Artifact-free Wavelet Denoising: Non-convex Sparse Regularization, Convex Optimization

Yin Ding and Ivan W. Selesnick

***Abstract*—Algorithms for signal denoising that combine wavelet-domain sparsity and total variation (TV) regularization are relatively free of artifacts, such as pseudo-Gibbs oscillations, normally introduced by pure wavelet thresholding. This paper formulates wavelet-TV (WATV) denoising as a unified problem. To strongly induce wavelet sparsity, the proposed approach uses non-convex penalty functions. At the same time, in order to draw on the advantages of convex optimization (unique minimum, reliable algorithms, simplified regularization parameter selection), the non-convex penalties are chosen so as to ensure the convexity of the total objective function. A computationally efficient, fast converging algorithm is derived.**

I. Introduction

Simple thresholding of wavelet coefficients is a reasonably effective procedure for noise reduction when the signal of interest possesses a sparse wavelet representation. However, wavelet thresholding often introduces artifacts such as spurious noise spikes and pseudo-Gibbs oscillations around discontinuities [13]. In contrast, total variation (TV) denoising [34] does not introduce such artifacts; but it often produces undesirable staircase artifacts.

Roughly stated, noise spikes resulting from wavelet thresholding are due to noisy wavelet coefficients exceeding the threshold, whereas pseudo-Gibbs artifacts are due to non-zero coefficients being erroneously set to zero. An approach to alleviate pseudo-Gibbs artifacts is to estimate thresholded coefficients by variational means [5], [14]. The use of TV minimization as a way to recover thresholded coefficients is particularly effective [10], [11], [18], [19]. In this approach, thresholding is performed first. Those coefficients that survive thresholding (i.e., significant coefficients) are maintained. Those coefficients that do not survive (i.e., insignificant coefficients) are then optimized to minimize a TV objective function. TV regularization was subsequently used with wavelet packets [26], curvelets [7], complex wavelets [25], shearlets [21], tetrolets [23], [37], and generalized in [20]. Powerful proximal algorithms have been developed for signal restoration with general hybrid regularization (including wavelet-TV) allowing constraints and non-Gaussian noise [32].

In this letter, we propose a unified wavelet-TV (WATV) approach that estimates all wavelet coefficients (significant and insignificant) simultaneously via the minimization of a single objective function. To induce wavelet-domain sparsity, we employ non-convex penalties due to their strong sparsity-inducing properties [29]–[31]. In general practice, when non-convex penalties are used, the convexity of the objective function is usually sacrificed. However, in our approach, we restrict the non-convex penalty so as to ensure the strict convexity of the objective function; then, the minimizer is unique and can be reliably obtained through convex optimization. The proposed WATV denoising method is relatively resistant to pseudo-Gibbs oscillations and spurious noise spikes. We derive a computationally efficient, fast converging optimization algorithm for the proposed objective function.

The formulation of convex optimization problems for linear inverse problems, where a non-convex regularizer is designed in such a way that the total objective function is convex, was proposed by Blake and Zimmerman [4] and Nikolova [27], [28], [31]. The use of semidefinite programming (SDP) to attain such a regularizer has been considered [35]. Recently, the concept has been applied to group-sparse denoising [12], TV denoising [36], and non-convex fused-lasso [3].

II. Problem Formulation

We consider the estimation of a signal $x \in \mathbb{R}^N$ observed in additive white Gaussian noise (AWGN),

$$y_n = x_n + v_n, \quad n = 0, 1, \ldots, N-1. \tag{1}$$

We denote the wavelet transform by W and the wavelet coefficients of signal x as $w = Wx$. We index the coefficients as $w_{j,k}$ where j and k are the scale and time indices, respectively. In this work, we use the translational-invariant (i.e., undecimated) wavelet transform [13], [24], which satisfies the Parseval frame condition,

$$W^{\mathsf{T}} W = I. \tag{2}$$

But the proposed WATV denoising algorithm can be used with any transform W satisfying (2).

The total variation (TV) of signal $x \in \mathbb{R}^N$ is defined as $\mathrm{TV}(x) := \|Dx\|_1$ where D is the first-order difference matrix,

$$D = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}, \tag{3}$$

and $\|x\|_1$ denotes the ℓ_1 norm of x, i.e., $\|x\|_1 = \sum_n |x_n|$. We also denote $\|x\|_2^2 := \sum_n |x_n|^2$. For a set of doubly-indexed wavelet coefficients, we denote $\|w\|_2^2 := \sum_{j,k} |w_{j,k}|^2$.

The authors are with the Department of Electrical and Computer Engineering, School of Engineering, New York University, 6 Metrotech Center, Brooklyn, NY 11201. Email: yd372@nyu.edu and selesi@nyu.edu.

MATLAB software available at http://eeweb.poly.edu/iselesni/WATV

The proposed WATV denoising method finds wavelet coefficients w by solving the optimization problem,

$$\hat{w} = \arg\min_{w} \Big\{ F(w) = \frac{1}{2}\|Wy - w\|_2^2 + \sum_{j,k} \lambda_j \phi(w_{j,k}; a_j) + \beta \|DW^{\mathsf{T}} w\|_1 \Big\}. \quad (4)$$

The estimate of the signal x is then given by the inverse wavelet transform of $\hat{w}$, i.e., $\hat{x} = W^{\mathsf{T}}\hat{w}$. The penalty term $\|DW^{\mathsf{T}}w\|_1$ is the total variation of the signal estimate $\hat{x}$. The regularization parameters are $\lambda_j > 0$ and $\beta > 0$. We allow the wavelet regularization and penalty parameters (λ_j and a_j) to vary with scale j.

A. Parameterized Penalty Function

In the proposed approach, the function $\phi(\,\cdot\,; a)\colon \mathbb{R} \to \mathbb{R}$ is a non-convex sparsity-inducing penalty function with parameter $a \geqslant 0$. We assume ϕ satisfies the following conditions:

1) ϕ is continuous on $\mathbb{R}$
2) ϕ is twice continuously differentiable, increasing, and concave on $\mathbb{R}_+$
3) $\phi(x; 0) = |x|$
4) $\phi(0; a) = 0$
5) $\phi(-x; a) = \phi(x; a)$
6) $\phi'(0^+; a) = 1$
7) $\phi''(x; a) \geqslant -a$ for all $x \neq 0$

The following proposition indicates a suitable range of values for the parameter a. This proposition combines Propositions 1 and 2 of [12].

Proposition 1. *Suppose the parameterized penalty function ϕ satisfies the above-listed properties and $\lambda > 0$. If*

$$0 \leqslant a < \frac{1}{\lambda}, \quad (5)$$

then the function $f\colon \mathbb{R} \to \mathbb{R}$,

$$f(x) = \frac{1}{2}(y - x)^2 + \lambda\, \phi(x; a), \quad (6)$$

is strictly convex and the function $\theta\colon \mathbb{R} \to \mathbb{R}$,

$$\theta(y; \lambda, a) = \arg\min_{x \in \mathbb{R}} \Big\{ \frac{1}{2}(y - x)^2 + \lambda\, \phi(x; a) \Big\}, \quad (7)$$

is a continuous nonlinear threshold function with threshold value λ. That is, $\theta(y; \lambda, a) = 0$ for all $|y| < \lambda$.

Note that (7) is the definition of the proximity operator, or proximal mapping, of ϕ [2], [15].

Several common penalties satisfy the above-listed properties, for example, the logarithmic and arctangent penalties (when suitably normalized), both of which have been advocated for sparse regularization [8], [29]. We use the arctangent penalty as it induces sparsity more strongly [12], [35],

$$\phi(x; a) = \begin{cases} \frac{2}{a\sqrt{3}}\Big(\operatorname{atan}\Big(\frac{1+2a|x|}{\sqrt{3}}\Big) - \frac{\pi}{6}\Big), & a > 0 \\ |x|\,, & a = 0. \end{cases} \quad (8)$$

Figure 1 illustrates this penalty and its corresponding threshold function θ, which is given by the solution to a cubic polynomial; see Eqn. (21) in [35].

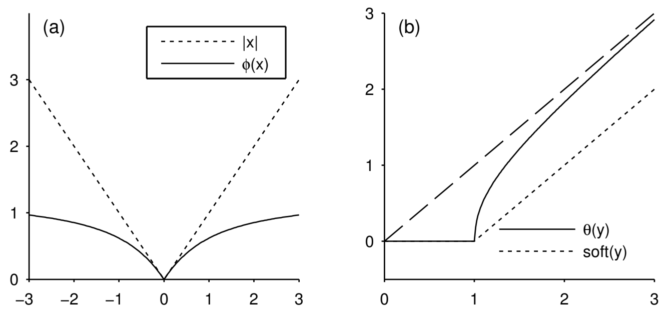

Fig. 1. (a) Non-convex penalty. (b) Threshold function ($\lambda = 1$, $a = 0.95$).

The parameter a (with $0 \leqslant a < 1/\lambda$) controls the non-convexity of the penalty ϕ; specifically, $\phi''(0^+; a) = -a$. This parameter also controls the shape of the threshold function θ; namely, the right-sided derivative at the threshold λ, is $\theta'(\lambda^+) = 1/(1 - a\lambda)$. If $a = 0$, then $\theta'(\lambda^+) = 1$ and θ is the soft-threshold function. If $a \approx 1/\lambda$ (and $a < 1/\lambda$), then θ is a continuous approximation of the hard-threshold function. But, if $a > 1/\lambda$, then f in (6) is not convex, the associated threshold function is not continuous, and the threshold value is not λ. As discussed in [35], the value $a = 1/\lambda$ is the critical value that maximizes the sparsity-inducing behavior of the penalty while ensuring convexity of f.

B. Convexity of the Objective Function

In the WATV denoising problem (4), we use the non-convex penalty ϕ to strongly promote sparsity, but we seek to ensure the strict convexity of the objective function F.

Proposition 2. *Suppose the parameterized penalty function ϕ satisfies the properties listed in Sec. II-A. The objective function* (4) *is strictly convex if*

$$0 \leqslant a_j < \frac{1}{\lambda_j} \quad (9)$$

for each scale j.

Proof. We write the objective function in (4) as

$$F(w) = \sum_{j,k} \Big\{ \frac{1}{2}([Wy]_{j,k} - w_{j,k})^2 + \lambda_j\, \phi(w_{j,k}; a_j) \Big\} + \beta\, \|DW^{\mathsf{T}}w\|_1. \quad (10)$$

By Proposition 1, condition (9) ensures each term in the curly braces is strictly convex. Also, the ℓ_1 norm term is convex. It follows that F, being the sum of strictly convex and convex functions, is strictly convex. □

When $\beta = 0$ in (4) and $a_j < 1/\lambda_j$ for all j, then $\hat{w}_{j,k} = \theta([Wy]_{j,k}; \lambda_j, a_j)$. That is, the solution is obtained by wavelet-domain thresholding.

C. Parameters

The proposed formulation (4) requires parameters λ_j, a_j, and β. We suggest using $a_j = 1/\lambda_j$ or slightly less (e.g.,

$0.95/\lambda_j$) to maximally induce sparsity while keeping F convex. To set λ_j and β, we consider separately the special cases of (*i*) $\beta = 0$ and (*ii*) $\lambda_j = 0$ for all j. In case (*i*), problem (4) reduces to pure wavelet denoising, for which we presume suitable parameters, denoted λ_j^{W}, are known. In case (*ii*), problem (4) reduces to pure TV denoising, for which we presume a suitable parameter, denoted β^{TV}, is known. We suggest setting

$$\lambda_j = \eta\, \lambda_j^{\mathsf{W}}, \quad \beta = (1-\eta)\, \beta^{\mathsf{TV}}, \tag{11}$$

where $0 < \eta < 1$ controls the relative weight of wavelet and TV regularization. We further suggest the smaller range of $0.9 < \eta < 1.0$, so that TV regularization makes a relatively minor adjustment of the wavelet solution, to reduce spurious noise spikes and pseudo-Gibbs' phenomena.

In this work, we set the wavelet thresholds $\lambda_j^{\mathsf{W}} = 2.5\sigma_j$, where σ_j^2 denotes the noise variance in wavelet scale j. For an undecimated wavelet transform satisfying (2), we have $\sigma_j = \sigma/2^{j/2}$ where σ^2 is the noise variance in the signal domain. For pure TV denoising, we use $\beta^{\mathsf{TV}} = \sqrt{N}\sigma/4$ in accordance with [17]. Therefore, we suggest using

$$\lambda_j = 2.5\, \eta\, \sigma/2^{j/2}, \quad \beta = (1-\eta)\sqrt{N}\sigma/4, \tag{12}$$

with a nominal value of $\eta = 0.95$. Additionally, we omit regularization of low-pass wavelet coefficients.

III. Optimization Algorithm

To solve the WATV denoising problem (4) we use the split augmented Lagrangian shrinkage algorithm (SALSA) [1] which in turn is based on the alternating direction method of multipliers (ADMM) [6], [22]. Proximal methods (e.g., Douglas-Rachford) could also be used [15]. We assume that a_j satisfies (9) for each j, so that F is strictly convex.

With variable splitting, problem (4) is expressed as a constrained problem,

$$\begin{aligned} &\arg\min_{u,w}\ g_1(w) + g_2(u) \\ &\text{subject to } u = w, \end{aligned} \tag{13}$$

where

$$g_1(w) = \frac{1}{2}\|Wy - w\|_2^2 + \sum_{j,k} \lambda_j\, \phi(w_{j,k}; a_j) \tag{14a}$$

$$g_2(u) = \beta \|DW^{\mathsf{T}} u\|_1. \tag{14b}$$

Both g_1 and g_2 are strictly convex, hence we may utilize the convergence theory of [22]. Even though ϕ is non-convex, the function g_1 is strictly convex because a_j satisfies (9).

The augmented Lagrangian is given by

$$L(w, u, \mu) = g_1(w) + g_2(u) + \frac{\mu}{2}\|u - w - d\|_2^2. \tag{15}$$

The solution of (13) can be found by iteratively minimizing with respect to w and u alternately, as proven in [22]. Thereby, we obtain an iterative algorithm to solve (4), where each iteration consists of three steps:

$$w = \arg\min_w \big\{ g_1(w) + \frac{\mu}{2}\|u - d - w\|_2^2 \big\}, \tag{16a}$$

$$u = \arg\min_u \big\{ g_2(u) + \frac{\mu}{2}\|u - d - w\|_2^2 \big\}, \tag{16b}$$

$$d = d - (u - w), \tag{16c}$$

for some $\mu > 0$. We initialize $u = Wy$ and $d = 0$. Note that both (16a) and (16b) are strictly convex. (Sub-problem (16a) is strictly convex because a_j satisfies (9).)

Sub-problem (16a) can be solved exactly and explicitly. Combining the quadratic terms, (16a) may be expressed as

$$w = \arg\min_w \sum_{j,k} \left\{ \frac{1}{2}(p_{j,k} - w_{j,k})^2 + \frac{\lambda_j}{\mu+1}\phi(w_{j,k}; a_j) \right\},$$

where $p = (Wy + \mu(u-d))/(\mu+1)$. Since this problem is separable, it can be further written as

$$w_{j,k} = \arg\min_{v \in \mathbb{R}} \left\{ \frac{1}{2}(p_{j,k} - v)^2 + \frac{\lambda_j}{\mu+1}\phi(v; a_j) \right\}, \tag{17}$$

for each j and k. This is the same form as (6). Consequently, the solution to sub-problem (16a) is given by

$$w_{j,k} = \theta\left(p_{j,k}; \frac{\lambda_j}{\mu+1}, a_j \right) \tag{18}$$

where θ is the threshold function (7).

Sub-problem (16b) can also be solved exactly. We use properties of proximity operators to express (16b) in terms of a TV denoising problem [15]. In turn, the TV denoising problem can be solved exactly by a fast finite-time algorithm [16]. Define function q as

$$\begin{aligned} q(v) &= \arg\min_u \left\{ \beta\|DW^{\mathsf{T}}u\|_1 + \frac{\mu}{2}\|u - v\|_2^2 \right\} && (19)\\ &= \arg\min_u \left\{ \psi(W^{\mathsf{T}}u) + \frac{1}{2}\|u - v\|_2^2 \right\} && (20)\\ &= \arg\min_u \left\{ h(u) + \frac{1}{2}\|u - v\|_2^2 \right\} && (21)\\ &= \operatorname{prox}_h(v) && (22)\\ &= v + W\big(\operatorname{prox}_\psi(W^{\mathsf{T}}v) - W^{\mathsf{T}}v\big) && (23) \end{aligned}$$

where

$$\psi(x) = \frac{\beta}{\mu}\|Dx\|_1, \quad h(u) = \psi(W^{\mathsf{T}}u), \tag{24}$$

and where we have used the 'semi-orthogonal linear transform' property of proximity operators to obtain (23) from (22), which depends on (2). See Table 1.1 and Example 1.7.4 in [15], and see [9], [33] for details. Note that

$$\begin{aligned} \operatorname{prox}_\psi(W^{\mathsf{T}}v) &= \arg\min_x \left\{ \psi(x) + \frac{1}{2}\|x - W^{\mathsf{T}}v\|_2^2 \right\} && (25)\\ &= \arg\min_x \left\{ \frac{\beta}{\mu}\|Dx\|_1 + \frac{1}{2}\|x - W^{\mathsf{T}}v\|_2^2 \right\} \\ &= \operatorname{tvd}\big(W^{\mathsf{T}}v,\ \beta/\mu\big) && (26) \end{aligned}$$

is total variation denoising. We use the fast finite-time exact software of [16]. Hence, (16b) can be implemented as

$$v = d + w \tag{27}$$

$$u = v + W\big(\operatorname{tvd}(W^{\mathsf{T}}v,\ \beta/\mu) - W^{\mathsf{T}}v\big). \tag{28}$$

Hence, solving problem (4) by (16a)-(16c), consists of iteratively thresholding (18) [to solve (16a)] and performing total variation denoising (28) [to solve (16b)]. Table I summarizes the whole algorithm. Each iteration involves one wavelet transform, one inverse wavelet transform, and one total variation denoising problem. (The term Wy in the algorithm

TABLE I
ITERATIVE ALGORITHM FOR WATV DENOISING (4).

Input: $y,\ \lambda_j,\ \beta,\ \mu$
Initialization: $u = Wy,\ d = 0$
$a_j = 1/\lambda_j$
Repeat:
$\quad p = (Wy + \mu(u - d))/(\mu + 1)$
$\quad w_{j,k} = \theta(p_{j,k}; \lambda_j/(\mu+1),\ a_j)$ for all j, k
$\quad v = d + w$
$\quad u = v + W\big(\mathrm{tvd}(W^{\mathsf{T}} v,\ \beta/\mu) - W^{\mathsf{T}} v\big)$
$\quad d = d - (u - w)$
Until convergence
Return: $x = W^{\mathsf{T}} w$

TABLE II
AVERAGE RMSE FOR DENOISING EXAMPLE

σ	Thresholding	Two-step WATV	Proposed WATV
1.0	0.44	0.39	0.37
2.0	0.81	0.69	0.67
4.0	1.54	1.44	1.28
8.0	2.90	2.75	2.46
16.0	5.25	4.77	4.19

may be precomputed.) Since both minimization problems, (16a) and (16b), are solved exactly, the algorithm is guaranteed to converge to the unique global minimizer according to the convergence theory of [22].

IV. EXAMPLE

We illustrate the proposed WATV denoising algorithm using noisy data of length 1024 (Fig. 2a), generated by Wavelab (http://www-stat.stanford.edu/%7Ewavelab/), with AWGN of $\sigma = 4.0$. We use a 5-scale undecimated wavelet transform with two vanishing moments. Hard-thresholding (with $\lambda_j^{\mathsf{W}} = 2.5\sigma_j$) results in noise spikes due to wavelet coefficients exceeding the threshold (Fig. 2b). Two-step WATV denoising [19] (wavelet hard-thresholding followed by TV optimization of coefficients thresholded to zero) does not correct noise spikes due to noisy coefficients that exceed the threshold, so we use higher thresholds of $\sigma_j\sqrt{2\log N}$ to avoid noise spikes from occurring in the first step. The result (Fig. 2c) is relatively free of noise spikes and psuedo-Gibbs oscillations.

The proposed WATV algorithm [with parameters (12), $\eta = 0.95$, and $a_j = 1/\lambda_j$] is resistant to spurious noise spikes, preserves sharp discontinuities, and results in a lower root-mean-square-error (RMSE) (Fig. 2d). The proposed algorithm was run for 20 iterations in a time of 0.15 seconds using Matlab 7.14 on a MacBook Air (1.3 GHz CPU).

Table II shows the RMSE for several values of the noise standard deviation (σ) for each method. Each RMSE value is obtained by averaging 20 independent realizations. The higher RMSE of two-step WATV appears to be due to distortion induced by the greater threshold value.

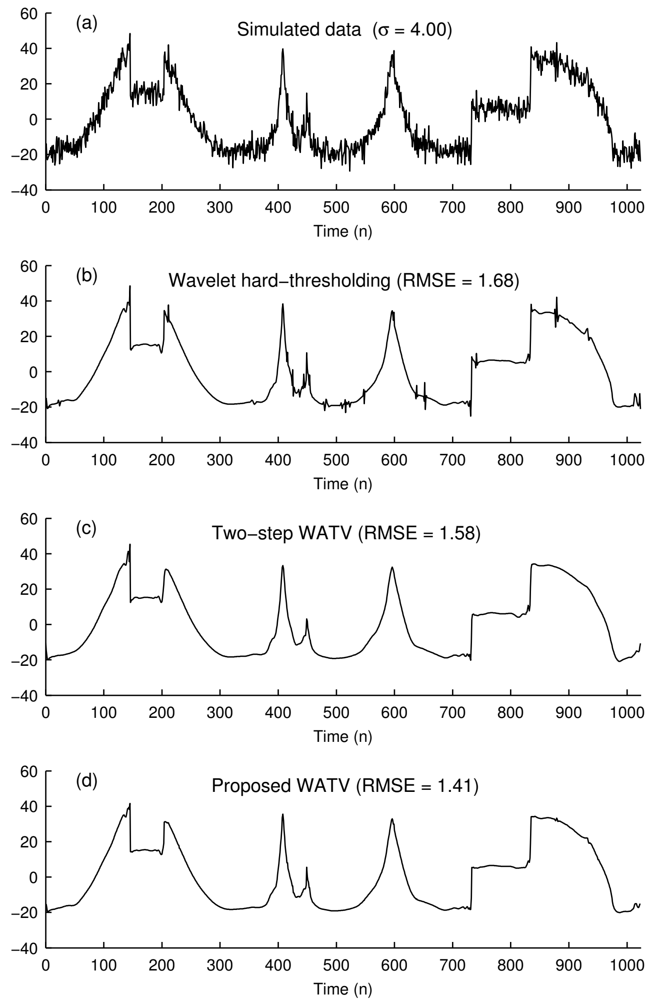

Fig. 2. (a) Simulated data. (b) Wavelet denoising (hard-thresholding). (c) Two-step WATV denoising [19]. (d) Proposed WATV denoising.

V. CONCLUSION

Inspired by [18], [19], this paper describes a wavelet-TV (WATV) denoising method that is relatively free of artifacts (pseudo-Gibbs oscillations and noise spikes) normally introduced by simple wavelet thresholding. The method is formulated as an optimization problem incorporating both wavelet sparsity and TV regularization. To strongly induce wavelet sparsity, the approach uses parameterized non-convex penalty functions. Proposition 2 gives the critical parameter value to maximally promote wavelet sparsity while ensuring convexity of the objective function as a whole. A fast iterative implementation is derived. We expect the approach to be useful with transforms other than the wavelet transform, provided they satisfy the Parseval frame property (2).

ACKNOWLEDGMENT

The authors gratefully acknowledge constructive comments by anonymous reviewers. In particular, a reviewer noted that (16b) can be expressed explicitly in terms of total variation denoising, an improvement upon the inexact solution in the original manuscript.

References

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.*, 19(9):2345–2356, September 2010.

[2] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.

[3] I. Bayram, P.-Y. Chen, and I. Selesnick. Fused lasso with a non-convex sparsity inducing penalty. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, May 2014.

[4] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[5] Y. Bobichon and A. Bijaoui. Regularized multiresolution methods for astronomical image enhancement. In *Proc. SPIE*, volume 3169 (Wavelet Applications in Signal and Image Processing V), pages 36–45, October 1997.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[7] E. J. Candés and F. Guo. New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction. *Signal Processing*, 82(11):1519–1543, 2002.

[8] E. J. Candès, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted l1 minimization. *J. Fourier Anal. Appl.*, 14(5):877–905, December 2008.

[9] L. Chaâri, N. Pustelnik, C. Chaux, and J.-C. Pesquet. Solving inverse problems with overcomplete transforms and convex optimization techniques. In *Proceedings of SPIE*, volume 7446 (Wavelets XIII), August 2009.

[10] T. F. Chan and H. M. Zhou. Total variation improved wavelet thresholding in image compression. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, volume 2, pages 391–394, September 2000.

[11] T. F. Chan and H.-M. Zhou. Total variation wavelet thresholding. *J. of Scientific Computing*, 32(2):315–341, August 2007.

[12] P.-Y. Chen and I. W. Selesnick. Group-sparse signal denoising: Non-convex regularization, convex optimization. *IEEE Trans. Signal Process.*, 62(13):3464–3478, July 2014.

[13] R. R. Coifman and D. L. Donoho. Translation-invariant de-noising. In A. Antoniadis and G. Oppenheim, editors, *Wavelet and statistics*, pages 125–150. Springer-Verlag, 1995.

[14] R. R. Coifman and A. Sowa. Combining the calculus of variations and wavelets for image enhancement. *J. of Appl. and Comp. Harm. Analysis*, 9(1):1–18, 2000.

[15] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke et al., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, 2011.

[16] L. Condat. A direct algorithm for 1-D total variation denoising. *IEEE Signal Processing Letters*, 20(11):1054–1057, November 2013.

[17] L. Dümbgen and A. Kovac. Extensions of smoothing via taut strings. *Electron. J. Statist.*, 3:41–75, 2009.

[18] S. Durand and J. Froment. Artifact free signal denoising with wavelets. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2001.

[19] S. Durand and J. Froment. Reconstruction of wavelet coefficients using total variation minimization. *SIAM J. Sci. Comput.*, 24(5):1754–1767, 2003.

[20] S. Durand and M. Nikolova. Denoising of frame coefficients using ℓ^1 data-fidelity term and edge-preserving regularization. *Multiscale Modeling & Simulation*, 6(2):547–576, 2007.

[21] G. R. Easley, D. Labate, and F. Colonna. Shearlet-based total variation diffusion for denoising. *IEEE Trans. Image Process.*, 18(2):260–268, February 2009.

[22] J. Eckstein and D. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 5:293–318, 1992.

[23] J. Krommweh and J. Ma. Tetrolet shrinkage with anisotropic total variation minimization for image approximation. *Signal Processing*, 90(8):2529–2539, 2010.

[24] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, Jr. Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Processing Letters*, 3(1):10–12, January 1996.

[25] J. Ma. Towards artifact-free characterization of surface topography using complex wavelets and total variation minimization. *Applied Mathematics and Computation*, 170(2):1014–1030, 2005.

[26] F. Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE Trans. Image Process.*, 11(12):1450–1456, December 2002.

[27] M. Nikolova. Estimation of binary images by minimizing convex criteria. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 108–112 vol. 2, 1998.

[28] M. Nikolova. Markovian reconstruction using a GNC approach. *IEEE Trans. Image Process.*, 8(9):1204–1220, 1999.

[29] M. Nikolova. Energy minimization methods. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, chapter 5, pages 138–186. Springer, 2011.

[30] M. Nikolova, M. K. Ng, and C. Tam. On ℓ_1 data fitting and concave regularization for image recovery. *SIAM J. Sci. Comput.*, 35(1):A397–A430, 2013.

[31] M. Nikolova, M. K. Ng, and C.-P. Tam. Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *IEEE Trans. Image Process.*, 19(12):3073–3088, December 2010.

[32] N. Pustelnik, C. Chaux, and J. Pesquet. Parallel proximal algorithm for image restoration using hybrid regularization. *IEEE Trans. Image Process.*, 20(9):2450–2462, September 2011.

[33] N. Pustelnik, J.-C. Pesquet, and C. Chaux. Relaxing tight frame condition in parallel proximal methods for signal restoration. *IEEE Trans. Signal Process.*, 60(2):968–973, February 2012.

[34] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[35] I. W. Selesnick and I. Bayram. Sparse signal estimation by maximally sparse convex optimization. *IEEE Trans. Signal Process.*, 62(5):1078–1092, March 2014.

[36] I. W. Selesnick, A. Parekh, and I. Bayram. Convex 1-D total variation denoising with non-convex regularization. *IEEE Signal Processing Letters*, 22(2):141–144, February 2015.

[37] L. Wang, L. Xiao, J. Zhang, and Z. Wei. New image restoration method associated with tetrolets shrinkage and weighted anisotropic total variation. *Signal Processing*, 93(4):661–670, 2013.