

Signal Restoration with Overcomplete Wavelet Transforms: Comparison of Analysis and Synthesis Priors

Ivan W. Selesnick^a and Mário A. T. Figueiredo^b

^aPolytechnic Institute of New York University, Brooklyn, NY 11201, USA

^bInstituto de Telecomunicações, Instituto Superior Técnico, 1049-001 Lisboa, Portugal

ABSTRACT

The variational approach to signal restoration calls for the minimization of a cost function that is the sum of a data fidelity term and a regularization term, the latter term constituting a ‘prior’. A synthesis prior represents the sought signal as a weighted sum of ‘atoms’. On the other hand, an analysis prior models the coefficients obtained by applying the forward transform to the signal. For orthonormal transforms, the synthesis prior and analysis prior are equivalent; however, for overcomplete transforms the two formulations are different. We compare analysis and synthesis ℓ_1 -norm regularization with overcomplete transforms for denoising and deconvolution.

Keywords: signal restoration, deconvolution, denoising, wavelets, sparsity, total variation

1. INTRODUCTION

The restoration, or reconstruction, of a signal from noisy measurements arises in many applications. We assume the classical observation model of the form

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

where \mathbf{H} represents a known linear operator, \mathbf{x} is the signal to be restored (estimated) and \mathbf{n} is additive white Gaussian noise. The operator \mathbf{H} can represent blurring, sub-sampling, masking, a Fourier transform (as in the case of magnetic resonance imaging), or any combination of these and other linear operators. If \mathbf{H} is the identity operator, then the problem is one of signal denoising. If \mathbf{H} represents convolution, then the problem is one of deconvolution. We do not assume that \mathbf{y} is of the same size as \mathbf{x} (the operator \mathbf{H} need not be square, it may be either a ‘tall’ or ‘wide’ matrix).

Arguably the most often used approach for restoring the signal \mathbf{x} is based on minimizing an objective function of the form

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda R(\mathbf{x}). \quad (2)$$

The first term, $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$, is known as the ‘data fidelity’ term because it measures the consistency between the estimated signal \mathbf{x} and the measured data \mathbf{y} ; its form results from the white Gaussian nature of the noise in (1). The second term, $R(\mathbf{x})$, is the ‘regularizer’ or ‘penalty’, and corresponds to the negative log prior from a Bayesian estimation perspective. The regularizer is chosen so as to reflect prior knowledge about \mathbf{x} and is designed to yield large values when the behaviour of \mathbf{x} is very different from its expected behaviour. For example, if it is known that \mathbf{x} should be a smooth (low-frequency) signal then $R(\mathbf{x})$ may be chosen to be the energy of the output of a high-pass filter. If it is known that \mathbf{x} should be a piecewise-constant signal, then is appropriate to set $R(\mathbf{x})$ to be the *total variation*⁴⁷ of \mathbf{x} . The parameter λ is the regularization parameter, which is used to control the degree of regularization.

Much recent research on the restoration of natural signals has shown that regularization based on sparsity can be very effective. This result relies on the sparse representation in an appropriate domain of the signal being restored. For example, the wavelet transform and related transforms provide moderately sparse representations for many signals and images. To exploit wavelet-domain sparsity, the regularizer, $R(\mathbf{x})$, should be a measure of

Author Email: selesi@poly.edu, mario.figueiredo@lx.it.pt

the sparsity of the wavelet coefficients of \mathbf{x} . For example, $R(\mathbf{x})$ may be chosen to be the number of non-zero wavelet coefficients (the ℓ_0 -norm). However, the ℓ_0 -norm is not convex and therefore the objective function $J(\mathbf{x})$ can be difficult to minimize in general. As an alternative, the ℓ_1 -norm is often used instead of the ℓ_0 -norm because it is convex and yet promotes sparsity. (But see Ref. 44 for algorithms and extensive examples on real images showing the superiority of the ℓ_0 -norm in comparison with the ℓ_1 -norm).

In performing wavelet-based signal/image restoration^{39,41,57} by minimizing an objective function of the form (2), several questions arise: the choice of the transform, the way sparsity is measured, the selection of the regularization parameter λ , and the availability of algorithms minimizing the objective function $J(\mathbf{x})$. In this paper, we exclusively use the ℓ_1 -norm, and we do not address the selection of λ .

The main goal of this paper is to compare two problem formulations for using wavelet-regularization signal restoration. Given a specific wavelet transform to be used for the regularization, we consider the ‘analysis-prior’ and the ‘synthesis-prior’ formulations of the problem. For some signal restoration problems, the two formulations give quite different results.

The *analysis prior* formulation uses an objective function of the form:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{A}\mathbf{x}\|_1. \quad (3)$$

Here, the regularization term is the ℓ_1 -norm of $\mathbf{A}\mathbf{x}$. The symbol \mathbf{A} represents the transform, and $\mathbf{A}\mathbf{x}$ is the transform of \mathbf{x} .

On the other hand, the *synthesis prior* formulation uses an objective function of the form:

$$J(\mathbf{w}) = \|\mathbf{y} - \mathbf{H}\mathbf{S}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1. \quad (4)$$

Again, the goal is to minimize $J(\mathbf{w})$, but now the restored signal, \mathbf{x} , is given by $\mathbf{S}\mathbf{w}$. The symbol \mathbf{S} represents the ‘inverse’ transform, and \mathbf{w} represents the transform coefficients (e.g. wavelet coefficients). In this formulation, the signal to be restored, \mathbf{x} , is modeled as a synthesis of atoms.

A basic, known, result is that if \mathbf{A} is an orthonormal transform ($\mathbf{A}\mathbf{A}^t = \mathbf{A}^t\mathbf{A} = \mathbf{I}$) then, with $\mathbf{S} = \mathbf{A}^t$, the analysis and synthesis priors are equivalent. The same is true more generally if \mathbf{A} is invertible. Their minimization gives the same restored signal \mathbf{x} . However, otherwise, the two formulations are different. Note: in the orthonormal denoising case ($\mathbf{H} = \mathbf{I}$ and $\mathbf{A}\mathbf{A}^t = \mathbf{A}^t\mathbf{A} = \mathbf{I}$), the solution is given directly by soft-thresholding in the transform domain.

In this paper, we focus on the use of overcomplete wavelet transforms (an expansive transform maps an N -point signal \mathbf{x} to M wavelet coefficients with $M > N$). Then, the matrix \mathbf{A} is tall, and the matrix $\mathbf{S} = \mathbf{A}^t$ is wide. For these wavelet transforms, the analysis prior and synthesis prior formulations give different results. The analysis formulation (3) and synthesis formulation (4) for signal restoration has been discussed by several authors, see Refs. 8,9,17,27,44. A focus of this paper is the way the low-pass coefficients of a wavelet transform are dealt with. Our experiments are based on the exclusion of the low-pass coefficients from the regularization term because, as we discuss below, the low-pass coefficients of many signals and images are often not sparse. An important topic is the development and implementation of efficient algorithms for minimizing the ℓ_1 regularized objective functions (3) and (4) and we refer the reader to papers, eg Refs. 1, 8,9,16,32.

2. ALGORITHMS

2.1 Analysis Prior

An algorithm for minimizing the analysis-prior objective function, (3), is given by the Chambolle-type algorithm:^{3,13}

$$\mathbf{b}^{(i)} = \mathbf{x}^{(i)} + \frac{1}{\alpha} \mathbf{H}^t (\mathbf{y} - \mathbf{H}\mathbf{x}^{(i)}) \quad (5)$$

$$\mathbf{z}^{(i+1)} = \left(c \mathbf{z}^{(i)} + \mathbf{A}(\mathbf{b}^{(i)} - \mathbf{A}^t \mathbf{z}^{(i)}) \right) ./ \left(\frac{2\alpha}{\lambda} |\mathbf{A}\mathbf{x}^{(i)}| + c \right) \quad (6)$$

$$\mathbf{x}^{(i+1)} = \mathbf{b}^{(i)} - \mathbf{A}^t \mathbf{z}^{(i+1)} \quad (7)$$

where the superscript i is the iteration index. The operations $./$ and $|\cdot|$ in (6) are element-wise division and element-wise absolute value. An informal derivation of the algorithm (5-7) is given in Appendix A. The derivation is based on Ref. 29, however, in Appendix A several steps of the conjugate gradient algorithm are replaced by a single update step. The algorithm (5-7) is a 1D version of Chambolle’s algorithm.^{3,13} To ensure convergence, we can set $\alpha > \text{maxeig}(\mathbf{H}^t\mathbf{H})$ and $c > \text{maxeig}(\mathbf{A}\mathbf{A}^t)$.

Because the objective function is convex, the initialization of \mathbf{x} and \mathbf{z} is not very important. The numerical experiments in this paper use $\mathbf{z}^{(0)} = \mathbf{0}$ and $\mathbf{x}^{(0)} = \mathbf{H}^t \mathbf{y}$.

Algorithm (5-7) is simple to implement; it requires only procedures to apply \mathbf{A} , \mathbf{A}^t , \mathbf{H} , and \mathbf{H}^t . No individual rows, columns, or elements of \mathbf{A} or \mathbf{H} are needed. However, the algorithm can converge very slowly. The development of fast algorithms to minimize the objective function (3), and related ones, is an active area of research and more sophisticated algorithms have been developed with faster convergence. In particular, some algorithms developed for total-variation regularized inverse problems can be adapted for the minimization of (3); other algorithms are developed specifically for the analysis prior formulation, see Refs. 1, 9, 32, 42, 46, 54, 58. Several works have specifically considered analysis (and synthesis) priors with overcomplete wavelet transforms, see Refs. 8–10, 12, 35, 38

2.2 Synthesis Prior

An algorithm for minimizing the synthesis-prior objective function (4), is given by iterative thresholding:

$$\mathbf{w}^{(i+1)} = \text{soft}\left(\mathbf{w}^{(i)} + \frac{1}{\alpha}\mathbf{S}^t\mathbf{H}^t(\mathbf{y} - \mathbf{H}\mathbf{S}\mathbf{w}^{(i)}), \lambda/(2\alpha)\right) \quad (8)$$

where the superscript i is the iteration index, and $\text{soft}(x, T)$ is the soft-threshold function with threshold T ,

$$\text{soft}(x, T) := \text{sign}(x) \max(0, |x| - T) \quad (9)$$

applied element-wise. For convergence of (8) to the minimizer of (4), we can set $\alpha \geq \text{maxeig}(\mathbf{S}^t\mathbf{H}^t\mathbf{H}\mathbf{S})$. Relatively recently, it has been found that convergence is ensured even if this lower bound is halved.^{7,19,34}

Algorithm (8) is known as *iterated soft-thresholding* (IST) or as the *thresholded-Landweber* (TL) algorithm.^{21,22,28,30} Like the algorithm (5-7), the IST algorithm is simple to implement, requiring only procedures to apply \mathbf{A} , \mathbf{A}^t , \mathbf{H} , and \mathbf{H}^t . However, for some problems this algorithm converges slowly. The use of the ℓ_1 norm as a regularizer for linear inverse problems is a central part of sparsity-based signal restoration, and the development of fast algorithms for the minimization of (4), and generalizations thereof, is an active area of research. Faster algorithms are given in Refs. 2, 5, 11, 26, 31, 36, 56.

Note that the objective function (4) is essentially a special case of (3). Algorithms for the minimization of (3) can be used to minimize (4), but not vice versa.

3. SUBBAND-DEPENDENT REGULARIZATION

In wavelet-domain signal restoration, it is common to model the statistical distribution of wavelet coefficients differently in different subbands. For example, the variance of wavelet coefficients is generally smaller in high-frequency subbands than in low-frequency subbands. Accordingly, in wavelet-based signal restoration via the minimization of an objective function, it is appropriate to individually weight subbands in the regularization term. To this end, we partition the wavelet coefficients \mathbf{w} into blocks, each block corresponding to a single subband:

$$\mathbf{w} = [\mathbf{w}_1^t \quad \cdots \quad \mathbf{w}_K^t \quad \mathbf{w}_{\text{LP}}^t]^t = [\mathbf{w}_1; \quad \cdots; \quad \mathbf{w}_K; \quad \mathbf{w}_{\text{LP}}]$$

where \mathbf{w}_{LP} represent the low-pass wavelet coefficients (the ‘scaling’ coefficients at the lowest-resolution scale). For convenience, we use ‘;’ to denote vertical concatenation. There are a total of $K+1$ subbands. For 1D wavelet transforms, there is usually a single subband at each scale. For 2D wavelet transforms, each scale usually has several subbands, but there is still usually a single low-pass subband.

The statistical distribution of wavelet coefficients of natural signals has been extensively studied. It has been found that the distribution of wavelet coefficients is generally heavy-tailed with a sharp peak at zero.⁴⁰ Equivalently, the wavelet coefficients are somewhat sparse. However, this statistical behaviour does not extend to the low-pass coefficients. The low-pass wavelet coefficients of a natural signal are generally far from sparse.

Because the low-pass wavelet coefficients are not sparse, it is not justified to penalize them in the objective function with the ℓ_1 -norm. It is for this reason that, in wavelet-domain thresholding for signal denoising, the low-pass wavelet coefficients are usually unprocessed (usually, no thresholding is applied to the low-pass coefficients). In this paper, we exclude the low-pass wavelet coefficients from the regularization term (the ℓ_1 norm).

In critically-sampled wavelet transforms with a sufficient number of levels, there will be only a few low-pass coefficients, therefore, whether one includes them in the regularization term will often make only a small difference in the result. However, in order that the regularization term follow more closely the behaviour of wavelet coefficients, in the numerical results in the following sections, we do not include the low-pass wavelet coefficients in the regularization term. It is for this reason that we explicitly partition the wavelet coefficients into subbands — we will treat the low-pass wavelet coefficients separately.

3.1 Analysis Prior

We partition the matrix \mathbf{A} into blocks, each block corresponding to a single subband:

$$\mathbf{A} = [\mathbf{A}_1; \cdots; \mathbf{A}_K; \mathbf{A}_{LP}]$$

so that the wavelet coefficients in subband \mathbf{w}_k are given by $\mathbf{A}_k \mathbf{x}$. Then, with $\lambda_{LP} = 0$, the analysis-prior objective function (3) becomes:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \sum_{k=1}^K \lambda_k \|\mathbf{A}_k \mathbf{x}\|_1 \quad (10)$$

with $\lambda_k \geq 0$ for $k = 1, \dots, K$. For the low-pass subband, we set the regularization parameter, λ_{LP} , equal to zero, because the low-pass wavelet coefficients are generally not sparse.

To minimize the objective function (10) we use algorithm (5-7), however, the expression involving λ in (6) must be interpreted as:

$$\frac{2\alpha}{\lambda} |\mathbf{A}\mathbf{x}^{(i)}| = \left[\frac{2\alpha}{\lambda_1} |\mathbf{A}_1 \mathbf{x}^{(i)}|; \frac{2\alpha}{\lambda_2} |\mathbf{A}_2 \mathbf{x}^{(i)}|; \cdots; \frac{2\alpha}{\lambda_K} |\mathbf{A}_K \mathbf{x}^{(i)}| \right].$$

Note that the objective function (10) is an example of a compound regularizer.⁶

3.2 Synthesis Prior

Similarly, for the synthesis prior, we partition \mathbf{S} into blocks as:

$$\mathbf{S} = [\mathbf{S}_1 \cdots \mathbf{S}_K \mathbf{S}_{LP}]$$

so that the signal is given by $\mathbf{x} = \mathbf{S}_1 \mathbf{w}_1 + \cdots + \mathbf{S}_K \mathbf{w}_K + \mathbf{S}_{LP} \mathbf{w}_{LP}$. Then the synthesis-prior objective function (4) becomes:

$$J(\mathbf{w}) = \|\mathbf{y} - \mathbf{H}\mathbf{S}\mathbf{w}\|_2^2 + \sum_{k=1}^K \lambda_k \|\mathbf{w}_k\|_1 \quad (11)$$

with $\lambda_k \geq 0$ for $k = 1, \dots, K$. An algorithm to minimize (11) is given by:

$$\mathbf{w}_k^{(i+1)} = \text{soft} \left(\mathbf{w}_k^{(i)} + \frac{1}{\alpha} \mathbf{S}^t \mathbf{H}^t (\mathbf{y} - \mathbf{H}\mathbf{S} \mathbf{w}^{(i)}), \lambda_k / (2\alpha) \right).$$

3.3 Remarks

If the wavelet transform is denoted \mathbf{W} , then for the analysis prior objective function (3) we will use $\mathbf{A} = \mathbf{W}$; for the synthesis prior objective function (4) we will use $\mathbf{S} = \mathbf{W}^t$. We will use the same regularization parameters for analysis and synthesis priors.

In this paper, we use wavelet transforms for which $\mathbf{W}^t\mathbf{W} = \mathbf{I}$. That is, $\mathbf{S}\mathbf{A} = \mathbf{I}$. In other words, the wavelet transforms we will use are either orthonormal (in the critically-sampled case) or tight frames (in the overcomplete case).

4. SIGNAL RESTORATION EXAMPLES

In this section we compare the analysis prior and synthesis prior for 1D signal restoration using the undecimated dyadic discrete wavelet transform¹⁸ (UDWT).^{*} In Section 4.1 we consider the denoising problem and in Section 4.2 we consider the deconvolution problem.

4.1 Denoising

We first consider the denoising of a piecewise constant signal using the undecimated Haar transform. Specifically, we use the 512-point ‘Blocks’ signal from WaveLab.²³ The noisy signal is obtained by adding zero-mean white Gaussian noise with standard deviation 0.01.

The minimization of the analysis prior objective function $J(\mathbf{x})$ in (3) was performed with 100 iterations of the algorithm in Section 2.1; and the minimization of the synthesis prior objective function $J(\mathbf{x})$ in (4) was performed with 100 iterations of the algorithm in Section 2.2.

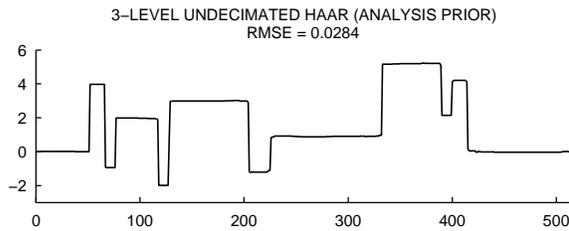
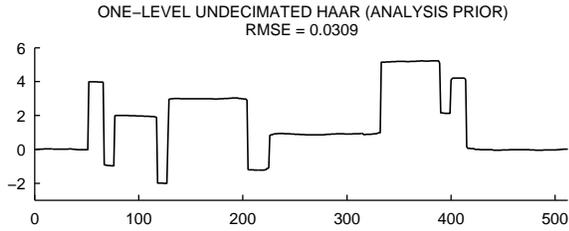
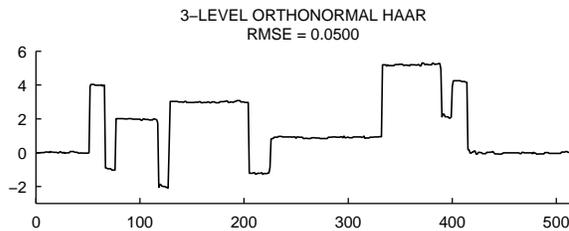
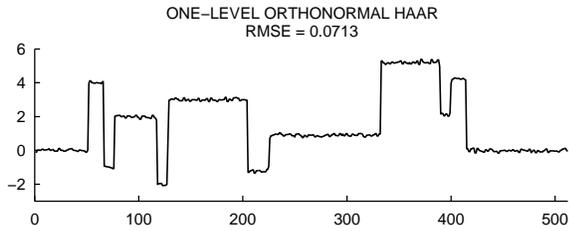
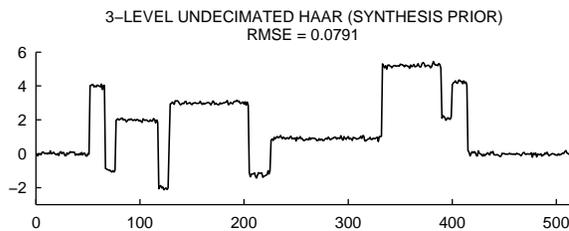
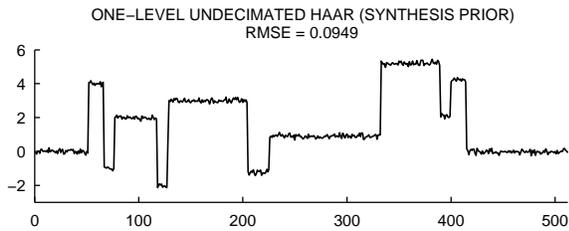
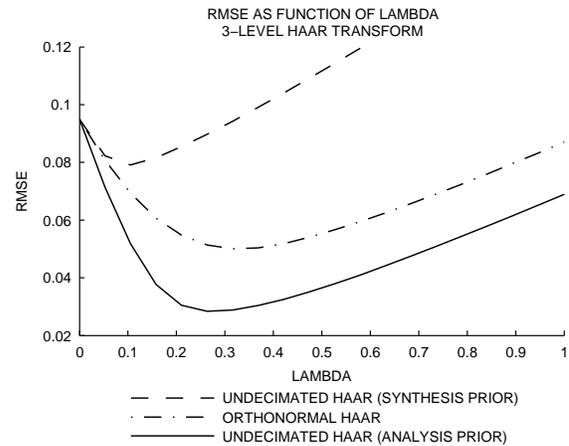
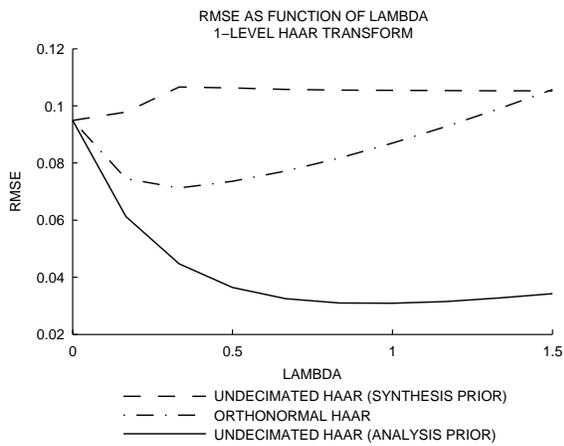
In addition to the undecimated Haar transform, we will use the orthonormal Haar transform for comparison. For the orthonormal Haar transform, the analysis and synthesis prior objective functions are equivalent (they give the same signal \mathbf{x} when minimized). Furthermore, the objective functions are exactly minimized by wavelet-domain soft-thresholding.

One-level Haar: The first example uses only a single decomposition level of the Haar transform, as that is the most basic form. Fig. 1a illustrates the RMSE as a function of the regularization parameter λ . (As there is a single decomposition, the number of subbands in the regularization term is $K = 1$. We do not include the low-pass coefficients; we set $\lambda_{LP} = 0$.) Note that denoising using the *synthesis* prior gives a poor result; in this example, the result obtained using the synthesis prior is slightly worse than the noisy signal itself. However, the result obtained using the *analysis* prior is good. Even though both objective functions use exactly the same transform (the one-level undecimated Haar transform), the results are very different. As illustrated in Fig. 1a, denoising using the one-level orthonormal Haar transform is better than the undecimated case with the synthesis prior, but is worse than the undecimated case with the analysis prior.

This example highlights the difference between the analysis and synthesis priors for signal restoration. This example also highlights the advantage of overcomplete transforms, but shows that how they are utilized (analysis versus synthesis) can make a substantial difference in the results.

The synthesis prior with the one-level UDWT performs poorly in this example because the noisy signal can be reasonably well represented using only the low-pass subband. The synthesis prior as we have formulated it, represents the signal as the output of the undecimated wavelet transform, yet it regularizes (penalizes) only the high-frequency subband. Therefore, minimizing $J(\mathbf{x})$ over the set of all wavelet coefficients including the low-pass coefficients, without regularizing (penalizing) the low-pass coefficients, will put most of the signal energy into the low-pass coefficients. There are several approaches to address this issue with the synthesis prior when it is used with overcomplete wavelet transforms. (1) One approach is to set λ_{LP} to a small positive value instead of setting it to zero. This way, the low-pass coefficients are slightly penalized. For some signals, this may work well, however, the low-pass coefficients rarely exhibit sparsity and therefore including the low-pass coefficients in the regularization term, which is chosen specifically to induce sparsity, is not well justified. (2) Another approach is to use more than a single decomposition level. In that case, the low-pass coefficients are a smaller fraction of the total set of coefficients. As we will see below, using more levels decomposition levels does improve the

^{*}The UDWT is also called the translation-invariant wavelet transform and the stationary wavelet transform.



(a) Denoising using one-level Haar transform.

(a) Denoising using three-level Haar transform.

Figure 1. Denoising of the 'blocks' signal using the one-level and three-level Haar transforms. The illustrated denoised signals correspond to the value of λ minimizing the RMSE for each method.

result obtained using the synthesis prior, but it is still worse than both the analysis prior and the orthonormal transform. (3) A third approach is to minimize the synthesis-prior objective function over only the wavelet coefficients $\mathbf{w}_1, \dots, \mathbf{w}_K$, and to fix \mathbf{w}_{LP} . In this case, \mathbf{w}_{LP} can be set equal to the low-pass coefficients obtained by applying the the wavelet transform to the degraded signal.¹⁰ Then, the absence of \mathbf{w}_{LP} from the regularization term is not problematic, however, the low-pass coefficients will be noisy. Otherwise, this approach requires a method for estimating \mathbf{w}_{LP} . For denoising, one could use the low-pass coefficients of the noisy signal, however, for deconvolution and general inverse problems, the choice of \mathbf{w}_{LP} constitutes another restoration/estimation problem.

For the analysis prior, excluding the low-pass coefficients from the regularization term by setting $\lambda_{LP} = 0$ does not introduce any problem. The low-pass coefficients are not required for the representation of \mathbf{x} because the objective function to be minimized depends directly on \mathbf{x} , not on a set of wavelet coefficients. The analysis-prior does not depend on the representation of a signal as a sum of weighted atoms, so it is not adversely affected by the lack of regularization of a subset of weighting coefficients.

The analysis-prior objective function, based on a one-level undecimated Haar transform with $\lambda_{LP} = 0$, can be written as

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{h} * \mathbf{x}\|_1$$

where the filter coefficients \mathbf{h} are

$$\mathbf{h} = \left[\frac{1}{2}, -\frac{1}{2} \right].$$

The regularization term consists of only a single high-pass subband, and the Haar high-pass filter is simply a first-order difference. Writing out the objective function explicitly gives,

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_k |x(k) - x(k-1)|. \quad (12)$$

The regularization function is well known as the *total variation* of the signal \mathbf{x} , denoted $\text{TV}(\mathbf{x})$:

$$\text{TV}(\mathbf{x}) := \sum_k |x(k) - x(k-1)|. \quad (13)$$

Minimizing (12) is known as TV denoising. Therefore, *1D TV denoising is a special case of wavelet-based denoising*, where specifically, the wavelet transform is the undecimated one-level Haar transform and the analysis-prior is used with $\lambda_{LP} = 0$.

The use of TV for signal denoising started in 1992.⁴⁷ It has been shown extensively that total variation is a very effective regularizer for inverse problems in signal and image processing,^{14,15,37,43,55} like denoising, deconvolution, demosaicing, inpainting, compressed-sensing, etc. TV regularization penalizes differences between adjacent samples and therefore encourages the solution to be piecewise constant. Therefore, it is especially effective for signals and images that are piecewise constant, but also works well for natural imagery. The study of TV regularization for solving inverse problems is a very active area of research, in particular, the development of fast algorithms for the minimization of TV-based objective functions. The relationship between wavelet (Haar) denoising and TV denoising has been described in Ref. 52, however, as illustrated in Fig. 1a, the relationship for 1D signals is an exact equivalence provided the analysis-prior formulation (3) of wavelet signal restoration is considered, rather than the synthesis formulation.

Figure 1a also illustrates that denoising by minimizing the analysis-prior does not introduce Gibbs-like artifacts. The denoising signal is without ringing around the discontinuities.

Three-level Haar: We repeat the denoising experiment illustrated in Fig. 1a, but now we use three-level Haar transforms instead of one-level Haar transforms. In this case, for the orthonormal Haar transform we set $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ and $\lambda_{LP} = 0$ as before. We regularize each subband with the same regularization parameter. (This corresponds to soft-thresholding each subband with the same threshold.) For the undecimated

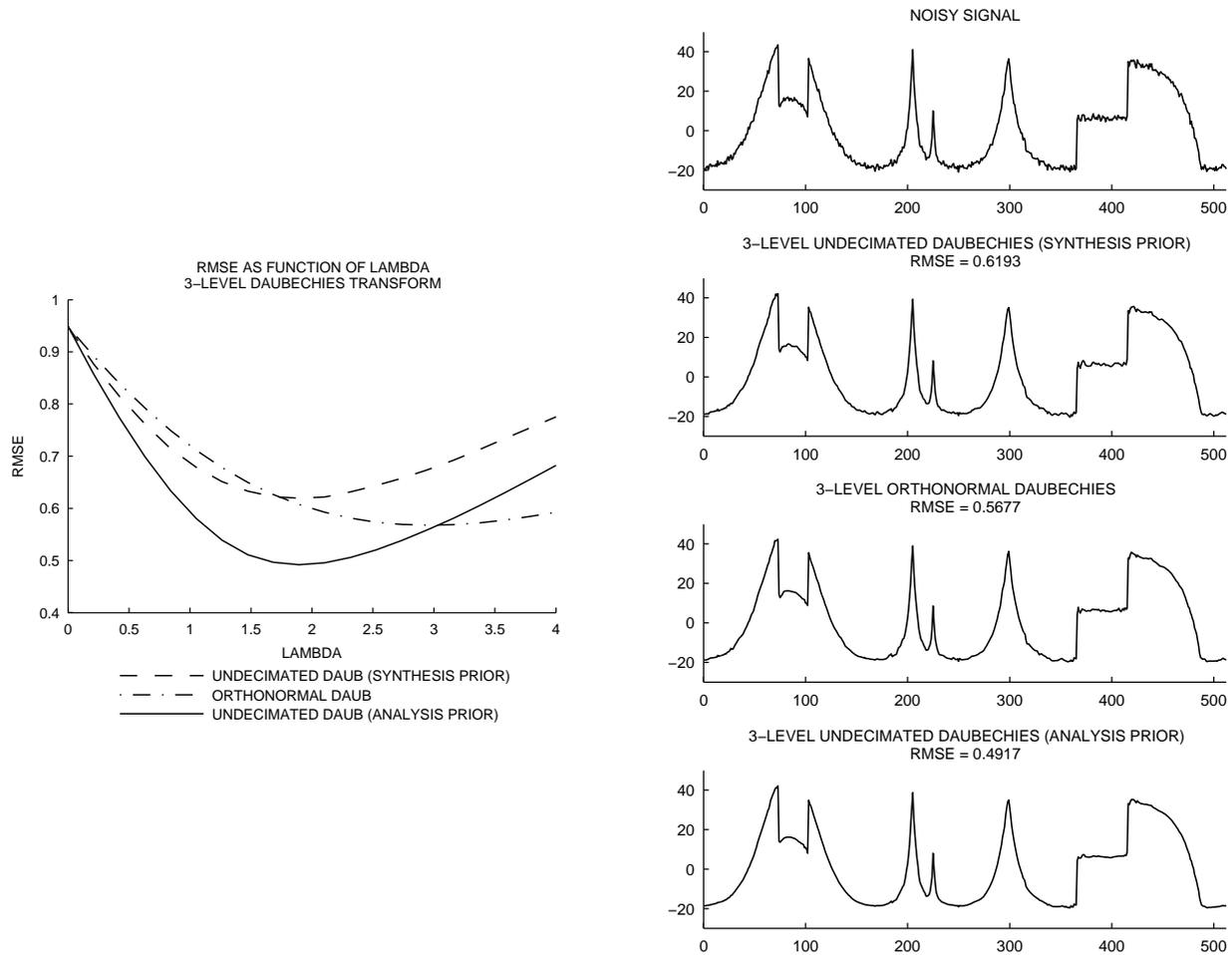


Figure 2. Denoising of the ‘piece-regular’ signal using the three-level Daubechies wavelet transforms with three vanishing moments. The illustrated denoised signals correspond to the value of λ minimizing the RMSE for each method.

Haar transform, the filters for each subband have a different norm; to account for this, we set the regularization parameters as $\lambda_1 = \lambda$, $\lambda_2 = \lambda/\sqrt{2}$, $\lambda_3 = \lambda/2$, and $\lambda_{LP} = 0$. This accounts for the noise variance being different in successive bands of the undecimated wavelet transform.

The results are illustrated in Fig. 1b. Each of the three schemes provide improved results with three levels compared to one level. Yet, the synthesis-prior result is very poor compared to the other two methods. Again, the analysis-prior gives the best result. Note that the analysis-prior is no longer equivalent to TV denoising, yet the result is very similar to that in Fig. 1a.

Three-level Daubechies: In this example, we consider the denoising of a piecewise smooth signal using the wavelet transform with Daubechies filters.²⁰ For denoising signals that are not piecewise constant, it is usually more effective to use a Daubechies wavelet transform with a few vanishing moments rather than the Haar transform (which has one vanishing moment).

Here, the test signal is the 512-point ‘piece-regular’ signal from WaveLab.²³ The noisy signal is obtained by adding zero-mean white Gaussian noise with standard deviation 1.0. The noisy signal is illustrated in Fig. 2. We perform denoising using the three-level orthonormal wavelet transform, and the three-level undecimated wavelet transform; the latter with both the analysis prior and the synthesis prior.

Figure 2 illustrates the RMSE as a function of the regularization parameter, λ . The plot shows that, of the three methods, the undecimated wavelet transform gives the best denoising result when used with the analysis prior, yet gives the worst denoising result when used with the synthesis prior. However, the difference between the three methods is not as substantial for this example as it was for the denoising of the ‘blocks’ signal using the Haar transform in Fig. 1. The restored signal, illustrated in Fig. 2 correspond to the value of λ minimizing the RMSE.

Denoising the ‘piece-regular’ signal using the undecimated Haar transform with the analysis prior (equivalently, TV denoising) would give a denoised signal with a ‘stair-case’ artifact. The stair-case effect is well documented in the TV denoising literature.^{24,25} On the other hand, denoising using an undecimated Daubechies wavelet transform with several vanishing moments, applied as an analysis prior, with $\lambda_{LP} = 0$, serves as a generalization of TV that accommodates signals that are piecewise smooth instead of piecewise constant. Higher order TV penalty functions have been studied previously⁵¹ in order to apply the TV concept effectively to piecewise smooth signals. This example illustrates that the undecimated wavelet transforms used with the analysis prior, is another way to extend the TV concept to higher orders.

The joint use of TV and wavelets/curvelets for denoising piecewise regular functions has been proposed by several authors.^{24,25,48–50} As TV regularization (for one-dimensional signals) is equivalent to wavelet-based regularization (with analysis prior and $\lambda_{LP} = 0$), the use of both TV and wavelets for regularizing an inverse problem can be interpreted as the use of two wavelet transforms simultaneously. The analysis prior formulation, (3), describes this case, where \mathbf{A} is a concatenation of the two wavelet transforms.

The equivalence between TV and the undecimated Haar transform (as an analysis prior) does not extend to two-dimensions. The 2D Haar transform gives three subbands at the first level, while 2D TV is based on a single penalty term. However, the use of overcomplete 2D wavelet transforms with analysis priors, may be an effective alternative to TV for images that are not piecewise constant.

4.2 Deconvolution

The deconvolution problem is considered here. We use the same two signals considered in Section 4.1, namely ‘blocks’ and ‘piece-regular’. We convolve each signal with a 10-point moving average with unity dc gain,

$$h(n) = 0.1, \quad n = 0, \dots, 9$$

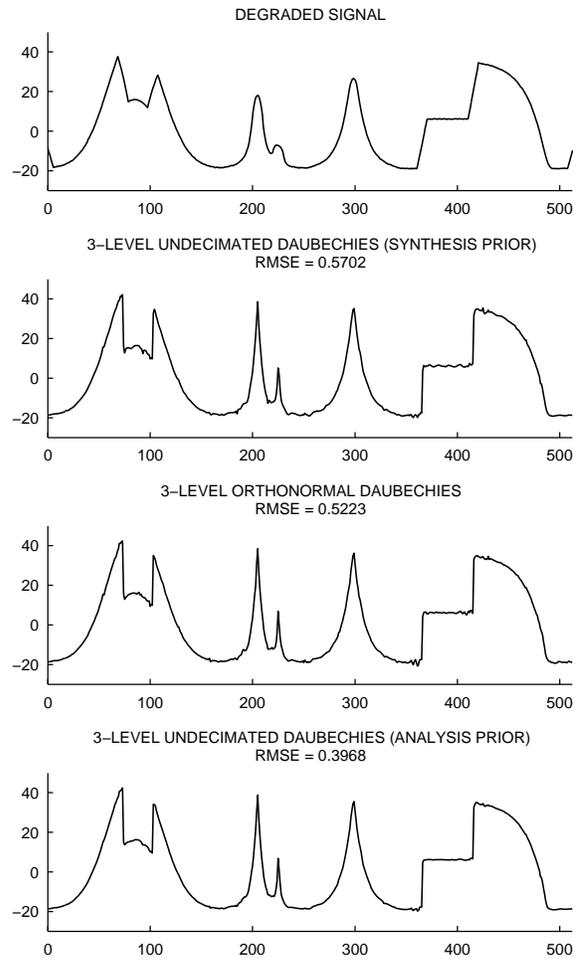
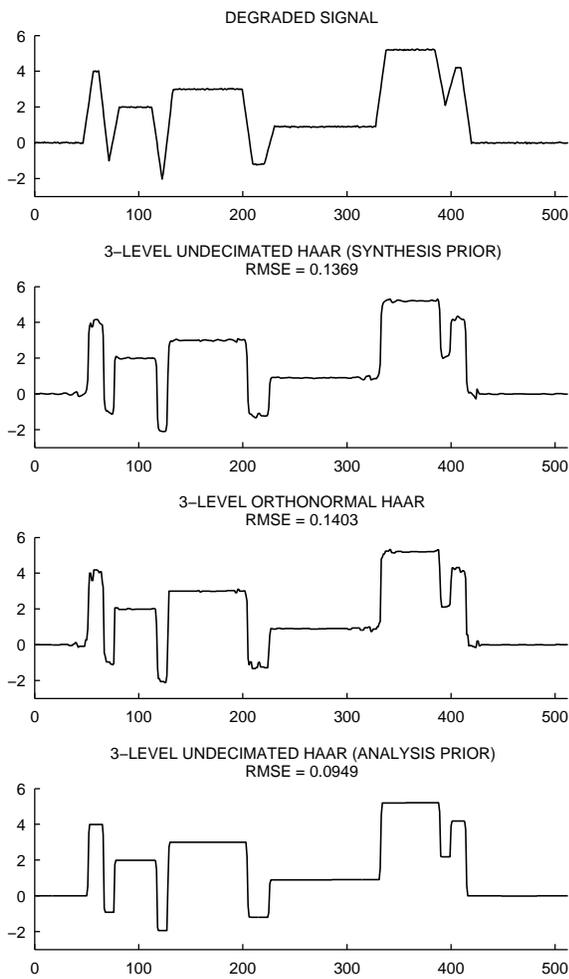
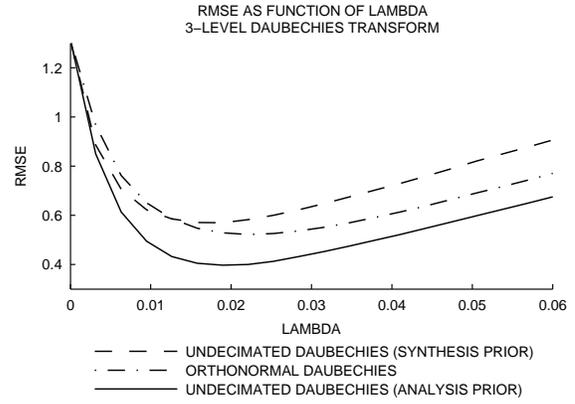
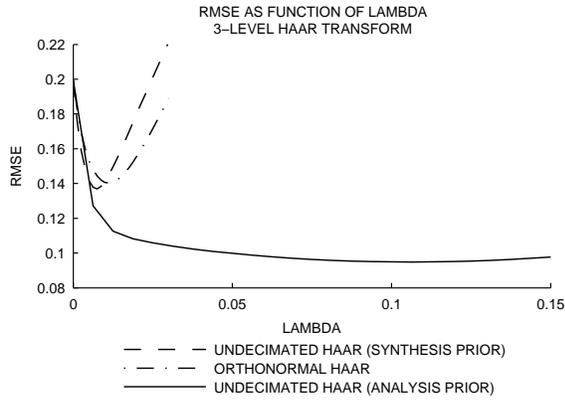
and add independent white Gaussian noise (for ‘blocks’ we use standard deviation of 0.02 while for ‘piece-regular’ we use a standard deviation of 0.1.) The degraded signals are illustrated in Fig. 3. Note that the blurring function $h(n)$ is far from being invertible.

For restoring the ‘blocks’ signal we use the three-level Haar transform (orthonormal and undecimated, the later with both the analysis and synthesis prior). Similarly, for restoring the ‘piece-regular’ signal we use the three-level Daubechies wavelet transform with three vanishing moments. We used 1000 iterations of the algorithms in Sections 2. Many more iterations are required than for denoising because of the ill-conditioning of \mathbf{H} . Faster algorithms will require fewer iterations, for example the algorithms of Refs. 1, 2, 56. However, minimizing execution time was not our main concern here because for these one-dimensional signals the basic algorithms in Sections 2 did not consume substantial execution time.

Figure 3 illustrates the RMSE as a function of λ . In this example, the undecimated wavelet transforms with the synthesis prior give similar performance as the orthonormal transforms. However, the undecimated wavelet transforms with the analysis prior gives markedly better results, moreover, the result has very little ringing around the discontinuities.

5. CONCLUSION

The use of wavelet-domain sparsity-based priors is an effective method for the restoration of many natural signals. For denoising, it is well-recognized that overcomplete wavelet transforms improve denoising performance.¹⁸ However, for the more general linear inverse problem (deconvolution, etc), the way in which the problem is posed can make a substantial impact on the results. The presented examples show that when a synthesis prior



(a) Deconvolution of 'blocks' signal.

(b) Deconvolution of 'piece-regular' signal.

Figure 3. Signal deconvolution: (a) Deconvolution of the 'blocks' signal using the three-level Haar transform. (b) Deconvolution of the 'piece-regular' signal using the three-level Daubechies wavelet transform (with 3 vanishing moments). The illustrated restored signals correspond to the value of λ minimizing the RMSE for each method.

is used and the low-pass wavelet coefficients are not regularized, then an overcomplete wavelet transform can yield results that are inferior to an otherwise similar orthonormal transform. We have suggested that the low-pass coefficients should not be regularized because the low-pass coefficients do not exhibit sparse behaviour. In contrast to the synthesis prior, the analysis prior formulation does not depend on the low-pass coefficients and is therefore not adversely affected when they are not regularized. This illustrates the type of problem for which the analysis prior formulation is more effective than the synthesis prior formulation. It is also interesting to note that in 1D the highly effective total variation prior is a special case of an analysis prior formulation where the overcomplete wavelet transform is the one-level undecimated Haar transform and where the low-pass coefficients are unregularized ($\lambda_{LP} = 0$). The effectiveness of TV, together with ineffectiveness of the synthesis prior formulation of the same transform (see Fig. 1a), suggests that the analysis prior might be an appropriate formulation for signal restoration more generally (see Fig. 2).

Note that the ℓ_1 -norm is quite simplistic when compared with other models.^{33,45} For example, when used for denoising with an orthonormal wavelet transform, the ℓ_1 -norm regularizer leads to pure soft-thresholding in the wavelet domain. It does not take into account the statistical dependence among neighboring coefficients (spatially and in successive scales) which is known to improve performance. It is expected that subband-adaptive (or context adaptive) regularizers would give better results. However, the purpose of this paper is to compare the analysis prior and synthesis prior formulations for the signal restoration problem. Improved priors may improve the performance of both the analysis and synthesis-prior formulations.

APPENDIX A. INFORMAL DERIVATION OF ANALYSIS PRIOR ALGORITHM

We provide an informal derivation of the simple algorithm presented in Section 2.1 for the minimization of (3). Using the majorization-minimization (MM) approach,⁴ an algorithm for finding the minimizer of (3) is:

$$\mathbf{b}^{(i)} = \mathbf{x}^{(i)} + \frac{1}{\alpha} \mathbf{H}^t (\mathbf{y} - \mathbf{H} \mathbf{x}^{(i)})$$

$$\mathbf{x}^{(i+1)} = \underset{\mathbf{x}}{\operatorname{argmin}} \left[\|\mathbf{b}^{(i)} - \mathbf{x}\|_2^2 + \frac{\lambda}{\alpha} \|\mathbf{A} \mathbf{x}\|_1 \right]$$

where $\alpha > \max \operatorname{eig}(\mathbf{H}^t \mathbf{H})$.

The second step is a denoising problem, for which an iterative algorithm is given by (23)-(24) in Section A.1. Using just one iteration of (23)-(24) per denoising step gives the algorithm (5)-(7).

A.1 Denoising

For denoising, we need to minimize the objective function

$$J(\mathbf{x}) = \|\mathbf{b} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{A} \mathbf{x}\|_1 \quad (14)$$

Taking the derivative of $J(\mathbf{x})$ gives $\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{x} - 2\mathbf{b} + \lambda \mathbf{A}^t \mathbf{\Lambda} \mathbf{A} \mathbf{x}$, where $\mathbf{\Lambda}$ is the diagonal matrix, $\mathbf{\Lambda} = \operatorname{diag}(1./|\mathbf{A} \mathbf{x}|)$, where $./$ represents element-wise division. Setting the derivatives to zero gives the matrix equation

$$\left(\mathbf{I} + \frac{\lambda}{2} \mathbf{A}^t \mathbf{\Lambda} \mathbf{A} \right) \mathbf{x} = \mathbf{b}. \quad (15)$$

The same system of equations is derived using the majorization-minimization approach^{4,29} which also provides convergence results. This is a non-linear system of equations for \mathbf{x} because $\mathbf{\Lambda}$ depends on \mathbf{x} . However, it suggests the iteration

$$\mathbf{\Lambda}^{(i)} = \operatorname{diag}(1./|\mathbf{A} \mathbf{x}^{(i)}|) \quad (16)$$

$$\mathbf{x}^{(i+1)} = \left(\mathbf{I} + \frac{\lambda}{2} \mathbf{A}^t \mathbf{\Lambda}^{(i)} \mathbf{A} \right)^{-1} \mathbf{b}. \quad (17)$$

On each iteration \mathbf{x} is obtained by solving a linear system of equations, and $\mathbf{\Lambda}$ is updated based on the new \mathbf{x} . An issue with iteration (17) is that each iteration requires the solution to a large (in practice, very large) system of linear equations. To alleviate that problem, an approximate solution can be found on each iteration using a few iterations of an iterative linear system solution method. Another issue is that in (16) we may have division by very small numbers. In fact, small numbers and zero are expected because $|\mathbf{Ax}|$ is expected to be sparse.

In Ref. 29 it is proposed that the matrix inversion lemma be used to rewrite (15) so as to avoid the division by very small numbers. From the matrix inversion lemma, we have

$$\left(\mathbf{I} + \frac{\lambda}{2} \mathbf{A}^t \mathbf{\Lambda} \mathbf{A}\right)^{-1} = \mathbf{I} - \mathbf{A}^t \left(\frac{2}{\lambda} \mathbf{\Lambda}^{-1} + \mathbf{A} \mathbf{A}^t\right)^{-1} \mathbf{A}.$$

So (15) is equivalent to

$$\mathbf{x} = \mathbf{b} - \mathbf{A}^t \left(\frac{2}{\lambda} \mathbf{\Lambda}^{-1} + \mathbf{A} \mathbf{A}^t\right)^{-1} \mathbf{A} \mathbf{b}$$

or

$$\mathbf{z} = \left(\frac{2}{\lambda} \mathbf{\Lambda}^{-1} + \mathbf{A} \mathbf{A}^t\right)^{-1} \mathbf{A} \mathbf{b}, \quad \mathbf{x} = \mathbf{b} - \mathbf{A}^t \mathbf{z}. \quad (18)$$

So the update (17) is equivalent to

$$\mathbf{z}^{(i+1)} = \left(\frac{2}{\lambda} \text{diag}(|\mathbf{Ax}^{(i)}|) + \mathbf{A} \mathbf{A}^t\right)^{-1} \mathbf{A} \mathbf{b} \quad (19)$$

$$\mathbf{x}^{(i+1)} = \mathbf{b} - \mathbf{A}^t \mathbf{z}^{(i+1)}. \quad (20)$$

This avoids division by small numbers, however it still calls for the solution to a large system of linear equations. It is proposed in Ref. 29 that a few iterations of the conjugate gradient method be used to solve (19). Another approach is the following one: To solve for \mathbf{z} in (18), we need to solve

$$\left(\frac{2}{\lambda} \mathbf{\Lambda}^{-1} + \mathbf{A} \mathbf{A}^t\right) \mathbf{z} = \mathbf{A} \mathbf{b}. \quad (21)$$

Adding $c\mathbf{z}$ to both sides and subtracting $\mathbf{A} \mathbf{A}^t \mathbf{z}$ from both sides gives the equivalent equation

$$\left(\frac{2}{\lambda} \mathbf{\Lambda}^{-1} + c\mathbf{I}\right) \mathbf{z} = \mathbf{A} \mathbf{b} + (c\mathbf{I} - \mathbf{A} \mathbf{A}^t) \mathbf{z}$$

which suggests the update rule

$$\mathbf{z}^{(k+1)} = \left(\frac{2}{\lambda} \mathbf{\Lambda}^{-1} + c\mathbf{I}\right)^{-1} \left(\mathbf{A} \mathbf{b} + (c\mathbf{I} - \mathbf{A} \mathbf{A}^t) \mathbf{z}^{(k)}\right) \quad (22)$$

which is easy because the matrix to be inverted is diagonal. In order to ensure convergence of $\mathbf{z}^{(k)}$ to the solution of (21), we should have $c > \text{maxeig}(\mathbf{A} \mathbf{A}^t)$. This bound is derived in Section A.2 using eigenvalue analysis.

With this, an algorithm for minimizing minimizing (14) is given by:

$$\mathbf{z}^{(i+1)} = \left(\mathbf{A} \mathbf{b} + (c\mathbf{I} - \mathbf{A} \mathbf{A}^t) \mathbf{z}^{(i)}\right) ./ \left(\frac{2}{\lambda} \text{diag}(|\mathbf{Ax}^{(i)}|) + c\right) \quad (23)$$

$$\mathbf{x}^{(i+1)} = \mathbf{b} - \mathbf{A}^t \mathbf{z}^{(i+1)} \quad (24)$$

where $c > \text{maxeig}(\mathbf{A} \mathbf{A}^t)$. The notation $./$ and $|\cdot|$ in (23) are element-wise division and element-wise absolute value. This is a one-dimensional version of Chambolle's algorithm.¹³ The derivation in Ref. 13 uses the dual formulation of the minimization problem to introduce the auxiliary variable; while here, the matrix inversion lemma introduces the auxiliary variable \mathbf{z} . This shows the relationship between Chambolle's algorithm and the MM-based algorithm of Ref. 29. Where Ref. 29 uses several conjugate gradient iterations, Ref. 13 uses a single iteration of the iteration (22).

A.2 The Selection of c

How should c be chosen so as to ensure the update (22) converges to the solution of (21)? It is known⁵³ that $\mathbf{x}^{(k+1)} = \mathbf{P}^{-1}((\mathbf{P} - \mathbf{A})\mathbf{x}^{(k)} + \mathbf{b})$ converges to the solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ if the spectral radius of $\mathbf{I} - \mathbf{P}^{-1}\mathbf{A}$ is less than 1. Therefore (22) converges to the solution of (21) if the eigenvalues of

$$\mathbf{M} := \mathbf{I} - \left(\frac{2}{\lambda}\mathbf{\Lambda}^{-1} + c\mathbf{I}\right)^{-1} \left(\frac{2}{\lambda}\mathbf{\Lambda}^{-1} + \mathbf{A}\mathbf{A}^t\right) \quad (25)$$

are all less than 1 in absolute value which is ensured if $c > \max\text{eig}(\mathbf{A}\mathbf{A}^t)$.

We want to show that the matrix \mathbf{M} in (25) has spectral radius less than 1. Assume:

1. \mathbf{V} is a symmetric positive semi-definite matrix.
2. c is greater than the maximum eigenvalue of $\mathbf{G}^t\mathbf{G}$.
3. The minimum eigenvalue of $\mathbf{G}^t\mathbf{G}$ is positive.

Then the eigenvalues of the matrix

$$\mathbf{M} := \mathbf{I} - (\mathbf{V} + c\mathbf{I})^{-1} (\mathbf{V} + \mathbf{G}^t\mathbf{G})$$

are real and lie between 0 and 1.

To show this, note

$$\mathbf{M} = \mathbf{I} - (\mathbf{V} + c\mathbf{I})^{-1} (\mathbf{V} + \mathbf{G}^t\mathbf{G}) = (\mathbf{V} + c\mathbf{I})^{-1} [(\mathbf{V} + c\mathbf{I}) - (\mathbf{V} + \mathbf{G}^t\mathbf{G})] = (\mathbf{V} + c\mathbf{I})^{-1} (c\mathbf{I} - \mathbf{G}^t\mathbf{G})$$

If c is greater than the spectral radius of $\mathbf{G}^t\mathbf{G}$, then $(c\mathbf{I} - \mathbf{G}^t\mathbf{G})$ is a positive matrix; its eigenvalues are real and positive. The matrix $(\mathbf{V} + c\mathbf{I})$ is positive definite because $c > 0$, so $(\mathbf{V} + c\mathbf{I})^{-1}$ is also positive definite. Since the product of two symmetric positive definite matrices is a matrix with real positive eigenvalues, the matrix \mathbf{M} has real positive eigenvalues. Now we need to show that the eigenvalues of \mathbf{M} are less than 1.

The matrix, $(\mathbf{V} + c\mathbf{I})^{-1} (\mathbf{V} + \mathbf{G}^t\mathbf{G})$, is also the product of two symmetric positive definite matrices. Therefore, it has real positive eigenvalues. Therefore, the eigenvalues of \mathbf{M} are less than 1. Note that using a larger value of c will lead to slower convergence.

For total variation regularization, \mathbf{A} is the first-order difference operator, \mathbf{D} . The eigenvalues of $\mathbf{D}\mathbf{D}^t$ are strictly between 0 and 4. Therefore, for Chambolle's algorithm $c = 4$.

REFERENCES

- [1] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. Preprint, retrieved from <http://iew3.technion.ac.il/Home/Users/becka.html>, 2009.
- [2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.*, 2(1):183–202, 2009.
- [3] J. Bect, L. Blanc-Féaud, G. Aubert, and A. Chambolle. A l^1 -unified variational framework for image restoration. In T. Pajdla and J. Matas, editors, *European Conference on Computer Vision, Lecture Notes in Computer Sciences*, volume 3024, pages 1–13, 2004.
- [4] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira. Total-variation image deconvolution: A majorization-minimization approach. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2006.
- [5] J. M. Bioucas-Dias and M. A. T. Figueiredo. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. on Image Processing*, 16(12):2992–3004, December 2007.
- [6] J. M. Bioucas-Dias and M. A. T. Figueiredo. An iterative algorithm for linear inverse problems with compound regularizers. In *Proc. IEEE Int. Conf. Image Processing*, October 2008.
- [7] K. Bredies and D. Lorenz. Linear convergence of iterative soft-thresholding. *Journal of Fourier Analysis and Applications*, 14(5-6):813–837, December 2008.

- [8] J.-F. Cai, S. Osher, and Z. Shen. Linearized Bregman iterations for frame-based image deblurring. *SIAM J. Imag. Sci.*, 2(1):226–252, 2009.
- [9] J.-F. Cai, S. Osher, and Z. Shen. Split Bregman methods and frame based image restoration. preprint, 2009.
- [10] J.-F. Cai and Z. Shen. Framelet based deconvolution. *J. Comp. Math.*, 2008. preprint.
- [11] E. Candes and J. Romberg. L1-magic: Recovery of sparse signals. <http://www.acm.caltech.edu/l1magic/>, 2005.
- [12] A. Chai and Z. Shen. Deconvolution: a wavelet frame approach. *Numer. Math.*, 106(4):529–587, 2007.
- [13] A. Chambolle. An algorithm for total variation minimization and applications. *J. of Math. Imaging and Vision*, 20:89–97, 2004.
- [14] T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. on Image Processing*, 10(2):231–241, February 2001.
- [15] R. Chartrand and V. Staneva. Total variation regularisation of images corrupted by non-Gaussian noise using a quasi-Newton method. *Image Processing, IET*, 2(6):295–303, December 2008.
- [16] C. Chau, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame based inverse problems. *Inverse Problems*, 23:1495–1518, June 2007.
- [17] S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.
- [18] R. R. Coifman and D. L. Donoho. Translation-invariant de-noising. In A. Antoniadis, editor, *Wavelets and Statistics*. Springer-Verlag Lecture Notes, 1995.
- [19] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.
- [20] I. Daubechies. *Ten Lectures On Wavelets*. SIAM, 1992.
- [21] I. Daubechies, M. Defriese, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math*, LVII:1413–1457, 2004.
- [22] I. Daubechies, G. Teschke, and L. Vese. On some iterative concepts for image restoration. In P. W. Hawkes, editor, *Advances in Imaging and Electron Physics*, volume 150, pages 2–51. Elsevier, 2008.
- [23] D. Donoho, A. Maleki, and M. Shahram. Wavelab 850. <http://www-stat.stanford.edu/~wavelab/>.
- [24] S. Durand and J. Froment. Artifact free signal denoising with wavelets. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2001.
- [25] S. Durand and J. Froment. Reconstruction of wavelet coefficients using total variation minimization. *SIAM J. Sci. Comput.*, 24(5):1754–1767, 2003.
- [26] M. Elad, B. Matalon, and M. Zibulevsky. Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization. *J. of Appl. and Comp. Harm. Analysis*, 23:346–367, 2007.
- [27] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23:947–968, 2007.
- [28] M. Figueiredo, J. Bioucas-Dias, and R. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Trans. on Image Processing*, 16(12):2980–2991, December 2007.
- [29] M. Figueiredo, J. Bioucas-Dias, J. P. Oliveira, and R. D. Nowak. On total-variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavelet denoising. In *Proc. IEEE Int. Conf. Image Processing*, 2006.
- [30] M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. on Image Processing*, 12(8):906–916, August 2003.
- [31] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.*, 1(4):586–598, December 2007.
- [32] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imag. Sci.*, 2(2):323–343, 2009.
- [33] J. A. Guerrero-Colon, L. Mancera, and J. Portilla. Image restoration using space-variant Gaussian scale mixtures in overcomplete pyramids. *IEEE Trans. on Image Processing*, 17(1):27–41, January 2008.

- [34] E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence. *SIAM J. on Optimization*, 19(3):1107–1130, 2008.
- [35] B.-B. Hao, M. Li, and X.-C. Feng. Wavelet iterative regularization for image restoration with varying scale parameter. *Signal Processing: Image Communication*, 23(6):433–441, 2008.
- [36] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale ℓ_1 -regularized least squares. *IEEE. J. Sel. Top. Signal Processing*, 1(4):606–617, December 2007.
- [37] S.-H. Lee and M. G. Kang. Total variation-based image noise reduction with generalized fidelity function. *IEEE Signal Processing Letters*, 14(11):832–835, November 2007.
- [38] M. Li, B. Hao, and X. Feng. Iterative regularization and nonlinear inverse scale space based on translation invariant wavelet shrinkage. *Int. J. Wavelets Multiresolut. Inf. Process.*, 6(1):83–95, January 2008.
- [39] I. Loris, G. Nolet, I. Daubechies, and F.A. Dahlen. Tomographic inversion using ℓ_1 -norm regularization of wavelet coefficients. *Geophys. J. Int.*, 170:359–379, 2007.
- [40] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1998.
- [41] R. Neelamani, H. Choi, and R. Baraniuk. ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Trans. on Signal Processing*, 52(2):418–433, Feb. 2004.
- [42] J. Oliveira, J. Bioucas-Dias, and M. A. T. Figueiredo. Adaptive total variation image deblurring: A majorization-minimization approach. *Signal Processing*, 89(9):1683–1693, September 2009.
- [43] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation based image restoration. *Multiscale Model. & Simul.*, 4(2):460–489, 2005.
- [44] J. Portilla and L. Mancera. L_0 -based sparse approximation: two alternative methods and some applications. In *Proceedings of SPIE*, volume 6701 (Wavelets XII), 2007.
- [45] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. on Image Processing*, 12(11):1338–1351, November 2003.
- [46] P. Rodríguez and B. Wohlberg. Efficient minimization method for a generalized total variation functional. *IEEE Trans. on Image Processing*, 18(2):322–332, February 2009.
- [47] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [48] J.-L. Starck, E. J. Candés, and D. L. Donoho. Very high quality image restoration by combining wavelets and curvelets. In *Proceedings of SPIE*, volume 4478 (Wavelet Applications in Signal and Image Processing IX), 2001.
- [49] J.-L. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representation and a variational approach. *IEEE Trans. on Image Processing*, 14(10), 2005.
- [50] J.-L. Starck, M. K. Nguyen, and F. Murtagh. Wavelets and curvelets for image deconvolution: a combined approach. *Signal Processing*, 83(10):2279–2283, 2003.
- [51] G. Steidl. A note on the dual treatment of higher-order regularization functionals. *Computing*, 76(1):135–148, 2006.
- [52] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDes. *SIAM J. Numer. Anal.*, 42:686–713, 2004.
- [53] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, 1986.
- [54] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. on Imaging Sciences*, 1(3):248–272, 2008.
- [55] Y. Wang and H. Zhou. Total variation wavelet-based medical image denoising. *Int. J. of Biomedical Imaging*, pages 1–6, 2006. Article ID 89095.
- [56] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. on Signal Processing*, 57(7):2479–2493, July 2009.
- [57] J. Xu and S. Osher. Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising. *IEEE Trans. on Image Processing*, 16(2):534–544, February 2007.
- [58] M. Zhu, S. J. Wright, and T. F. Chan. Duality-based algorithms for total-variation-regularized image restoration. *J. Comp. Optimization and Applications*, 2008.