Wavelet Shrinkage With Consistent Cycle Spinning Generalizes Total Variation Denoising

Ulugbek Kamilov, *Student Member, IEEE*, Emrah Bostan, *Student Member, IEEE*, and Michael Unser, *Fellow, IEEE*

***Abstract*—We introduce a new wavelet-based method for the implementation of Total-Variation-type denoising. The data term is least-squares, while the regularization term is gradient-based. The particularity of our method is to exploit a link between the discrete gradient and wavelet shrinkage with cycle spinning, which we express by using redundant wavelets. The redundancy of the representation gives us the freedom to enforce additional constraints (e.g., normalization) on the solution to the denoising problem. We perform optimization in an augmented-Lagrangian framework, which decouples the difficult n-dimensional constrained-optimization problem into a sequence of n easier scalar unconstrained problems that we solve efficiently via traditional wavelet shrinkage. Our method can handle arbitrary gradient-based regularizers. In particular, it can be made to adhere to the popular principle of least total variation. It can also be used as a *maximum a posteriori* estimator for a variety of priors. We illustrate the performance of our method for image denoising and for the statistical estimation of sparse stochastic processes.**

***Index Terms*—Signal denoising, total variation, wavelet regularization, cycle spinning, augmented Lagrangian.**

I. Introduction

WAVELET regularization [1] has been shown to be particularly effective in reducing noise, while preserving important signal features. The performance of the method can be further improved, at little computational cost, by using the technique known as cycle spinning [2]–[4]. Cycle spinning compensates for the lack of shift-invariance of the wavelet basis by considering different shifts of the signal. Total variation (TV) [5] regularization is another widely used denoising method, which penalizes random oscillations in the signal, while allowing the discontinuities. Interestingly, for 1-D signals, Haar-wavelet shrinkage with cycle spinning has been shown to be closely related to TV regularization [6].

In this letter, we exploit the link between the two estimation methods to derive a new wavelet-based method for efficiently solving TV-type denoising problems in one or 2-D. The key observation is that TV-regularized least-squares minimization can be reformulated as a constrained optimization problem in the wavelet domain. We use the augmented-Lagrangian method [7] to cast the problem as a sequence of unconstrained problems that can be solved by simple soft-thresholding. By replacing the soft-thresholding function by another scalar function or by a precomputed lookup table, we can efficiently extend our algorithm beyond traditional ℓ_1 regularizers to general, possibly non-convex, potential functions.

This letter is organized as follows: In Section II, we review wavelet and TV regularization methods. Our main results are presented in Section III. In Section IV, we perform numerical experiments and illustrate the applications of the method.

Manuscript received November 09, 2011; revised January 05, 2012; accepted January 16, 2012. Date of publication January 27, 2012; date of current version February 09, 2012. This work was supported by the European Commission under Grant ERC-2010-AdG 267439-FUN-SP. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Kenneth E. Barner.

The authors are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne VD, Switzerland (e-mail: ulugbek.kamilov@epfl.ch; emrah.bostan@epfl.ch; michael.unser@epfl.ch).

Digital Object Identifier 10.1109/LSP.2012.2185929

II. The Basic Methods

Consider the signal denoising problem for which the goal is to estimate an unknown signal $\mathbf{x} \in \mathbb{R}^n$ from a noisy observation

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

where the vector $\mathbf{n} \in \mathbb{R}^n$ is assumed to be an i.i.d. Gaussian-noise vector. We assume periodic boundary conditions and vectorize 2-D images in a column-major order.

A. Wavelet-Based Denoising

In a wavelet based framework, the unknown signal to estimate is expressed as $\mathbf{x} = \mathbf{W}^T\mathbf{w}$, where $\mathbf{W}$ represents an orthonormal wavelet transform. Then, the estimation can be performed by the regularized least-squares problem

$$\hat{\mathbf{w}}(\mathbf{y}) = \underset{\mathbf{w}\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{w} - \mathbf{W}\mathbf{y}\|_2^2 + \tau\Phi(\mathbf{w}) \right\} \tag{1}$$

where the function $\Phi(\,\cdot\,)$ is a *regularizer* that promotes solutions with certain desirable properties, and $\tau \in \mathbb{R}_+$ is the *regularization parameter*. For example, the popular wavelet-shrinkage algorithm can be obtained by using the non-smooth convex function $\Phi(\,\cdot\,) = \|\cdot\|_1$ that favors sparse wavelet-domain solutions and admits the closed form solution

$$\hat{\mathbf{w}}(\mathbf{y}) = \eta(\mathbf{W}\mathbf{y}; \tau) \tag{2}$$

where the soft-thresholding function $\eta(\,\cdot\,)$ is applied componentwise [1]. The final signal estimate is obtained by computing $\hat{\mathbf{x}} = \mathbf{W}^T\hat{\mathbf{w}}$. If the wavelet representation $\mathbf{w}$ has both coarse (low-pass) and detail (high-pass) coefficients, the shrinkage is performed only on the detail coefficients.

The performance of wavelet shrinkage can be improved via the cycle-spinning method. Consider the transform

$$\mathbf{H} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_K \end{bmatrix} \tag{3}$$

1070-9908/$31.00 © 2012 IEEE

with $\mathbf{W}_k = \mathbf{W}\mathbf{S}_k$, where $\mathbf{S}_k$ represent different circular shifts required to make the transform $\mathbf{W}$ integer-shift-invariant. The transform $\mathbf{H}$ consists of K unitary transforms $\mathbf{W}_k$ and thus defines a *K-tight frame*, which implies the Parseval-like equality

$$\|\mathbf{x}\|_2^2 = \frac{1}{K}\|\mathbf{H}\mathbf{x}\|_2^2. \tag{4}$$

The dual of $\mathbf{H}$ is defined as $\mathbf{H}^\dagger = (1/K)[\mathbf{W}_1^T \dots \mathbf{W}_K^T]$, where the operation $\mathbf{W}_k^T$ simply corresponds to first inverting the wavelet coefficients by applying $\mathbf{W}^T$ and then performing inverse circular shift $\mathbf{S}_{-k}$. The estimation is performed by

$$\hat{\mathbf{w}}(\mathbf{y}) = \underset{\mathbf{w}\in\mathbb{R}^{nK}}{\operatorname{argmin}} \left\{ \frac{1}{2K}\|\mathbf{w} - \mathbf{H}\mathbf{y}\|_2^2 + \tau\Phi(\mathbf{w}) \right\} \tag{5}$$

where the regularizer $\Phi(\,\cdot\,)$ operates only on the detail wavelet coefficients. The final reconstruction is obtained via

$$\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{H}^\dagger\hat{\mathbf{w}}(\mathbf{y}). \tag{6}$$

For $\Phi(\,\cdot\,) = \|\cdot\|_1$, this formulation results in the method described in [2]. More recently, this approach has become popular in signal denoising and inverse problems to improve the performance of the wavelet-regularization methods at little computational cost [3], [4]. In this letter we refer to that type of cycle spinning as *conventional cycle spinning*, to distinguish it from our method.

B. Total-Variation Regularization

TV denoising works by solving the optimization problem

$$\hat{\mathbf{x}}(\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^{n}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \tau\|\mathrm{L}\{\mathbf{x}\}\|_1 \right\} \tag{7}$$

where $\mathrm{L}\{\mathbf{x}\} \in \mathbb{R}^n$ is a vector that contains the Euclidean norm of the discrete gradient of $\mathbf{x}$ at all indices $i \in [1 \dots n]$, defined component-wise as

$$[\mathrm{L}\{\mathbf{x}\}]_i = \begin{cases} |x_{i+1} - x_i| & 1\text{−D signals} \\ \sqrt{[\mathrm{D}_h\mathbf{x}]_i^2 + [\mathrm{D}_\mathrm{v}\mathbf{x}]_i^2} & 2\text{−D signals} \end{cases}. \tag{8}$$

Here, D_h and D_v are matrices corresponding to the first order finite difference in the horizontal and vertical directions of the 2-D image, respectively. Although solving (7) is more difficult than solving (1) due to the coupling introduced by the operator $\mathrm{L}\{\cdot\}$, algorithms such as FISTA [8] are capable of efficiently performing TV denoising.

By comparing the equation for the discrete gradient of the signal $\mathbf{x}$ at some pixel i with the Haar-wavelet expansion, which involves sums and differences [9], we can guess that TV denoising is related to wavelet shrinkage. In fact, it has been shown for 1D signals that a single iteration of a numerical TV denoising scheme (from [6]) is equivalent to Haar-wavelet shrinkage with conventional cycle spinning.

III. Proposed Method

A. Motivation

In the sequel, we will introduce a new, simple, and generic wavelet-based method for solving (7) that can be viewed as an extension to the conventional cycle spinning. As mentioned, due to the tightness of the frame, equality (4) always holds. However, as the transform is redundant, the converse equality $\|\mathbf{w}\|_2^2 = K\|\mathbf{H}^\dagger\mathbf{w}\|_2^2$ is not *necessarily true*. To restore a norm equivalence, we must impose consistency in the optimization problem (5). It turns out that the norm equivalence holds if and only if the solution $\mathbf{w} \in \mathbb{R}^{nK}$ is within the space spanned by the columns of $\mathbf{H}$. In other words, for any potential $\mathbf{w}$ there must exist $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{w} = \mathbf{H}\mathbf{x}$. An alternative way to represent this constraint is to enforce $\mathbf{H}\mathbf{H}^\dagger\mathbf{w} = \mathbf{w}$ on the solution of (5).

B. Constrained Formulation

We would like to find the solution of the optimization problem

$$\hat{\mathbf{x}}(\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^{n}}{\operatorname{argmin}}\, \mathcal{J}(\mathbf{x}) \tag{9a}$$

with

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \tau\sum_{i=1}^{n}\phi([\mathrm{L}\{\mathbf{x}\}]_i) \tag{9b}$$

which reduces to the TV problem in (7) for the *potential function* $\phi(\,\cdot\,) = |\cdot|$. When the objective $\mathcal{J}(\,\cdot\,)$ is non-convex, it is difficult to find a global minimizer and we are satisfied with a local optimum.

Consider the shift-invariant transform $\mathbf{H}$ in (3), given that $\mathbf{W}$ now represents a single-level 1-D Haar transform. Note that a 1-D Haar representation of a 2-D signal can be computed by applying the transform either horizontally or vertically. Hence, for 1-D signals we have $K = 2$ and for 2-D images $K = 4$ (two transformations for each dimension of the gradient). It is then clear that each element of the discrete gradient can be uniquely mapped to the detail coefficients of $\mathbf{w} = \mathbf{H}\mathbf{x}$ by replacing each finite difference by some corresponding Haar coefficient.

Assume that coarse and detail Haar coefficients correspond to the elements w_i of $\mathbf{w}$ for $i \in [1 \dots nK/2]$ and $i \in [1 + nK/2 \dots nK]$, respectively. Furthermore, assume we store horizontal coefficients before vertical ones in the 2-D case. Then, the constrained optimization problem

$$\hat{\mathbf{w}}(\mathbf{y}) = \underset{\mathbf{w}\in\mathbb{R}^{nK}}{\operatorname{argmin}}\, \mathcal{F}(\mathbf{w}) \text{ s.t. } \mathbf{H}\mathbf{H}^\dagger\mathbf{w} = \mathbf{w} \tag{10a}$$

with

$$\mathcal{F}(\mathbf{w}) = \frac{1}{2K}\|\mathbf{w} - \mathbf{H}\mathbf{y}\|_2^2 + \tau\Phi(\mathbf{w}), \tag{10b}$$

is equivalent to (9a) when the regularizer $\Phi(\,\cdot\,)$ is given by

$$\Phi(\mathbf{w}) = \begin{cases} \displaystyle\sum_{i=1}^{n}\phi(\sqrt{2}|w_{n+i}|) & 1\text{−D} \\ \displaystyle\sum_{i=1}^{n}\phi\left(\sqrt{2}\sqrt{(w_{2n+i})^2 + (w_{3n+i})^2}\right) & 2\text{−D}. \end{cases}$$

The final estimate of the signal can be obtained simply by applying the dual $\mathbf{H}^\dagger$ as in (6). The role of the constraint $\mathbf{H}\mathbf{H}^\dagger\mathbf{w} = \mathbf{w}$ in (10a) is to enforce the solution to be within the space spanned by the columns of $\mathbf{H}$. This implies the equivalence of the problems (9a) and (10a) because, in the feasible region $\{\mathbf{w} : \mathbf{H}\mathbf{H}^\dagger\mathbf{w} = \mathbf{w}\}$, the cost functions $\mathcal{J}(\,\cdot\,)$ in (9b) and $\mathcal{F}(\,\cdot\,)$ in (10b) coincide with one another.

C. Augmented-Lagrangian Formulation

It would be difficult to handle directly constrained-optimization problems such as (10a). Consequently, we replace it with an equivalent unconstrained optimization problem. One traditional technique for doing this is the *augmented-Lagrangian method* [7]. To apply it, we replace the objective function $\mathcal{F}(\cdot)$ with the new penalty function

$$\mathcal{L}(\mathbf{w}, \mathbf{x}, \boldsymbol{\lambda}; \mu) = \mathcal{F}(\mathbf{w}) + \frac{\mu}{2}\|\mathbf{w} - \mathbf{H}\mathbf{x}\|_2^2 - \boldsymbol{\lambda}^T(\mathbf{w} - \mathbf{H}\mathbf{x}) \quad (11)$$

where $\mu \in \mathbb{R}_+$ is called the *penalty parameter* and $\boldsymbol{\lambda} \in \mathbb{R}^{nK}$ is the vector of Lagrange multipliers. The consistency condition $\mathbf{w} = \mathbf{H}\mathbf{x}$ asserted by the penalty function is equivalent to $\mathbf{w} = \mathbf{H}\mathbf{H}^\dagger\mathbf{w}$. The method works by iteratively minimizing $\mathcal{L}(\cdot)$ with respect to $(\mathbf{w}, \mathbf{x})$ while keeping μ fixed and updating $\boldsymbol{\lambda}$ according to a simple rule

$$\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t - \mu(\hat{\mathbf{w}}^{t+1} - \mathbf{H}\hat{\mathbf{x}}^{t+1}) \quad (12)$$

where $t > 0$ is the iteration number. To solve the minimization of the objective (11), we apply the *alternating-direction method of multipliers* (ADMM) [10] and alternate between solving the problem for $\mathbf{w}$ with $\mathbf{x}$ fixed and *vice versa*.

The solution with respect to $\mathbf{w}$ is obtained by rewriting the problem as

$$\begin{aligned}\hat{\mathbf{w}}^{t+1} &= \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{w}, \hat{\mathbf{x}}^t, \boldsymbol{\lambda}^t; \mu) \\ &= \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{w} - \tilde{\mathbf{w}}\|_2^2 + \tilde{\tau}\Phi(\mathbf{w}) \right\} \quad (13)\end{aligned}$$

where $\tilde{\mathbf{w}} = (\mathbf{H}\mathbf{y} + \mu K\mathbf{H}\hat{\mathbf{x}}^t + K\boldsymbol{\lambda}^t)/(1 + \mu K)$, $\tilde{\tau} = \tau K/(1 + \mu K)$, and $\Phi(\cdot)$ is the same as in (10b). This problem is solved efficiently in one step by applying n times the generalized shrinkage function

$$\Psi_\phi(\mathbf{y}; \tau) = \underset{\mathbf{x} \in \mathbb{R}^{K/2}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \tau\phi\left(\sqrt{2}\|\mathbf{x}\|_2\right) \right\}. \quad (14)$$

The range of the function $\Psi_\phi(\cdot)$ is $\mathbb{R}$ for 1-D signals and $\mathbb{R}^2$ for 2-D images. For the case of TV denoising, the shrinkage function admits the closed form

$$\Psi_{|\cdot|}(\mathbf{y}; \tau) = \max\left\{ \|\mathbf{y}\|_2 - \sqrt{2}\tau, 0 \right\} \frac{\mathbf{y}}{\|\mathbf{y}\|_2}. \quad (15)$$

In the general case, we propose to precompute the solution of (14) and store it in a lookup table.

The solution with respect to $\mathbf{x}$ is obtained as

$$\begin{aligned}\hat{\mathbf{x}}^{t+1} &= \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{L}(\hat{\mathbf{w}}^{t+1}, \mathbf{x}, \boldsymbol{\lambda}^t; \mu) \\ &= \mathbf{H}^\dagger\hat{\mathbf{w}}^{t+1} - \frac{\mathbf{H}^\dagger\boldsymbol{\lambda}^t}{\mu}. \quad (16)\end{aligned}$$

The resulting Algorithm 1 is conceptually similar to conventional cycle spinning: it denoises in the shifted wavelet bases and combines the solutions. However, we are constraining our solutions to be consistent across shifted bases. This ensures that our algorithm avoids the oscillations caused by changing wavelet bases that have been observed by various authors (cf. [4]). Moreover, our algorithm splits the minimization into n-scalar shrinkages, which results in a favorable algorithmic complexity of $O(n)$ per iteration.

Algorithm 1

input: $\mathbf{y}, \hat{\mathbf{x}}^0 \in \mathbb{R}^n, \tau, \mu \in \mathbb{R}_+$
set: $t = 0, \boldsymbol{\lambda}^0 = \mathbf{0}$
repeat
  $\hat{\mathbf{w}}^{t+1} = (\mathbf{H}\mathbf{y} + \mu K\mathbf{H}\hat{\mathbf{x}}^t + K\boldsymbol{\lambda}^t)/(1 + \mu K)$
  $\hat{\mathbf{w}}^{t+1} = \Psi_\phi(\hat{\mathbf{w}}^{t+1}; \tau K/(1 + \mu K))$
  $\hat{\mathbf{x}}^{t+1} = \mathbf{H}^\dagger(\hat{\mathbf{w}}^{t+1} - \boldsymbol{\lambda}^t/\mu)$
  $\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t - \mu(\hat{\mathbf{w}}^{t+1} - \mathbf{H}\hat{\mathbf{x}}^{t+1})$
  $t = t + 1$
until stopping criterion
return $\hat{\mathbf{x}}$

Note: The function $\Psi_\phi(\cdot)$ is applied only on the detail coefficients of $\hat{\mathbf{w}}$.

IV. Experimental Results

A. Model-Based Estimation

In a first experiment, we consider the problem of estimating the samples of sparse Lévy processes corrupted by noise. The defining property of Lévy processes [11] is that they have stationary and independent increments, which means that the discrete signal $\mathbf{x}$ can be decoupled by application of a finite-difference operator. We draw the increments of the signal $\mathbf{x}$ from a Cauchy distribution, which has recently been demonstrated to be highly compressible [12].

The *maximum a posteriori* (MAP) estimator for such processes can be easily obtained by choosing the potential function $\phi([\mathrm{L}\{\mathbf{x}\}]_i) = \log([\mathrm{L}\{\mathbf{x}\}]_i^2 + 1)$. Although the objective in this estimation problem is nonconvex, our algorithm, tailored with a lookup table for the generalized shrinkage function $\Psi(\cdot)$, can easily be applied. We compare denoising performance of the MAP estimator against TV regularization, conventional cycle spinning, wavelet shrinkage with full decomposition, and LMMSE estimator. Various noise levels are taken into account, and for each noise level, we perform 100 random realizations of the problem with signals of length 2048. For each realization, the regularization parameters of the methods are optimized for the best MSE performance. We let all of the iterative methods run for 200 iterations and obtained the results presented in Fig. 1.

It can be seen from Fig. 1 that the MAP estimator obtained via our algorithm outperforms the other standard estimators over the whole range of noise considered. Meanwhile, our experiment confirms the common observation that linear estimators (LMMSE) are poorly suited for sparse signals. Clearly, the performance of wavelet shrinkage via thresholding of the Haar coefficients is significantly improved by using conventional cycle spinning. Also, we note a very similar performance of TV and conventional cycle spinning for certain noise levels, with TV offering a better overall performance.

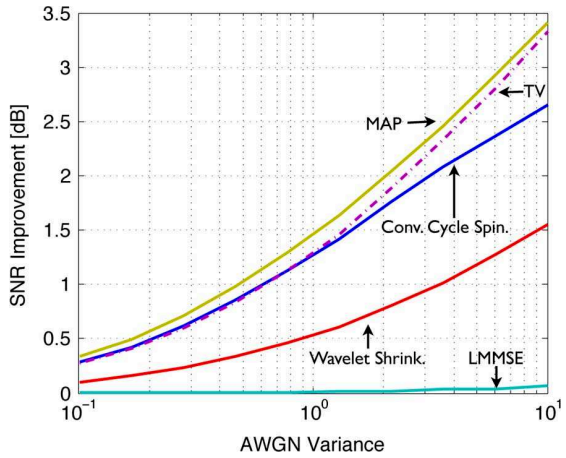

Fig. 1. Comparison of various sparse-estimation methods for α-stable Lévy processes. See text for a detailed explanation of the experiment.

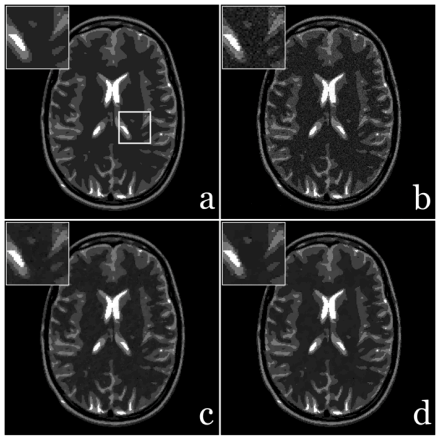

Fig. 2. Comparison of denoising results for the brain phantom: (a) original image; (b) noisy data, SNR $=$ 15.02 dB; (c) TV regularization, SNR $=$ 21.61 dB; (d) Log regularization, SNR $=$ 23.69 dB.

B. Image Denoising

It has been recently demonstrated that the recoverability of sparse signals can be significantly improved by extending the traditional ℓ_1-based reconstruction to nonconvex objectives [13], [14]. Following this philosophy, we consider the potential function $\phi([\mathrm{L}\{\mathbf{x}\}]_i) = \log([\mathrm{L}\{\mathbf{x}\}]_i^2 + \epsilon)$, where ϵ is a scalar parameter which controls the desired sparsity of the solution. We compare the SNR performance of log-regularized denoising and TV denoising, with $\epsilon = 10^{-4}$ and $\mu = 1/2$. Both methods are run for 500 iterations. Again, the regularization parameters are optimized via oracle for the highest-possible SNR. We compared the performance of our method, for TV denoising, with FISTA and observed that both algorithms are comparable in terms of speed[1]. In Fig. 2, we provide the results of this image-denoising experiment applied to the realistic analytical brain phantom [15]. It is a piecewise constant image, which implies that its gradient is sparse.

[1]See supplementary material for additional comparison plots available at http://people.epfl.ch/ulugbek.kamilov.

The results in Fig. 2 support the observation that log-penalized problems yield sparser solutions compared to ℓ_1 minimization. In the present example, the SNR is improved by more than 2 dB.

V. Conclusion

In this letter, we derived a simple denoising algorithm for signals and images. The algorithm exploits the link between two popular estimation methods: TV and wavelet regularization. We demonstrated that solving the TV problem can be achieved by constraining the solution set of Haar shrinkage with cycle spinning. We solved this constrained problem in an augmented-Lagrangian framework, which reduces the method into a sequence of scalar shrinkages. By precomputing the solution of the scalar subproblems, our algorithm can be extended to handle arbitrary potential functions, which makes it practically interesting for sparse estimation beyond ℓ_1.

References

[1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
[2] R. R. Coifman and D. L. Donoho, *Springer Lecture Notes in Statistics*. New York: Springer-Verlag, 1995, pp. 125–150, ch. Translation-invariant de-noising.
[3] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, 2003.
[4] M. Guerquin-Kern, M. Häberlin, K. P. Prüssmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1649–1660, 2011.
[5] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
[6] G. Steidl, J. Weickert, T. Brox, P. Mrazek, and M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs," *SIAM J. Numer. Anal.*, vol. 42, no. 2, pp. 686–713, 2004.
[7] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.
[8] A. Beck and M. Teboulle, "Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.
[9] S. Mallat, *A Wavelet Tool of Signal Processing: The Sparse Way*, 3rd ed. San Diego, CA: Academic, 2009.
[10] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programm.*, vol. 55, pp. 293–318, 1992.
[11] D. Applebaum, *Lévy Processes and Stochastic Calculus*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[12] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5193–5201, 2011.
[13] R. Chartrand, "Nonconvex compressive sensing and reconstruction of gradient-sparse images: Random vs. tomographic Fourier sampling," in *Proc. 15th IEEE Int. Conf. Image Processing (ICIP)*, Oct. 2008, pp. 2624–2627.
[14] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *J. Fourier Anal. Applicat.*, vol. 14, no. 5–6, pp. 877–905, 2007.
[15] M. Guerquin-Kern, L. Lejeune, K. P. Prüssmann, and M. Unser, "Realistic analytical phantoms for parallel magnetic resonance imaging," *IEEE Trans. Med. Imag.*, doi:10.1109/TMI.2011.2174158, to be published.