

# Convex 1-D Total Variation Denoising with Non-convex Regularization

Ivan W. Selesnick, Ankit Parekh, and İlker Bayram

**Abstract**—Total variation (TV) denoising is an effective noise suppression method when the derivative of the underlying signal is known to be sparse. TV denoising is defined in terms of a convex optimization problem involving a quadratic data fidelity term and a convex regularization term. A non-convex regularizer can promote sparsity more strongly, but generally leads to a non-convex optimization problem with non-optimal local minima. This letter proposes the use of a non-convex regularizer constrained so that the total objective function to be minimized maintains its convexity. Conditions for a non-convex regularizer are given that ensure the total TV denoising objective function is convex. An efficient algorithm is given for the resulting problem.

## I. INTRODUCTION

Total variation (TV) is a widely used regularizer in sparse signal and image processing [6], [20]; especially when it is known the signal to be recovered has a sparse derivative (or sparse gradients), i.e., when the signal is piecewise constant (PWC). One-dimensional signals of this form arise, for example, in geoscience, astrophysics, and biophysics [11].

TV denoising is defined in terms of a convex optimization problem involving a quadratic data fidelity term and a convex regularization term. Interestingly, for 1-D TV denoising, the exact solution can be obtained using very fast direct algorithms that terminate in a finite number of steps [5], [7], [12].

In this letter, we consider a modification of the 1-D TV denoising problem where the non-smooth convex regularizer is replaced by a non-convex one. This modification is motivated by the fact that non-convex regularizers can better recover flat signal regions [4], [15]–[17]. The mathematical properties of the solutions to non-convex regularized signal restoration problems are discussed by Nikolova [15]–[17].

Generally, the use of a non-convex regularizer (as opposed to a convex one) leads to the formulation of the signal recovery problem as a non-convex optimization problem. In turn, spurious (non-optimal) local minima exist in which iterative optimization algorithms may become entrapped. In addition, the solution to a non-convex problem can be highly sensitive to small changes in the data: an infinitesimal change in the data may lead to a large change in the output, as is the case with the hard threshold function. This sensitivity also complicates the selection of an appropriate value of the regularization

parameter. Therefore, we consider the question of how to maintain the convexity of the TV denoising objective function when the regularizer is taken to be non-convex.

This letter provides a condition on a non-convex regularizer for 1-D TV denoising that ensures the total objective function (comprising data fidelity and regularization terms) is strictly convex. A fast iterative algorithm is described to perform convex TV denoising with a non-convex regularizer set accordingly. Results of the proposed algorithm are compared with standard 1-D TV denoising on simulated data.

The idea of specifying non-convex penalties in the formulation of convex optimization problems for linear inverse problems was proposed by Blake and Zimmerman [2] and by Nikolova [13], [14], [17]. This approach has recently been considered in [22] where the convexity condition is cast as a semidefinite program (SDP), in [1] which considers a non-convex extension of fused-lasso, and in [3] which addresses translation-invariant group-sparse denoising.

## II. PROBLEM FORMULATION

Let  $y \in \mathbb{R}^N$  be a piecewise constant signal observed in additive noise. Consider the objective function  $F: \mathbb{R}^N \rightarrow \mathbb{R}$ ,

$$F(x) = \frac{1}{2} \|y - x\|_2^2 + \lambda \sum_n \phi([Dx]_n) \quad (1)$$

where  $\lambda > 0$  is a regularization parameter,  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  is a sparsity promoting penalty function (regularizer), and  $D$  is the  $(N - 1) \times N$  matrix

$$D = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & & \ddots & \\ & & & & \ddots & \\ & & & & & -1 & 1 \end{bmatrix}. \quad (2)$$

The notation  $[Dx]_n$  represents component  $n$  of vector  $Dx$ .

One-dimensional total variation (TV) denoising is defined as minimizing  $F$  with respect to  $x \in \mathbb{R}^N$ ,

$$x^* = \arg \min_{x \in \mathbb{R}^N} F(x) \quad (3)$$

where  $\phi$  is taken to be  $\phi(x) = |x|$ . In this case,  $F$  is strictly convex on  $\mathbb{R}^N$  and hence its minimizer is unique. However, it has been shown in the literature that non-convex penalties have advantages in comparison with convex penalties, in terms of more accurately recovering signals with flat regions [15]–[17].

Here, we consider how to set a non-convex penalty function  $\phi$  to promote sparsity of  $Dx$  while keeping  $F$  strictly convex. Then, the minimizer will be unique, the denoising process

I. Selesnick and A. Parekh are with the Dept. of Electrical and Computer Engineering, New York University, 6 Metrotech Center, Brooklyn, NY 11201. İ Bayram is with Istanbul Technical University, Istanbul, Turkey. MATLAB software is available at [http://eeweb.poly.edu/iselesni/TVD\\_nonconvex/](http://eeweb.poly.edu/iselesni/TVD_nonconvex/)

This research was supported by the NSF under Grant No. CCF-1018020. Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

will be continuous/stable (i.e., infinitesimal changes in  $y$  do not produce large changes in  $x^*$ ), and convex optimization techniques can be used to reliably obtain the minimizer.

#### A. Non-convex penalty functions

We assume  $\phi$  is continuous, symmetric, twice differentiable on  $\mathbb{R} \setminus \{0\}$ , increasing on  $\mathbb{R}_+$ , and concave on  $\mathbb{R}_+$ . Examples of such  $\phi$  are the logarithmic penalty

$$\phi(x; a) = \frac{1}{a} \log(1 + a|x|), \quad a > 0 \quad (4)$$

and the arctangent penalty [22]

$$\phi(x; a) = \frac{2}{a\sqrt{3}} \left( \tan^{-1} \left( \frac{1 + 2a|x|}{\sqrt{3}} \right) - \frac{\pi}{6} \right), \quad a > 0. \quad (5)$$

For later, we note for both penalties, that

$$\inf_{x \neq 0} \phi''(x; a) = \phi''(0^+; a) = -a. \quad (6)$$

### III. CONVEXITY CONDITION

To find a condition on  $\phi$  ensuring  $F$  in (1) is strictly convex, we write  $F$  as

$$F(x) = F_0(x) + F_1(x) \quad (7)$$

where

$$F_0(x) = \frac{1}{4} \|y - x\|_2^2 + \lambda \sum_{n \text{ even}} \phi([Dx]_n) \quad (8)$$

$$F_1(x) = \frac{1}{4} \|y - x\|_2^2 + \lambda \sum_{n \text{ odd}} \phi([Dx]_n). \quad (9)$$

Note that if both  $F_0$  and  $F_1$  are strictly convex, then  $F$  is strictly convex. Hence, it suffices to find  $\phi$  such that  $F_0$  and  $F_1$  are strictly convex. Due to the similarity of  $F_1$  and  $F_0$ , it suffices to find  $\phi$  such that  $F_0$  is strictly convex.

We write  $F_0$  as

$$\begin{aligned} F_0(x) &= \sum_{n \text{ even}} \left( \frac{1}{4} (y_n - x_n)^2 + \frac{1}{4} (y_{n+1} - x_{n+1})^2 \right. \\ &\quad \left. + \lambda \phi(x_{n+1} - x_n) \right) \\ &= \left( \sum_{n \text{ even}} \frac{1}{4} x_n^2 + \frac{1}{4} x_{n+1}^2 + \lambda \phi(x_{n+1} - x_n) \right) + L(x) \end{aligned}$$

where  $L(x)$  is linear in  $x$ . We write

$$F_0(x) = \left( \sum_{n \text{ even}} B(x_n, x_{n+1}) \right) + L(x) \quad (10)$$

where  $B: \mathbb{R}^2 \rightarrow \mathbb{R}$  is defined as

$$B(u_0, u_1) = \frac{1}{4} u_0^2 + \frac{1}{4} u_1^2 + \lambda \phi(u_1 - u_0). \quad (11)$$

For the purpose of analyzing the convexity of  $F_0$ , the linear function  $L$  can be disregarded. Hence, if  $B$  is strictly convex on  $\mathbb{R}^2$ , then  $F_0$ , being a sum of strictly convex functions, will be strictly convex on  $\mathbb{R}^N$ .

To find a condition on  $\phi$  so as to ensure  $B$  is strictly convex, we define  $H: \mathbb{R}^2 \rightarrow \mathbb{R}$  as

$$H(v_0, v_1) = B\left(\frac{v_1 - v_0}{2}, \frac{v_1 + v_0}{2}\right). \quad (12)$$

Then  $B$  is strictly convex if and only if  $H$  is strictly convex. From (11), we write

$$H(v_0, v_1) = \frac{1}{4} \left( \frac{v_1 - v_0}{2} \right)^2 + \frac{1}{4} \left( \frac{v_1 + v_0}{2} \right)^2 + \lambda \phi(v_0) \quad (13)$$

$$= \frac{1}{8} v_0^2 + \frac{1}{8} v_1^2 + \lambda \phi(v_0) + L_2(v_0, v_1) \quad (14)$$

where  $L_2(v_0, v_1)$  is linear in  $(v_0, v_1)$ . Since we are concerned with the convexity of  $H$ , the linear function  $L_2$  can be disregarded, as above.

The function  $H$  is strictly convex if and only if  $\psi: \mathbb{R} \rightarrow \mathbb{R}$ , defined as  $\psi(x) = x^2/8 + \lambda \phi(x)$ , is strictly convex. By Theorem 6.4 of [10],  $\psi$  is strictly convex if it has a strictly increasing right-derivative. By assumption,  $\phi$  is twice differentiable on  $\mathbb{R} \setminus \{0\}$ ; hence,  $H$  is strictly convex if and only if

$$\phi'(0^+) \geq \phi'(0^-) \quad (15)$$

and

$$\frac{1}{4} + \lambda \phi''(x) > 0, \quad \text{for all } x \neq 0. \quad (16)$$

In fact, from the assumptions on  $\phi$  as stated in Sec. II-A, condition (15) already follows. (If  $\phi$  is increasing on  $\mathbb{R}_+$  and twice differentiable on  $\mathbb{R} \setminus \{0\}$ , then  $\phi'(0^+) \geq 0$ . If also  $\phi$  is symmetric, then  $\phi'(0^-) = -\phi'(0^+) \leq 0$ .)

Condition (16) is the positivity of the second derivative of  $H(v_0, v_1)$  with respect to  $v_0$ . We write the condition as

$$\inf_{x \neq 0} \phi''(x) > -\frac{1}{4\lambda}, \quad (17)$$

which constitutes a constraint on the non-convexity of  $\phi$ . For various standard penalties, including (4) and (5), the second derivative  $\phi''(x)$  is most negative at  $x = 0^+$ . For such penalties, condition (17) can be written

$$\phi''(0^+) > -\frac{1}{4\lambda}. \quad (18)$$

For the logarithmic and arctangent penalties, (4) and (5), parameterized by  $a$ , we use (6) to obtain the condition

$$a < \frac{1}{4\lambda}. \quad (19)$$

### IV. OPTIMALITY CONDITION

If  $F$  is strictly convex, then  $x^* \in \mathbb{R}^N$  is its unique minimizer if and only if  $0 \in \partial F(x^*)$  where  $\partial F$  is the subgradient of  $F$ . If  $\phi$  is such that  $F$  in (1) is convex, then  $\partial F$  is given by

$$\partial F(x) = \{x - y + \lambda D^T u : u_n \in d_\phi([Dx]_n), u \in \mathbb{R}^{N-1}\},$$

where  $d_\phi$  is a set-valued function on  $\mathbb{R}$ , defined as

$$d_\phi(u) = \begin{cases} \{\phi'(u)\}, & u \neq 0 \\ [-1, 1], & u = 0. \end{cases} \quad (20)$$

Hence, the condition  $0 \in \partial F(x^*)$  can be written as

$$\frac{1}{\lambda} (y - x^*) \in \{D^T u : u_n \in d_\phi([Dx^*]_n), u \in \mathbb{R}^{N-1}\}. \quad (21)$$

Let  $S$  be a matrix of size  $(N - 1) \times N$  such that  $SD^T = I$ . It can be taken to be the discrete anti-derivative (cumulative summation operator), defined by

$$[Sx]_n = \sum_{k \leq n} x_k. \quad (22)$$

Then it follows from (21) that

$$\frac{1}{\lambda}[S(y - x^*)]_n \in d_\phi([Dx^*]_n). \quad (23)$$

Condition (23) can be used to validate the optimality of a candidate  $x$  and to gauge the convergence of an algorithm minimizing  $F$ . The condition (23) implies that the points

$$z_n = ([Dx^*]_n, [S(y - x^*)]_n/\lambda) \in \mathbb{R}^2 \quad (24)$$

lie on the graph of  $d_\phi$ , as illustrated in Fig. 2 below.

## V. ALGORITHM

Following the procedure in [8], we use the majorization-minimization (MM) approach to derive a fast-converging algorithm. A majorizer of  $\phi$  is given by  $g: \mathbb{R} \times \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ ,

$$g(x; v) = \frac{\phi'(v)}{2} \left( \frac{x^2}{v} - v \right) + \phi(v). \quad (25)$$

(See, for example, Fig. 11 in [21].) Hence, a majorizer of  $F$  in (1) is given by

$$G(x; v) = \frac{1}{2} \|y - x\|_2^2 + \lambda \sum_n g([Dx]_n, [Dv]_n) \quad (26)$$

$$= \frac{1}{2} \|y - x\|_2^2 + \frac{\lambda}{2} x^T D^T [W(Dv)] Dx + c(v) \quad (27)$$

where  $W$  is a diagonal matrix,

$$[W(Dv)]_{n,n} = \frac{\phi'([Dv]_n)}{[Dv]_n} \quad (28)$$

and  $c(v)$  does not depend on  $x$ .

Using (27) in the MM update iteration,

$$x^{(k+1)} = \arg \min_x G(x; x^{(k)}), \quad (29)$$

leads to the iteration

$$x^{(k+1)} = (I + \lambda D^T [W(Dx^{(k)})] D)^{-1} y. \quad (30)$$

As noted in [8], as the algorithm converges to a solution for which  $Dx$  is sparse, elements of  $W$  go to infinity. To avoid the numerical problems related to this, as in [8], we use the matrix inverse lemma to write

$$(I + \lambda D^T W D)^{-1} = I - D^T \left( \frac{1}{\lambda} W^{-1} + D D^T \right)^{-1} D. \quad (31)$$

The iteration (30) can then be written as

$$x^{(k+1)} = y - D^T \left( \frac{1}{\lambda} [W(Dx^{(k)})]^{-1} + D D^T \right)^{-1} D y. \quad (32)$$

We initialize the iteration with  $x^{(0)} = y$ . Note that the system matrix in (32) is tridiagonal; hence, the iteration can be implemented with very high computational efficiency using a fast solver [19, Sect 2.4]. Due to MM, each iteration monotonically decreases the cost function value. We have found that 20 iterations of (32) are usually sufficient.

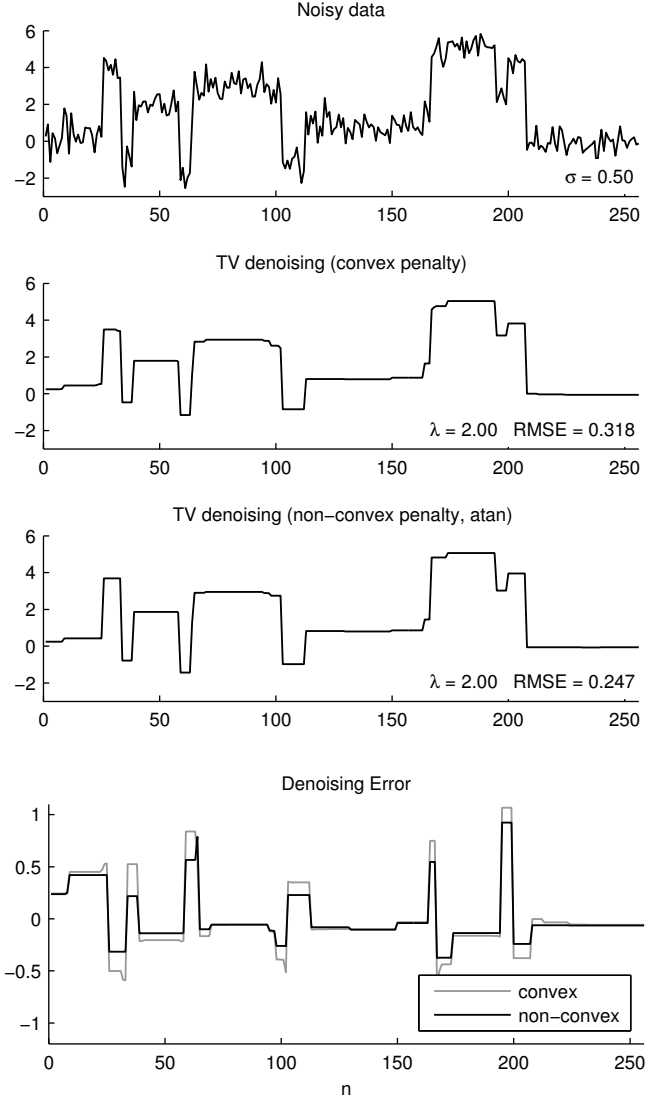


Fig. 1. Total variation denoising with convex and non-convex penalties.

Note that  $g(x; v)$  in (25) is undefined for  $v = 0$ . Hence if  $[Dx^{(k)}]_n = 0$  for some iteration  $k$  and some index  $n$ , then the majorizer is undefined. This manifests itself as a ‘division-by-zero’ error in (28). Due to the use of the matrix inverse lemma, this becomes a ‘multiplication-by-zero’ and causes no numerical problem in the algorithm (32). However, it complicates the proof of convergence of the algorithm. We do not prove its convergence. We remark (i) in practice, convergence is not adversely affected by this issue, (ii) optimality can be verified using (23), and (iii) this issue has been discussed in the literature [8], [9], [18] where it was found not to impede the convergence of affected algorithms in practice.

## VI. EXAMPLES

### A. Example 1

Total variation denoising with convex and non-convex regularization is illustrated in Fig. 1. The noisy data is obtained using additive white Gaussian noise (AWGN) ( $\sigma = 0.5$ ) on a PWC signal,  $s \in \mathbb{R}^N$ , of length  $N = 256$  (‘blocks’ generated by the Wavelab function, `MakeSignal`). For both convex and

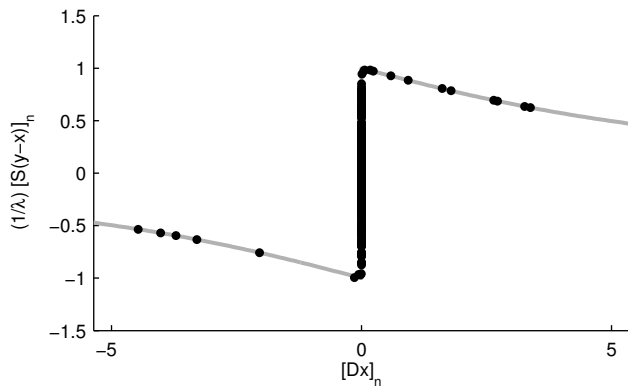


Fig. 2. Optimality condition (23) for Example 1.

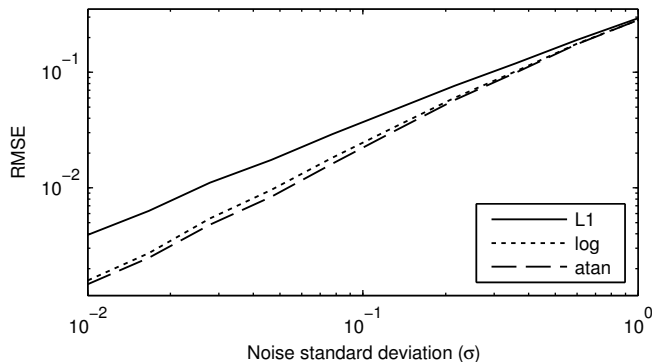


Fig. 3. Example 2. RMSE as a function of noise level for randomly generated PWC signals of length 1000. Non-convex penalties yield a lower RMSE than convex penalties.

non-convex cases, we set  $\lambda = \sqrt{N}\sigma/4$ , consistent with the range suggested in [7] for standard (convex) TV denoising. We set the non-convexity parameter to its maximal value,  $a = 1/(4\lambda)$ . We use 20 iterations of (32).

The improvement of non-convex regularization is reflected in the lower RMSE of 0.25 compared to 0.32. For further comparison, Fig. 1 shows the error,  $x^* - s$ , for both convex and non-convex regularized solutions. The convex solution underestimates the true first-order difference signal more so than the non-convex one.

The optimality of the non-convex solution acquired using iteration (32) is validated using (23). The condition is graphically illustrated as a scatter plot in Fig. 2. The preponderance of points on the vertical line,  $[Dx]_n = 0$ , reflects the sparsity of  $Dx^*$ .

### B. Example 2

In this example, we consider the relative performance of convex and non-convex regularized TV denoising as a function of noise power. We generate random realizations of PWC signals. Each realization is of length 1000 and has 15 step-edges. The step-edges are uniformly distributed over the duration of the signal. The amplitudes of the steps are uniformly distributed in  $[-5, 5]$ . We corrupt each realization with AWGN,  $\mathcal{N}(0, \sigma^2)$ . TV denoising is applied to each noise-corrupted realization using  $\lambda = \sqrt{N}\sigma/4$  and  $a = 1/(4\lambda)$  as

above. Figure 3 illustrates the RMSE as a function of  $\sigma$ . It can be seen that non-convex regularization offers the most improvement at low noise levels.

## VII. CONCLUSION

TV denoising is a basic method for the estimation of PWC signals in noise. This letter gives a modification of the standard TV denoising problem where the regularizer is non-convex yet constrained so that the total objective function is convex. The improvement is not as dramatic as that offered by non-convex regularization without such a constraint – see [16] for examples. However, due to the convexity, the solution is reliably obtained via convex optimization, the solution depends continuously on the data, and the regularization parameter can be set as in the convex case (e.g., [7]).

## REFERENCES

- [1] 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.
- [2] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.
- [3] 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.
- [4] E. Chouzenoux, A. Jeziarska, J. Pesquet, and H. Talbot. A majorize-minimize subspace approach for  $\ell_2 - \ell_0$  image regularization. *SIAM J. Imag. Sci.*, 6(1):563–591, 2013.
- [5] L. Condat. A direct algorithm for 1-D total variation denoising. *IEEE Signal Processing Letters*, 20(11):1054–1057, November 2013.
- [6] C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot. Dual constrained TV-based regularization on graphs. *SIAM J. Imag. Sci.*, 6(3):1246–1273, 2013.
- [7] L. Dümbgen and A. Kovac. Extensions of smoothing via taut strings. *Electron. J. Statist.*, 3:41–75, 2009.
- [8] M. Figueiredo, J. Bioucas-Dias, and R. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Trans. Image Process.*, 16(12):2980–2991, December 2007.
- [9] J.-J. Fuchs. Convergence of a sparse representations algorithm applicable to real or complex data. *IEEE. J. Sel. Top. Signal Processing*, 1(4):598–605, December 2007.
- [10] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer, 2001.
- [11] M. A. Little and N. S. Jones. Generalized methods and solvers for noise removal from piecewise constant signals: Part I – background theory. *Proc. R. Soc. A*, 467:3088–3114, 2011.
- [12] E. Mammen and S. van de Geer. Locally adaptive regression splines. *The Annals of Statistics*, 25(1):387–413, February 1997.
- [13] M. Nikolova. Estimation of binary images by minimizing convex criteria. In *IEEE Int. Conf. Image Proc.*, pages 108–112 vol. 2, 1998.
- [14] M. Nikolova. Markovian reconstruction using a GNC approach. *IEEE Trans. Image Process.*, 8(9):1204–1220, 1999.
- [15] M. Nikolova. Local strong homogeneity of a regularized estimator. *SIAM J. Appl. Math.*, 61(2):633–658, 2000.
- [16] M. Nikolova. Energy minimization methods. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, chapter 5, pages 138–186. Springer, 2011.
- [17] 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.
- [18] 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.
- [19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C: the art of scientific computing (2nd ed.)*. Cambridge University Press, 1992.
- [20] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [21] I. Selesnick. Penalty and shrinkage functions for sparse signal processing. *Connexions*, 2012. <http://cnx.org/content/m45134/1.1/>.
- [22] I. W. Selesnick and I. Bayram. Sparse signal estimation by maximally sparse convex optimization. *IEEE Trans. Signal Process.*, 62(5):1078–1092, March 2014.