A Direct Algorithm for 1D Total Variation Denoising

Laurent Condat

A Direct Algorithm for 1D Total Variation Denoising

Laurent Condat, *Member, IEEE*

***Abstract*— A very fast noniterative algorithm is proposed for denoising or smoothing one-dimensional discrete signals, by solving the total variation regularized least-squares problem or the related fused lasso problem. A C code implementation is available on the web page of the author.**

***Index Terms*— Total variation, denoising, nonlinear smoothing, fused lasso, regularized least-squares, nonparametric regression, taut string algorithm, accelerated Douglas-Rachford algorithm, convex nonsmooth optimization, splitting**

I. INTRODUCTION

The problem of smoothing a signal, to remove or at least attenuate the noise it contains, has numerous applications in communications, control, machine learning, and many other fields of engineering and science [1]. In this paper, we focus on the numerical implementation of total variation (TV) denoising for one-dimensional (1D) discrete signals; that is, we are given a (noisy) signal $y = (y[1], \ldots, y[N]) \in \mathbb{R}^N$ of size $N \geq 1$, and we want to efficiently compute the denoised signal $x \in \mathbb{R}^N$, defined implicitely as the solution to the minimization problem

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \frac{1}{2}\sum_{k=1}^{N}\big|y[k]-x[k]\big|^2+\lambda\sum_{k=1}^{N-1}\big|x[k+1]-x[k]\big|, \quad (1)$$

for some regularization parameter $\lambda \geq 0$ (whose choice is a difficult problem by itself [2]). We recall that, as the functional to minimize is strongly convex, the solution x to the problem exists and is unique, whatever the data y. The TV denoising problem has received large attention in the communities of signal and image processing, inverse problems, sparse sampling, statistical regression analysis, optimization theory, among others. It is not the purpose of this paper to review the properties of the nonlinear TV denoising filter, as numerous papers can be found on this vast topic; see, e.g., [3]–[8] for various insights.

To solve the convex nonsmooth optimization problem (1), we mostly find in the literature iterative fixed-point methods [9], [10]. Until not so long ago, such methods applied to TV regularization had rather high computational complexity [11]–[15], but the growing interest for related ℓ_1-norm problems in compressed sensing, sparse recovery, or low rank matrix completion [16]–[18], has yielded advances in the field. Recent iterative methods based on operator splitting, which exploit both the primal and dual formulations of the problems and use variable stepsize strategies or Nesterov-style accelerations, are quite efficient when applied to TV-based problems [19]–[23]. Graph cuts methods can also be used to solve (1) or its extension on graphs [24]. They actually solve a quantized version of (1): the minimizer x is not searched in $\mathbb{R}^N$ but in $\varepsilon\,\mathbb{Z}^N$, for some $\varepsilon > 0$, with complexity $O(\log_2(1/\varepsilon)N)$. If ε is small enough, the exact solution in $\mathbb{R}^N$ can be obtained from the quantized one, as shown by Hochbaum [25], [26]. In this paper, we present a novel and very fast algorithm to compute the denoised signal x solution to (1), exactly, in a direct, noniterative, way, possibly in-place. It is appropriate for real-time processing of an incoming stream of data, as it locates the jumps in x one after the other by forward scans, almost online. The possibility of such an algorithm sheds light on the relatively local nature of the TV denoising filter [27].

L. Condat is with the Image Team of the GREYC laboratory, a joint CNRS-ENSICAEN-Univ. of Caen research unit in Caen, France. Contact: see http://www.greyc.ensicaen.fr/∼lcondat/.

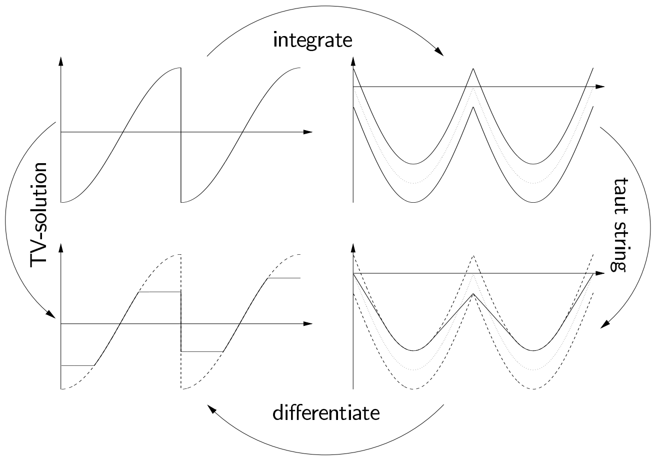


Fig. 1. Total variation denoising can be interpreted as pulling the antiderivative of the signal taut in a tube around it. The proposed algorithm is different from the so-called *taut string algorithm* implementing this principle. This figure is borrowed from a PDF slide of a talk by M. Grasmair in 2007, entitled "dual settings for total variation regularization".

After this work was completed, the author found that, actually, there already exists a direct, linear time, method for 1D TV denoising, called the *taut string algorithm* [28], see also [29]–[32]. To understand its principle, define the sequence of running sums r by $r[k] = \sum_{i=1}^{k} y[i]$ for $1 \leq k \leq N$, and consider the problem:

$$\underset{s\in\mathbb{R}^{N+1}}{\text{minimize}}\ \sum_{k=1}^{N}\sqrt{1+\big|s[k]-s[k-1]\big|^2}\quad\text{subject to}$$
$$s[0]=0,\ s[N]=r[N],\ \text{and}\ \max_{1\leq k\leq N}\big|s[k]-r[k]\big|\leq\lambda. \quad (2)$$

Then, the problems (1) and (2) are equivalent, in the sense that their respective solutions x and s are related by $x[k] = s[k]-s[k-1]$, for $1 \leq k \leq N$ [28], [33]. Thus, the formulation (2) allows to express the TV solution x as the discrete derivative of a string threaded through a tube around the discrete primitive of the data, and pulled taut such that its length is minimized. This principle is illustrated in Fig. 1. Its implementation consists in alternating between the computation of the greatest convex minorant and least concave majorant of the tube walls $r + \lambda$ and $r - \lambda$. The taut string method seems to have been largely ignored, as iterative methods are regularly proposed for 1D TV denoising [34]–[37]. The proposed algorithm is different, as it does not manipulate any running sum and only performs forward scans; the signal x is constructed definitively segment by segment.

The paper is organized as follows. In Sect. II, we describe and discuss the new algorithm. In Sect. III, we suggest some applications.

II. Proposed Method

We first introduce the *dual* problem to the *primal* problem (1) [13], [21], [22]:

$$\underset{u\in\mathbb{R}^{N+1}}{\text{minimize}} \quad \sum_{k=1}^{N} \big|y[k]-u[k]+u[k-1]\big|^2 \quad \text{s.t.}$$
$$|u[k]|\leq \lambda,\ \forall k=1,\dots,N-1,\ \text{ and } u[0]=u[N]=0. \quad (3)$$

Once the solution u to the dual problem is found, one recovers the primal solution x by

$$x[k]=y[k]-u[k]+u[k-1],\ \forall k=1,\dots,N. \quad (4)$$

Actually, the method of [13] and its accelerated version [20] solve (3) iteratively, using forward-backward splitting [9].

The Karush-Kuhn-Tucker conditions caracterize the unique solutions x and u [22]. They yield, in addition to (4),

$$u[0]=u[N]=0 \quad \text{and} \quad \forall k=1,\dots,N-1,$$
$$\begin{cases} u[k]\in[-\lambda,\lambda] & \text{if } x[k]=x[k+1],\\ u[k]=-\lambda & \text{if } x[k]<x[k+1],\\ u[k]=\lambda & \text{if } x[k]>x[k+1]. \end{cases} \quad (5)$$

Hence, the proposed algorithm consists in running forwardly through the samples $y[k]$; at location k, it tries to prolongate the current segment of x by $x[k+1]=x[k]$. If this is not possible without violating (4) and (5), it goes back to the last location where a jump can be introduced in x, validates the current segment until this location, starts a new segment, and continues. In more details, the proposed direct algorithm is as follows:

1D TV Denoising Algorithm:
(a) Set $k=k_0=k_-=k_+\leftarrow 1$, $v_{\min}\leftarrow y[1]-\lambda$, $v_{\max}\leftarrow y[1]+\lambda$, $u_{\min}\leftarrow\lambda$, $u_{\max}\leftarrow -\lambda$.
(b) If $k=N$, set $x[N]\leftarrow v_{\min}+u_{\min}$ and terminate. Else, we are at location k and we are building a segment starting at k_0, with value $v=x[k_0]=\dots=x[k]$. v is unknown but we know $v_{\min}$ and $v_{\max}$ such that $v\in[v_{\min},v_{\max}]$. $u_{\min}$ and $u_{\max}$ are the values of $u[k]$ in case $v=v_{\min}$ and $v=v_{\max}$, respectively. Now, we are trying to prolongate the segment with $x[k+1]=v$, by updating the four variables $v_{\min}$, $v_{\max}$, $u_{\min}$, $u_{\max}$, for the location $k+1$. The three possible cases $(b1)$, $(b2)$, $(b3)$ are:
$(b1)$ If $y[k+1]+u_{\min}<v_{\min}-\lambda$, we cannot update $u_{\min}$ without violating (4) and (5), because $v_{\min}$ is too high. This means that the assumption $x[k_0]=\dots=x[k+1]$ was wrong, so that the segment must be broken, and the negative jump necessarily takes place at the last location k_- where $u_{\min}$ was equal to λ. Thus, we set $x[k_0]=\dots=x[k_-]\leftarrow v_{\min}$, $k=k_0=k_-=k_+\leftarrow k_-+1$, $v_{\min}\leftarrow y[k]$, $v_{\max}\leftarrow y[k]+2\lambda$, $u_{\min}\leftarrow\lambda$, $u_{\max}\leftarrow-\lambda$.
$(b2)$ Else, if $y[k+1]+u_{\max}>v_{\max}+\lambda$, then by the same reasoning, a positive jump must be introduced at the last location k_+ where $u_{\max}$ was equal to $-\lambda$. Thus, we set $x[k_0]=\dots=x[k_+]\leftarrow v_{\max}$, $k=k_0=k_-=k_+\leftarrow k_++1$, $v_{\min}\leftarrow y[k]-2\lambda$, $v_{\max}\leftarrow y[k]$, $u_{\min}\leftarrow\lambda$, $u_{\max}\leftarrow-\lambda$.
$(b3)$ Else, no jump is necessary yet, and we can continue with $k\leftarrow k+1$. So, we set $u_{\min}\leftarrow u_{\min}+y[k]-v_{\min}$ and $u_{\max}\leftarrow u_{\max}+y[k]-v_{\max}$. It may be necessary to update the bounds $v_{\min}$ and $v_{\max}$:
$(b31)$ If $u_{\min}\geq\lambda$, set $v_{\min}\leftarrow v_{\min}+(u_{\min}-\lambda)/(k-k_0+1)$, $u_{\min}\leftarrow\lambda$, $k_-\leftarrow k$.
$(b32)$ If $u_{\max}\leq-\lambda$, set $v_{\max}\leftarrow v_{\max}+(u_{\max}+\lambda)/(k-k_0+1)$, $u_{\max}\leftarrow-\lambda$, $k_+\leftarrow k$.
(c) If $k<N$, go to (b). Else, we have to test if the hypothesis of a segment $x[k_0]=\dots=x[N]$ does not violate the condition $u[N]=0$. The three possible cases are:
$(c1)$ If $u_{\min}<0$, then $v_{\min}$ is too high and a negative jump is necessary: set $x[k_0]=\dots=x[k_-]\leftarrow v_{\min}$, $k=k_0=k_-\leftarrow k_-+1$, $v_{\min}\leftarrow y[k]$, $u_{\min}\leftarrow\lambda$, $u_{\max}\leftarrow y[k]+\lambda-v_{\max}$, and go to (b).
$(c2)$ Else, if $u_{\max}>0$, then $v_{\max}$ is too low and a positive jump is necessary: set $x[k_0]=\dots=x[k_-]\leftarrow v_{\max}$, $k=k_0=k_+\leftarrow k_++1$, $v_{\max}\leftarrow y[k]$, $u_{\max}\leftarrow-\lambda$, $u_{\min}\leftarrow y[k]-\lambda-v_{\min}$, and go to (b).
$(c3)$ Else, set $x[k_0]=\dots=x[N]\leftarrow v_{\min}+u_{\min}/(k-k_0+1)$ and terminate.

We note that the dual solution u is not computed. We can still recover it recursively from x using (4). We also remark that the case $\lambda=0$ is correctly handled and yields $x=y$.

The worst case complexity of the algorithm is $O(N+N-1+\dots+1)=O(N^2)$. Indeed, every added segment has size at least one, but the algorithm may have to scan all the remaining samples to validate it in one of the steps $(b1)$, $(b2)$, $(c1)$, $(c2)$. However, this worst case scenario is encountered only when x is a ramp with *very small* slope of order N^{-2}, except at the boundaries; for instance, consider that $\lambda=1$ and $y[1]=-2$, $y[k]=\alpha(k-2)$ for $2\leq k\leq N-1$, $y[N]=\alpha(N-3)+2$, where $\alpha=4/((N-2)(N-3))$. The solution x is such that $x[1]=y[1]+1$, $x[k]=y[k]$ for $2\leq k\leq N-1$, $x[N]=y[N]-1$. Actually, such a pathological case, for which there is no interest in applying TV denoising, is only a curiosity, and the complexity is $O(N)$ in all practical situations, as the segments of x are validated with a delay which does not depend on N.

The algorithm was implemented in C, compiled with gcc 4.4.1, and run on a Apple laptop with a 2.4 GHz Intel Core 2 Duo processor. We obtained computation times around 30 milliseconds for $N=10^6$, with various test signals and noise levels. Importantly, the computation time is insensitive to the value of λ.

For illustration purpose, we consider the example of a discrete Lévy process, which is a stochastic process with independent increments [8], corrupted by additive white Gaussian noise (AWGN). More precisely, $y[k]=x_*[k]+e[k]$ for $1\leq k\leq N=1000$, where the $e[k]\sim\mathcal{N}(0,1)$ are independent and identically distributed (i.i.d.), and the ground truth x_* has a fixed value $x_*[1]$ and i.i.d. random increments $d[k]=x_*[k]-x_*[k-1]$ for $2\leq k\leq N$. We chose a sparse Bernoulli-Gaussian law for the increments, since TV denoising approaches the optimal minimum mean square error (MMSE) estimator for such piecewise constant signals [7], [8]; that is, the probability density function of $d[k]$ is

$$p\,\delta(t)+\frac{(1-p)}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}},\quad t\in\mathbb{R}, \quad (6)$$

where $p=0.95$, $\sigma=4$, and $\delta(t)$ is the Dirac distribution. We found empirically that the root mean square error (RMSE) $\|x_*-x\|_2/\sqrt{N}$ is minimized for $\lambda=2$. The computation time of x, averaged over several runs and realizations, was 30 microseconds. One realization of the experiment is depicted in Fig. 2.

III. Further Applications

Besides denoising of 1D signals, the proposed algorithm can be used as a black box to solve other problems.

A. The Fused Lasso

The *fused lasso signal approximator*, introduced in [38], yields a solution that has sparsity in both the coefficients and their successive differences. It consists in solving the problem

$$\underset{z\in\mathbb{R}^N}{\text{minimize}}\ \frac{1}{2}\sum_{k=1}^{N}\big|z[k]-y[k]\big|^2+\lambda\sum_{k=1}^{N-1}\big|z[k+1]-z[k]\big|+\mu\sum_{k=1}^{N}\big|z[k]\big|, \quad (7)$$

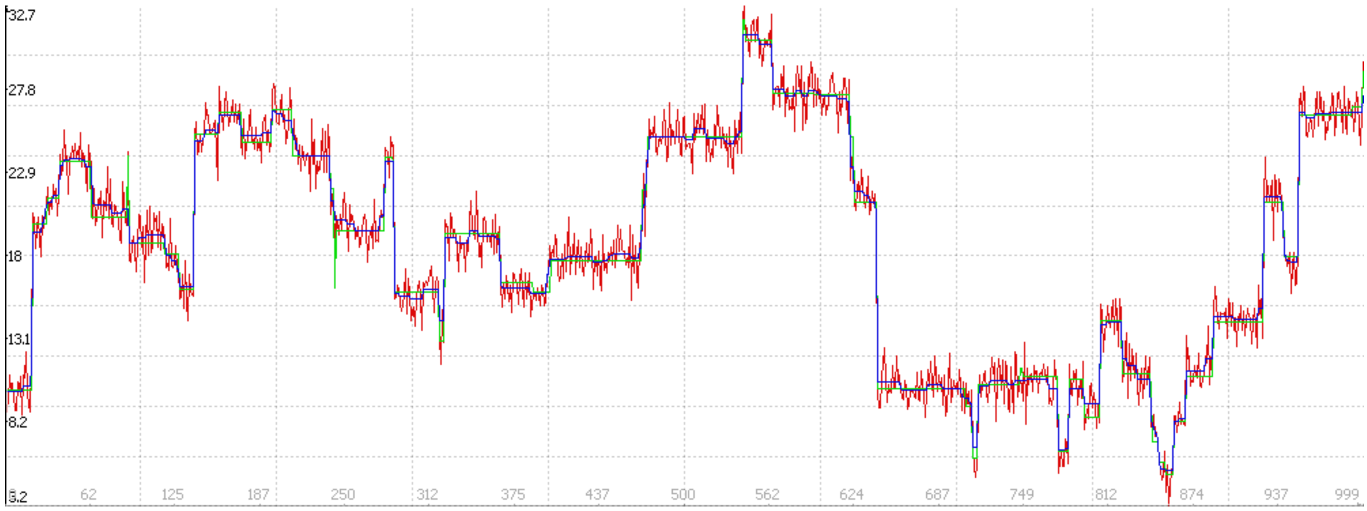


Fig. 2. In this example, y (in red) is a piecewise constant process of size $N = 1000$ (unknown ground truth, in green) corrupted by additive Gaussian noise of unit variance. The TV-denoised signal x (in blue), solving (1) with $\lambda = 2$ exactly, was computed by the proposed algorithm in 30 microseconds.

for some $\lambda \geq 0$ and $\mu \geq 0$. The fused lasso has many applications, e.g. in bioinformatics [39]–[41]. As shown in [40], the complexity of the fused lasso is the same as TV denoising, since the solution z can be obtained by simple soft-thresholding from the solution x of (1):

$$z[k] = \begin{cases} x[k] - \mu.\text{sign}(x[k]) & \text{if } |x[k]| > \mu \\ 0 & \text{otherwise} \end{cases}. \tag{8}$$

It is straightforward to add soft-thresholding steps to the proposed algorithm to solve the generalization (7) of (1), for essentially the same computation time.

B. Using the Algorithm as a Proximity Operator

As is classical in convex analysis, we introduce the set $\Gamma_0(\mathbb{R}^N)$ of proper, lower semi-continuous, convex functions from $\mathbb{R}^N$ to $\mathbb{R} \cup \{+\infty\}$ [9]. Many problems in signal and image processing can be formulated as finding a minimizer $x \in \mathbb{R}^N$ of the sum of functions $F_i \in \Gamma_0(\mathbb{R}^N)$, where each F_i is introduced to enforce some constraint or promote some property on the solution [6], [9], [16], [18]. To solve such problems, convex nonsmooth optimization theory provides us with first-order proximal splitting methods [9], [22], which call the gradient operator or the *proximity operator* of each function F_i, individually and iteratively. The Moreau proximity operator of a function $F \in \Gamma_0(\mathbb{R}^N)$ is defined as

$$\text{prox}_F : s \in \mathbb{R}^N \mapsto \underset{s' \in \mathbb{R}^N}{\text{argmin}} \, \frac{1}{2}\|s - s'\|^2 + F(s'). \tag{9}$$

Thus, if we define $TV : r \in \mathbb{R}^N \mapsto \sum_{k=1}^{N-1} |r[k+1] - r[k]|$, we can rewrite (1) as $x = \text{prox}_{\lambda TV}(y)$. In other words, the proposed algorithm computes the proximity operator of the 1D TV. Hence, we are equipped to solve any convex minimization problem which can be expressed in terms of the 1D TV. For instance, we can denoise an image y of size $N_1 \times N_2$ by applying the proximity operator of the 2D anisotropic TV:

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \; \underbrace{\frac{1}{2}\|x - y\|^2 + \lambda \sum_{k_1=1}^{N_1} TV_{v,k_1}(x)}_{F_1(x)} + \underbrace{\lambda \sum_{k_2=1}^{N_2} TV_{h,k_2}(x)}_{F_2(x)}, \tag{10}$$

where $TV_{v,k_1}(x)$ and $TV_{h,k_2}(x)$ are the TV of the k_1-th column and k_2-th row of the image x, seen as 1D signals, respectively, and $N = N_1 N_2$. To find a minimizer of the sum of two proximable functions F_1 and F_2 of $\Gamma_0(\mathbb{R}^N)$, we propose a new splitting algorithm as follows:

Accelerated Douglas-Rachford Algorithm (ADRA)
Fix $\gamma > 0$, $x_0, s_0 \in \mathbb{R}^N$, and iterate, for $n = 0, 1, \ldots$

$$\left| \begin{array}{l} r_{n+1} = s_n - x_n + \text{prox}_{\gamma F_1}(2x_n - s_n), \\ s_{n+1} = r_{n+1} + \frac{n}{n+3}(r_{n+1} - r_n), \\ x_{n+1} = \text{prox}_{\gamma F_2}(s_{n+1}). \end{array} \right.$$

Establishing convergence properties of splitting algorithms is a hot topic in the community of convex optimization, and the ongoing concern of the author [22]. Although there is currently no convergence proof of x_n to the minimizer x of $F_1 + F_2$ as $n \to +\infty$ with ADRA, it was found empirically to converge and to be remarkably effective for the problem (10), with $\gamma = 1$ and $s_0 = x_0 = y$. For the example illustrated in Fig. 3, we considered the classical *Lena* image of size 512×512, with gray values in $[0, 255]$, corrupted by AWGN of std. dev. 30. When used to solve (10) with $\lambda = 30$, ADRA consists in applying the 1D TV denoising algorithm on the rows and columns of the image, iteratively. Remarkably, the convergence is very fast and the image x_5 after five iterations is visually identical to the image x obtained at convergence, with a RMSE of 0.5 gray levels, for a computation time of 0.27s. This is about four times less than the times reported in [24] with state-of-the-art graph-cuts approaches and a similar quantization step of 1 gray level. Still, the latter remain faster if a quantization step of 2^{-16}, corresponding to machine precision, is to be reached.

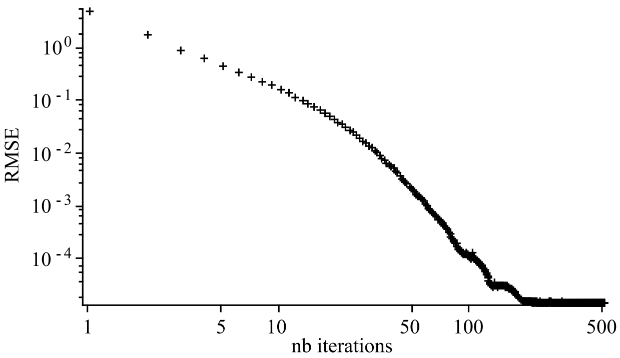
IV. Conclusion

In this article, we proposed a direct and very fast algorithm for denoising 1D signals by total variation (TV) minimization or fused lasso approximation. Since the algorithm computes the proximity operator of the 1D TV, it can be used as a basic unit within iterative splitting methods, like the new proposed accelerated Douglas-Rachford algorithm, to solve more complex inverse multidimensional problems.

This work opens the door for a variety of extensions and applications. Future work will include the extension of the algorithm to generalized forms of the TV, where the two-tap finite difference is replaced by another discrete differential operator, to favor piecewise

(a)

(b)



(c)

Fig. 3. (a) The image *Lena* corrupted by AWGN of std. dev. 30; (b) the image after 5 iterations of the proposed ADRA method for 2D anisotropic TV denoising; (c) log-log plot of the RMSE between the iterate x_n and the solution x of (10) with $\lambda = 30$, in term of the number n of iterations.

polynomial reconstruction of higher degree or other types of signals [33], [35]. Also, the algorithm should be extended to complex-valued or multi-valued signals [41]. The extension to data of higher dimensions, like 2D images or graphs, deserves further investigation as well [31]. Furthermore, we should consider replacing the quadratic data fidelity term by other penalties, like the anti-log-likelihood of Poisson noise [32].

Besides, path-following, a.k.a. homotopy, algorithms have been proposed for ℓ_1 penalized problems; they can find the smallest value of λ and the associated x in (1) such that x has at most m segments, with complexity $O(mN)$ [18], [29], [40], [42]–[44]. The relationship between such algorithms, the approach in [45], and the proposed one should be studied.

References

[1] M. A. Little and N. S. Jones, "Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory & II. New methods," *Proc. R. Soc. A*, vol. 467, 2011.

[2] C. Deledalle, S. Vaiter, G. Peyré, J. Fadili, and C. Dossal, "Unbiased risk estimation for sparse analysis regularization," Jan. 2012, preprint hal-00662718.

[3] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.

[4] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[5] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, June 2007.

[6] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, "An introduction to total variation for image analysis," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, vol. 9. De Gruyter, Radon Series Comp. Appl. Math., 2010, pp. 263–340.

[7] R. Gribonval, "Should penalized least squares regression be interpreted as maximum a posteriori estimation?" *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2405–2410, May 2011.

[8] U. Kamilov, A. Amini, and M. Unser, "MMSE denoising of sparse Lévy processes via message passing," in *Proc. of IEEE ICASSP*, Mar. 2012.

[9] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York: Springer-Verlag, 2010.

[10] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[11] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, pp. 227–238, 1996.

[12] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 231–241, Feb. 2001.

[13] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1–2, pp. 89–97, 2004.

[14] T. Chan, S. Esedoglu, F. Park, and M. Yip, "Recent developments in total variation image restoration," in *Mathematical Models of Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. Springer Verlag, 2005.

[15] B. Wohlberg and P. Rodriguez, "An iteratively reweighted norm algorithm for minimization of total variation functionals," *IEEE Signal Processing Lett.*, vol. 14, no. 12, pp. 948–951, 2007.

[16] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal Processing Mag.*, vol. 27, no. 3, pp. 76–88, 2010.

[17] S. Becker, E. J. Candès, and M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Prog. Comp.*, vol. 3, no. 3, pp. 165–218, Sept. 2011.

[18] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," preprint hal-00613125, to appear in Foundations and Trends in Machine Learning.

[19] M. Zhu and T. F. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," 2008, uCLA CAM Report 08-34.

[20] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Processing*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[21] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[22] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," 2011, preprint hal-00609728.

[23] S. Bonettini and V. Ruggiero, "On the convergence of primal–dual hybrid gradient algorithms for total variation image restoration," to appear in *J. Math. Imaging and Vision*.

[24] A. Chambolle and J. Darbon, "On total variation minimization and surface evolution using parametric maximal flows," *International Journal of Computer Vision*, vol. 84, no. 3, 2009.

[25] ——, "A parametric maximum flow approach for discrete total variation regularization," in *Image Processing and Analysis with Graphs: Theory and Practice*, ser. Digital Imaging and Computer Vision, R. Lukac, Ed. CRC Press / Taylor and Francis, 2012.

[26] D. S. Hochbaum, "An efficient and effective tool for image segmentation, total variations and regularization," in *Proc. of Int. Conf. on Scale Space and Variational Methods in Computer Vision (SSVM)*, 2011, vol. LNCS 6667, pp. 338–349.

[27] C. Louchet and L. Moisan, "Total variation as a local filter," *SIAM J. on Imaging Sciences*, vol. 4, no. 2, pp. 651–694, 2011.

[28] P. L. Davies and A. Kovac, "Local extremes, runs, strings and multiresolution," *The Annals of Statistics*, vol. 29, no. 1, pp. 1–65, 2001.

[29] E. Mammen and S. van de Geer, "Locally adaptive regression splines," *The Annals of Statistics*, vol. 25, no. 1, pp. 387–413, 1997.

[30] M. Grasmair, "The equivalence of the taut string algorithm and BV-regularization," *J. Math. Imaging Vision*, vol. 27, no. 1, pp. 59–66, 2007.

[31] W. Hinterberger, M. Hintermüller, K. Kunisch, M. von Oehsen, and O. Scherzer, "Tube methods for BV regularization," *Journal of Mathematical Imaging and Vision*, vol. 19, no. 3, pp. 219–235, 2003.

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

[33] G. Steidl, S. Didas, and J. Neumann, "Splines in higher order TV regularization," *International Journal of Computer Vision*, vol. 70, no. 3, pp. 241–255, 2006.

[34] A. Barbero and S. Sra, "Fast Newton-type methods for total variation regularization," in *Proc. of Int. Conf. Machine Learning (ICML)*, June 2011, pp. 313–320.

[35] F. I. Karahanoglu, I. Bayram, and D. V. D. Ville, "A signal processing approach to generalized 1D total variation," *IEEE Trans. Signal Processing*, vol. 59, no. 11, pp. 5265–5274, Nov. 2011.

[36] U. Kamilov, E. Bostan, and M. Unser, "Generalized total variation denoising via augmented Lagrangian cycle spinning with Haar wavelets," in *Proc. of IEEE ICASSP*, Mar. 2012.

[37] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM algorithm for a class of total variation regularized estimation problems," in *Proc. of IFAC Symposium on System Identification*, July 2012.

[38] R. Tibshirani, M.Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. The Royal Statistical Society Series B*, vol. 67, no. 1, pp. 91–108, 2005.

[39] R. Tibshirani and P. Wang, "Spatial smoothing and hot spot detection for CGH data using the fused lasso," *Biostatistics*, vol. 9, pp. 18–29, 2008.

[40] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 302–332, 2007.

[41] K. Bleakley and J.-P. Vert, "The group fused lasso for multiple change-point detection," June 2011, preprint hal-00602121.

[42] H. Höfling, "A path algorithm for the fused lasso signal approximator," *Journal of Computational and Graphical Statistics*, vol. 19, no. 4, pp. 984–1006, 2010.

[43] R. J. Tibshirani and J. Taylor, "The solution path of the generalized lasso," *The Annals of Statistics*, vol. 39, no. 3, pp. 1335–1371, 2011.

[44] H. Zhou and K. Lange, "A path algorithm for constrained estimation," 2011, preprint arXiv:1103.3738.

[45] I. Pollak, A. S. Willsky, and Y. Huang, "Nonlinear evolution equations as fast and exact solvers of estimation problems," *IEEE Trans. Signal Processing*, vol. 53, no. 2, pp. 484–498, Feb. 2005.