of underlying Besov space $B_q^\alpha(L_p)$ depends on the properties of the data to be denoised. For real world images, we found that $B_\infty^1(L_1)$ yields good performance as well as simple numerical implementation.

Figure 1 compares the PSNR ($10\log_{10}(\frac{255^2}{\text{MSE}})$) of various denoising methods for a real image. As the wavelet bases,

[1] It can be shown that the realizations of this GGD model belong to the Besov space almost surely [11, 12].

Figure 1: Image denoising using projections onto Besov balls. (a) 512×512 original image with 256 gray levels. (b) Noisy image ($\sigma = 35$, PSNR = 17.2dB). (c) MATLAB's spatially adaptive Wiener filter (3×3 window, PSNR = 27.0dB). (d) Hard thresholding in D_{12} domain using threshold chosen to maximize PSNR (PSNR = 26.1dB), (e) Confidence tube denoising algorithm of Ishwar et al. [8] using D_6, D_8, D_{10}, D_{12} and D_{14} (PSNR = 27.5dB). (f) Besov ball projection in $B^1_\infty(L_1)$ using the same 5 wavelet bases (PSNR = 28.3dB).

we used Daubechies orthonormal wavelet filters. Here, D_n refers to the Daubechies length-n filter. We observe that the proposed denoising algorithm using 5 different wavelet bases outperforms other existing multiple-basis denoising methods both visually and in MSE. In particular, the Besov ball algorithm outperforms the recently proposed confidence tube algorithm [8] by almost 1dB. In the denoising simulations involving many different images and different noise variances, the Besov ball algorithm demonstrated excellent performance, especially when the SNR was low.

6. CONCLUSIONS

In this paper, we have proposed a novel "Besov ball" denoising algorithm that uses multiple wavelet bases. The performance of the algorithm is superior to other state-of-the-art multiple-basis denoising algorithms.

The basis of the proposed denoising algorithm is the relationship between the Besov norm and the normalized likelihood under a wavelet-domain statistical model. This relationship generalizes the Besov space theory for practical image processing problems.

Currently we are working to generalize the concept of Besov balls to other, more accurate statistical image models, obtaining a set or space containing real-world images with a measure telling how each of the elements resembles a real, photograph-like image [11].

7. REFERENCES

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

[2] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 719–746, March 1992.

[3] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," in *Proc. IEEE TFTS Symposium*, Pittsburgh, PA, Oct. 6-9 1998, pp. 633–636.

[4] E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *Proc. 31st Asilomar Conference*, Pacific Grove, CA, Nov. 1997, pp. 673–678.

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

[6] S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empirical Wiener filtering," in *Proceedings of SPIE*, San Diego, July 1997, vol. 3169, pp. 389–399.

[7] H. Choi and R. Baraniuk, "Analysis of wavelet-domain Wiener filters," in *Proc. IEEE TFTS Symposium*, Pittsburgh, PA, Oct. 6-9 1998, pp. 613–616.

[8] P. Ishwar, K. Ratakonda, P. Moulin, and N. Ahuja, "Image denoising using multiple compaction domains," in *Proc. ICASSP 98*, Seattle, WA, May 1998, pp. 1889–1892.

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

[10] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.

[11] H. Choi and R. Baraniuk, "Wavelet-domain statistical models and Besov spaces," in *Proc. of SPIE Technical conference on Wavelet Applications in Signal Processing VII*, Denver, July 1999.

[12] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Roy Statist. Soc. Ser. B*, vol. 60, pp. 725–749, 1998.

[13] D. C. Youla and H. Webb, "Image restoration by method of convex projections: Part 1—Theory," *IEEE Trans. Medical Imaging*, vol. MI-1, no. 2, pp. 81–94, October 1982.