TOTAL VARIATION DENOISING WITH OVERLAPPING GROUP SPARSITY

Ivan W. Selesnick and Po-Yu Chen

Polytechnic Institute of New York University, Brooklyn, NY 11201, USA

ABSTRACT

This paper describes an extension to total variation denoising wherein it is assumed the first-order difference function of the unknown signal is not only sparse, but also that large values of the first-order difference function do not generally occur in isolation. This approach is designed to alleviate the staircase artifact often arising in total variation based solutions. A convex cost function is given and an iterative algorithm is derived using majorization-minimization. The algorithm is both fast converging and computationally efficient due to the use of fast solvers for banded systems.

***Index Terms*—** total variation, sparse signal processing, L1 norm, group sparsity, denoising, filter, convex optimization.

1. INTRODUCTION

Total variation (TV) [27] is commonly used as a penalty function in sparse signal processing. For example, total variation has been used extensively for denoising [5, 7, 8, 27], deconvolution [3, 22, 23], reconstruction [32], nonlinear decomposition [29,31], and compressed sensing [33]. Numerous algorithms have been developed for TV-regularized inverse problems, e.g. [5–7, 26, 32, 35]. For 1-D TV denoising, the exact solution can be obtained by a direct algorithm [10].

However, total variation has some shortcomings. Signals produced via TV-based processing often exhibit stair-case artifacts (which appear as artificial contours in images). For this reason, several generalizations and extensions of total variation have been proposed in the literature [4, 15, 18, 19, 26]. While total variation is suitable for piecewise-constant signals (i.e. signals with a sparse derivative function), for signals that are locally approximated by higher order polynomials, generalized forms of total variation, such as those referenced, are more appropriate.

This paper describes an extension of total variation that aims to take into account group sparsity characteristics of signal derivatives. That is, it is assumed here that the signal of interest has a derivative that is not only sparse, but exhibits a simple form of structured sparsity. Specifically, it is assumed that large values of the derivative are not isolated, but usually arise near, or adjacent to, other large values. In this sense, points where the signal value changes rapidly, have a clustering or grouping property. Such a signal is approximately piecewise constant; however, the edges (step changes) of the signal are not exact discontinuities, but instead extend over some interval.

The approach described in this paper, like conventional total variation methods, is based on the minimization of a convex non-differential cost function. The group/clustering behavior of the signal derivative is promoted by a suitable penalty function. It is an objective of this work that the processing be translation-invariant. It is not assumed that the groups are known in advance of the processing (this paper focuses on 1D signal denoising). Moreover, it is not intended that the grouping behavior be clearly defined in the signal, as many natural signals are not so simply described. The grouping property referred to here, is the general tendency of large values to congregate rather than to occur in isolation. For this reason, the groups arising in the problem formulation are fully overlapping (as in sliding window processing, with the window being translated one sample at a time).

The iterative algorithm developed in this paper is derived using the majorization-minimization (MM) optimization method [13]. The algorithm is formulated so that the main computational step consists of solving a tridiagonal system of equations per iteration, which can be done very efficiently with fast solvers for banded systems of linear equations. The algorithm does not require any parameters (step size, etc.). The problem formulation and algorithm are illustrated with two examples: an artificial signal and a row extracted from a standard test image.

1.1. Related Work

Previous generalizations of total variation (e.g., above cited works) generally focus on improving its performance for signals that have a higher order smoothness than piecewise constant signals; for example, by replacing the first-order difference by which (discrete) TV is defined, by a higher order filter. On the other hand, the generalization discussed in this work, like conventional TV, is suitable for signals that are largely constant (flat), but for which changes in value are not exact discontinuities.

This work uses group sparsity concepts that have been previously used for sparse signal processing. Group lasso [34], a generalization of the lasso [30], is suitable when the signal to be estimated is known to be group sparse with non-overlapping groups and the group structure is known *a priori*. In contrast, for general signal denoising and restoration, the groups (clusters) of large values may arise anywhere in the domain of the signal. In this case, if the group structure were defined *a priori*, a group of large values may straddle two of the predefined groups. Hence, it is suitable to formulate the problem in terms of overlapping groups, as in Refs. [1, 2, 9, 11, 12, 16, 17, 24]. The penalty function (4), below, is of the form that is most studied and utilized in these works. This paper utilizes the overlapping-group sparsity-promoting penalty function (4) for the purpose of total variation denoising. In addition, the algorithm derived below, is distinct from previous algorithms for overlapping sparsity. Previous algorithms call for auxiliary variables (via variable splitting, replication, etc.) proportional to the overlapping factor, which entails additional memory proportional to the group size. The MM algorithm below does not utilize auxiliary variables nor does it require excess memory.

We note that, for overlapping group sparsity, an alternative to the penalty function (4), is the one proposed in [21]. As future work, it will be interesting to compare group-sparse total variation denoising using each of the two formulations of overlapping group sparsity.

This work is supported by the NSF under Grant No. CCF-1018020.
Email: selesi@poly.edu, poyupaulchen@gmail.com

1.2. Notation

An N-point signal $x(n)$, $n = 0, \ldots, N-1$, is represented as the column vector,

$$\mathbf{x} = [x(0), \ldots, x(N-1)]^T \in \mathbb{R}^N.$$

The first-order difference matrix is represented by $\mathbf{D}$, i.e.,

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{bmatrix}. \tag{1}$$

The first-order difference of an N-point signal $\mathbf{x}$ is given by $\mathbf{Dx}$ where $\mathbf{D}$ is of size $(N-1) \times N$.

A K-point group of the vector $\mathbf{v}$ will be denoted by

$$\mathbf{v}_{n,K} = [v(n), \ldots, v(n+K-1)] \in \mathbb{R}^K. \tag{2}$$

This is a block of K contiguous samples of $\mathbf{v}$ starting at index n.

2. GROUP-SPARSE TOTAL VARIATION DENOISING

It is assumed the unknown signal $\mathbf{x} \in \mathbb{R}^N$ is observed in additive independent white Gaussian noise $\mathbf{w}$. As described in the Introduction, it is assumed that the derivative (first-order difference, given by $\mathbf{Dx}$) of $\mathbf{x}$ has a group sparse behavior. Given data, $\mathbf{y} = \mathbf{x} + \mathbf{w}$, an estimate of $\mathbf{x}$ can be obtained as the solution to the optimization problem

$$\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathbb{R}^N} \left\{ F(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\, \phi(\mathbf{Dx}) \right\} \tag{3}$$

where ϕ is a penalty function that promotes group sparsity. Note that $\mathbf{v} = \mathbf{Dx} \in \mathbb{R}^{(N-1)}$, hence $\phi : \mathbb{R}^{(N-1)} \to \mathbb{R}$.

In this work, we use the function ϕ defined by

$$\phi(\mathbf{v}) = \sum_n \left[\sum_{k=0}^{K-1} |v(n+k)|^2 \right]^{1/2}. \tag{4}$$

This regularizer is commonly used to promote group sparsity [1, 2, 11, 12, 16, 17, 24]. The group size is denoted by K. (A different regularizer to promote group sparsity is given in [21].)

In (4), for $n+k$ outside the index range of $\mathbf{v}$, we take $v(n+k)$ as zero.

If $K = 1$, then $\phi(\mathbf{v}) = \|\mathbf{v}\|_1$ and problem (3) is the standard 1D total variation denoising problem. If $K > 1$, then the function $\phi(\mathbf{v})$ is a convex measure of group sparsity. We refer to problem (3) as group-sparse total variation (GS-TV) denoising.

2.1. Majorization-Minimization Algorithm

We use majorization-minimization (MM) as in [28] to derive a computationally efficient, fast converging, algorithm to minimize $F(\mathbf{x})$. Using (2), the penalty function $\phi(\mathbf{v})$ can be written as

$$\phi(\mathbf{v}) = \sum_n \|\mathbf{v}_{n,K}\|_2. \tag{5}$$

To find a majorizor of $F(\mathbf{x})$ defined in (3), we first find a majorizor of $\phi(\mathbf{v})$. To this end, note that

$$\frac{1}{2\|\mathbf{u}\|_2}\|\mathbf{v}\|_2^2 + \frac{1}{2}\|\mathbf{u}\|_2 \geq \|\mathbf{v}\|_2 \tag{6}$$

for all $\mathbf{v}$ and $\mathbf{u} \neq \mathbf{0}$ with equality when $\mathbf{u} = \mathbf{v}$. Using (6) for each group, a majorizor of $\phi(\mathbf{v})$ is given by

$$g(\mathbf{v}, \mathbf{u}) = \frac{1}{2} \sum_n \left[\frac{1}{\|\mathbf{u}_{n,K}\|_2} \|\mathbf{v}_{n,K}\|_2^2 + \|\mathbf{u}_{n,K}\|_2 \right]$$

with

$$g(\mathbf{v}, \mathbf{u}) \geq \phi(\mathbf{v}), \qquad g(\mathbf{u}, \mathbf{u}) = \phi(\mathbf{u}) \tag{7}$$

provided $\|\mathbf{u}_{n,K}\|_2 \neq 0$ for all n. Note that $g(\mathbf{v}, \mathbf{u})$ is quadratic in $\mathbf{v}$. It can be written as

$$g(\mathbf{v}, \mathbf{u}) = \frac{1}{2}\mathbf{v}^T \mathbf{\Lambda}(\mathbf{u})\, \mathbf{v} + C \tag{8}$$

where C does not depend on $\mathbf{v}$, and where $\mathbf{\Lambda}(\mathbf{u})$ is a diagonal matrix given, after some manipulations, by

$$[\mathbf{\Lambda}(\mathbf{u})]_{n,n} = \sum_{j=0}^{K-1} \left[\sum_{k=0}^{K-1} |u(n-j+k)|^2 \right]^{-1/2}. \tag{9}$$

The entries of $\mathbf{\Lambda}$ can be easily computed by point-wise squaring, a K-point moving sum, point-wise square-root, and a second K-point moving sum. Using (8), a majorizor of $F(\mathbf{x})$ is given by

$$\begin{aligned} G(\mathbf{x}, \mathbf{u}) &= \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\, g(\mathbf{Dx}, \mathbf{Du}) && (10) \\ &= \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{\lambda}{2}\mathbf{x}^T \mathbf{D}^T \mathbf{\Lambda}(\mathbf{Du})\, \mathbf{Dx} + \lambda C, && (11) \end{aligned}$$

i.e.,

$$G(\mathbf{x}, \mathbf{u}) \geq F(\mathbf{x}), \qquad G(\mathbf{u}, \mathbf{u}) = F(\mathbf{u}). \tag{12}$$

To minimize $F(\mathbf{x})$, the majorization-minimization (MM) approach defines an iterative algorithm via:

$$\mathbf{x}^{(i+1)} = \operatorname*{argmin}_{\mathbf{x}}\; G(\mathbf{x}, \mathbf{x}^{(i)})$$

where i is the iteration index. The iteration is initialized with some $\mathbf{x}^{(0)}$. Here, the MM iteration gives

$$\mathbf{x}^{(i+1)} = \operatorname*{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\, \mathbf{x}^T \mathbf{D}^T \mathbf{\Lambda}(\mathbf{Dx}^{(i)})\, \mathbf{Dx}, \tag{13}$$

which has the solution

$$\mathbf{x}^{(i+1)} = \Big(\mathbf{I} + \lambda \mathbf{D}^T \mathbf{\Lambda}(\mathbf{Dx}^{(i)})\, \mathbf{D} \Big)^{-1} \mathbf{y} \tag{14}$$

where the diagonal matrix $\mathbf{\Lambda}(\mathbf{Dx}^{(i)})$ depends on $\mathbf{Dx}^{(i)}$ per (9).

Note that some of the diagonal entries of $\mathbf{\Lambda}(\mathbf{Dx}^{(i)})$ generally go to infinity as $\mathbf{Dx}^{(i)}$ becomes sparse. This is a source of numerical inaccuracy in the update equation (14). To avoid this issue, we use the matrix inverse lemma as suggested in [14]. Using the matrix inverse lemma, the inverse in (14) can be written as

$$\begin{aligned} &\Big(\mathbf{I} + \lambda \mathbf{D}^T \mathbf{\Lambda}(\mathbf{Dx}^{(i)})\, \mathbf{D} \Big)^{-1} \\ &\qquad = \mathbf{I} - \mathbf{D}^T \left(\frac{1}{\lambda} \mathbf{\Lambda}^{-1}(\mathbf{Dx}^{(i)}) + \mathbf{DD}^T \right)^{-1} \mathbf{D}. \end{aligned} \tag{15}$$

Using (15), the update (14) can be written as

$$\mathbf{x}^{(i+1)} = \mathbf{y} - \mathbf{D}^T \Big(\underbrace{\frac{1}{\lambda} \mathbf{\Lambda}^{-1}(\mathbf{Dx}^{(i)}) + \mathbf{DD}^T}_{\text{banded}} \Big)^{-1} \mathbf{Dy}. \tag{16}$$

Algorithm: GS-TV denoising

input: $\mathbf{y}$, K, λ

1. $\mathbf{x} = \mathbf{y}$ (initialization)
2. $\mathbf{b} = \mathbf{D}^T \mathbf{y}$

repeat

3. $\mathbf{u} = \mathbf{D}\mathbf{x}$
4. $[\mathbf{\Lambda}]_{n,n} = \sum_{j=0}^{K-1} \left[\sum_{k=0}^{K-1} |u(n-j+k)|^2 \right]^{-1/2}$
5. $\mathbf{F} = \frac{1}{\lambda} \mathbf{\Lambda}^{-1} + \mathbf{D}\mathbf{D}^T$ ($\mathbf{F}$ is tridiagonal)
6. $\mathbf{x} = \mathbf{y} - \mathbf{D}^T (\mathbf{F}^{-1} \mathbf{b})$ (use fast solver)

until convergence

return: $\mathbf{x}$

Table 1: Group-sparse total variation (GS-TV) denoising algorithm.

The update equation (16) constitutes an iterative algorithm for solving the group-sparse total variation (GS-TV) denoising problem (3). The algorithm[1] is summarized in Table 1.

Note that the update equation (16) requires the solution to a large system of linear equations. However, the system matrix is banded (in fact, tridiagonal), hence the solution can be computed with high computational efficiency [25, Sect 2.4]. Note that the algorithm requires no user parameters (no step size parameters, etc.).

Convergence. Due to the derivation of the GS-TV algorithm using majorization-minimization, it is guaranteed that the cost function decreases at each iteration. However, the convergence of the algorithm to the minimizer is not so easily proven due to the 'singularity issue' that is known to arise in algorithms of this general form [13]. In [13, 22], the singularity issue and convergence of this type of MM algorithm was analyzed in detail and it was found that, with suitable initialization, the singularity issue generally does not hinder the convergence.

In the GS-TV algorithm (Table 1), the singularity issue may arise if it is not properly taken into account. Specifically, if an entire group of $\mathbf{u}$ equals zero, then the calculation of $\mathbf{\Lambda}(\mathbf{u})$ in (9) results in a 'divide-by-zero'. For this reason, it is important that the algorithm be initialized with a vector, $\mathbf{u}^{(0)} = \mathbf{D}\mathbf{x}^{(0)}$, for which all groups are non-zero. In case a divide-by-zero does occur during the course of the algorithm in the calculation of $[\mathbf{\Lambda}(\mathbf{u})]_{n,n}$ for some n, then it is suitable to assign a value of 'infinity' to that entry of $\mathbf{\Lambda}$. Note that $\mathbf{\Lambda}$ is subsequently used in the algorithm as $\mathbf{\Lambda}^{-1}$ only. Hence, $[\mathbf{\Lambda}(\mathbf{u})]_{n,n}^{-1}$ should be set to zero in this case. Once a group becomes equal to zero on some iteration, then it will remain zero for all subsequent iterations. This phenomenon ('zero-locking') is also recognized in algorithms of this general form [13]. However, it does not usually hinder the convergence of the GS-TV algorithm because this algorithm gradually reduces values to zero, rather than thresholds values directly to zero.

While the MM procedure was used to derive the proposed algo-

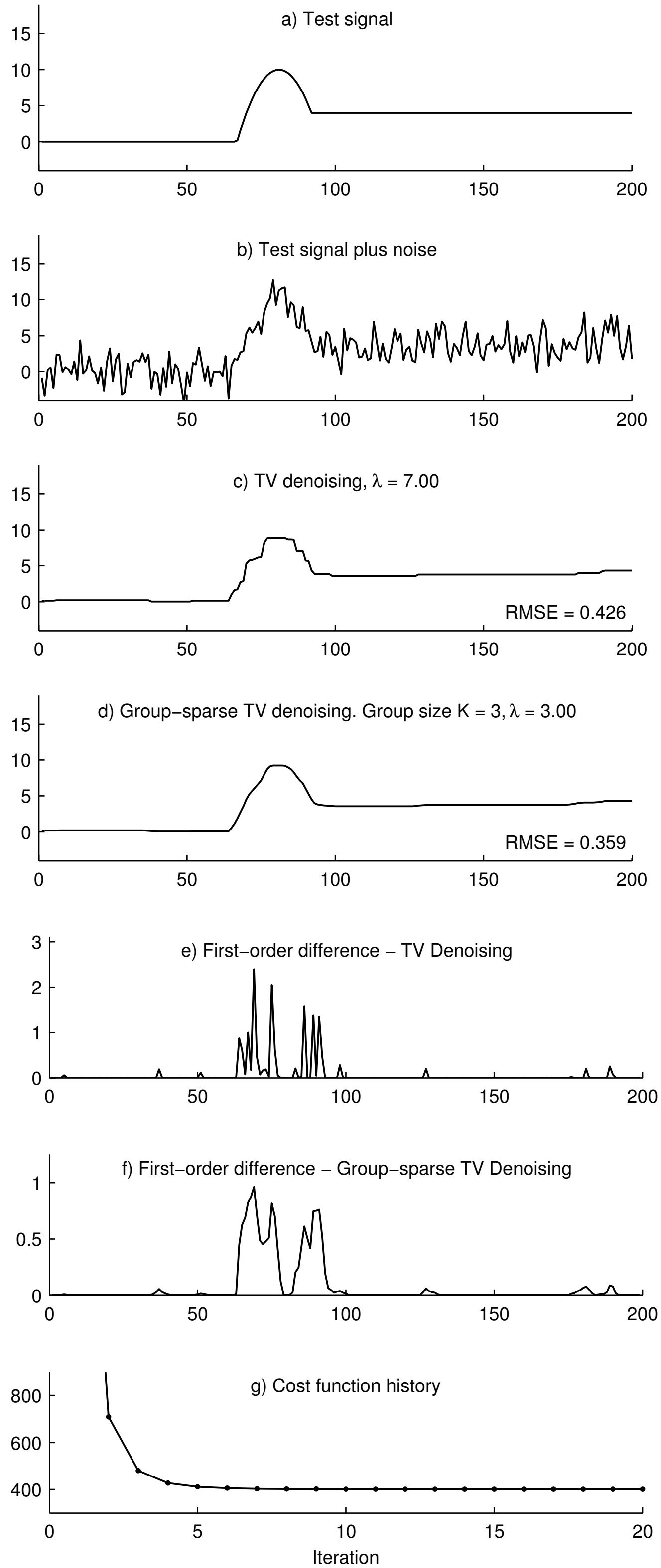

Fig. 1: Example 1. (a) Noise-free signal, (b) noisy data, (c) TV denoising, and (d) group-sparse TV denoising. First-order difference function for (e) TV denoising and (f) group-sparse TV denoising.

[1]A MATLAB implementation of the algorithm is available online at `http://eeweb.poly.edu/iselesni/gstv`. The MATLAB program uses sparse matrix structures so that fast solvers for banded systems are invoked by default.

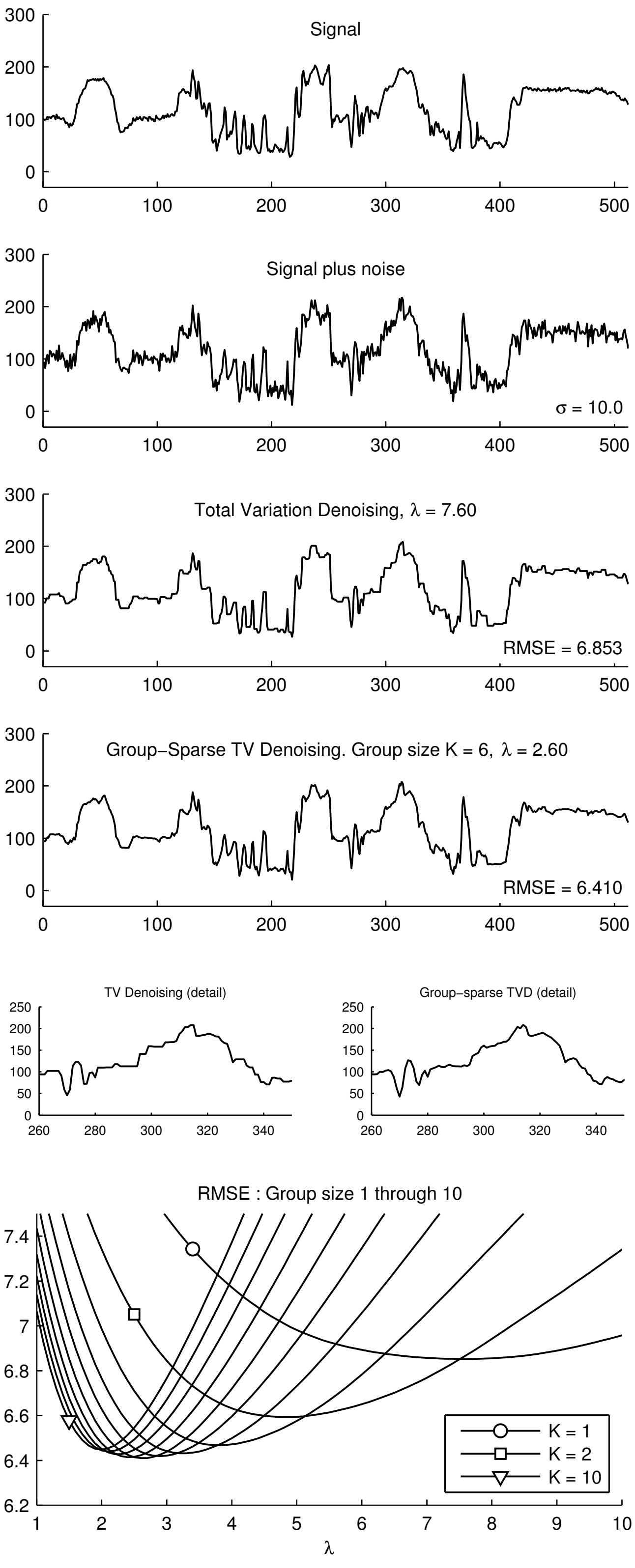

Fig. 2: Example 2: TV and group-sparse TV denoising. The signal is row 256 of the 'lena' image (512×512). Group-sparse TV denoising exhibits fewer stair-case artifacts and has improved RMSE compared with TV denoising.

rithm, other optimization approaches for signal restoration and general linear inverse problems may lead to equivalent algorithms, e.g., half-quadratic minimization [20].

2.2. Examples

Example 1. Figure 1 illustrates group-sparse total variation denoising. The simple synthetic test signal consists of a second order polynomial segment and two constant-valued intervals, illustrated in Fig. 1a. The signal is corrupted by additive white Gaussian noise. The result of TV denoising, illustrated in Fig. 1c, exhibits stair-case artifacts in the polynomial segment, as expected of TV denoising. Group-sparse TV denoising, shown in Fig. 1d, substantially reduces the stair-case behavior, and yields a smaller root-mean-square-error (RMSE). Here, the group size was set to $K = 3$.

To highlight the distinction between the two denoised signals, Figs. 1e and f show the absolute value of the first-order difference function, $|\mathbf{Dx}|$, for each of the two solutions. Total variation promotes sparsity of $|\mathbf{Dx}|$ but does not promote any grouping or clustering tendencies — large values in Fig. 1e are adjacent to small values. In contrast, large values in Fig. 1f are generally adjacent to other large values. The group-sparse penalty function has the effect of smoothing the sparse derivative signal.

The rapid convergence of the algorithm is illustrated in Fig. 1g, which shows the cost function value, $F(\mathbf{x}^{(i)})$, as a function of iteration i.

Example 2. Figure 2 illustrates group-sparse TV denoising on a single row of a standard test image (row 256 of 'lena'). Compared to TV, group-sparse TV leads to a result with less artificial blockiness in the denoising signal. (The signal around $n = 300$ is shown in detail.) In addition, GS-TV reduces the RMSE compared with TV denoising (6.41 compared with 6.85). To examine the effect of group size K and regularization parameter λ, we computed the RMSE as a function of λ for group sizes from 1 through 10. The result is illustrated in Fig. 2 (bottom). It can be seen that the minimal RMSE is obtained for group size $K = 6$ and $\lambda = 2.6$. These are the values used to illustrate GS-TV in the figure. Note that the GS-TV solution is not totally different from the TV solution — they are both based on sparsity of the first-order difference function.

3. CONCLUSION

This paper describes an extension to total variation denoising wherein it is assumed that the first-order difference function of the unknown signal is not only sparse, but also exhibits a basic form of structured sparsity: large values of the first-order difference function are not expected to occur in isolation. It is intended that this approach alleviates the staircase (blocking) artifact often arising in total variation based solutions. A convex cost function is given and an iterative algorithm is derived using majorization-minimization. The algorithm is both fast converging and computationally efficient due to the use of fast solvers for banded systems. On the whole, TV and GS-TV are not totally dissimilar, hence GS-TV is expected to retain the effectiveness of TV in sparse signal processing applications, such as compressed sensing, etc. As noted by a reviewer, a question remains regarding how suitable parameters K and λ can be chosen based on minimal knowledge of the signal characteristics.

Further extensions of this work are of interest. Group-sparse TV denoising can also be performed for images and multidimensional data. Non-convex penalty functions can be used so as to enhance group-sparsity. We are currently developing these extensions.

4. REFERENCES

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Structured sparsity through convex optimization. Technical report, Hal-00621245, 2011.

[2] I. Bayram. Mixed norms with overlapping groups as signal priors. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, pages 4036–4039, May 2011.

[3] J. Bect, L. Blanc-Féaud, G. Aubert, and A. Chambolle. A l^1-unified variational framework for image restoration. In T. Pajdla and J. Matas, editors, *European Conference on Computer Vision, Lecture Notes in Computer Sciences*, volume 3024, pages 1–13, 2004.

[4] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM J. Imag. Sci.*, 3(3):492–526, 2010.

[5] A. Chambolle. An algorithm for total variation minimization and applications. *J. of Math. Imaging and Vision*, 20:89–97, 2004.

[6] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.

[7] T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. Image Process.*, 10(2):231–241, February 2001.

[8] R. Chartrand and V. Staneva. Total variation regularisation of images corrupted by non-Gaussian noise using a quasi-Newton method. *Image Processing, IET*, 2(6):295–303, December 2008.

[9] P.-Y. Chen and I. W. Selesnick. Translation-invariant shrinkage/thresholding of group sparse signals. 2013. Preprint, http://eeweb.poly.edu/iselesni/ogs/.

[10] L. Condat. A direct algorithm for 1D total variation denoising. Technical report, Hal-00675043, 2012. http://hal.archives-ouvertes.fr/.

[11] W. Deng, W. Yin, and Y. Zhang. Group sparse optimization by alternating direction method. *Rice University CAAM Technical Report TR11-06, 2011*, 2011.

[12] M. Figueiredo and J. Bioucas-Dias. An alternating direction algorithm for (overlapping) group regularization. In *Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, 2011.

[13] 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.

[14] M. Figueiredo, J. Bioucas-Dias, J. P. Oliveira, and R. D. Nowak. On total-variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavelet denoising. In *Proc. IEEE Int. Conf. Image Processing*, 2006.

[15] Y. Hu and M. Jacob. Higher degree total variation (HDTV) regularization for image recovery. *IEEE Trans. Image Process.*, 21(5):2559–2571, May 2012.

[16] R. Jenatton, J.-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. *J. Mach. Learning Research*, 12:2777–2824, October 2011.

[17] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for sparse hierarchical dictionary learning. In *Proc. Int. Conf. on Machine Learning (ICML)*, 2010.

[18] F. I. Karahanoglu, I. Bayram, and D. Van De Ville. A signal processing approach to generalized 1-d total variation. *IEEE Trans. Signal Process.*, 59(11):5265–5274, November 2011.

[19] S.-H. Lee and M. G. Kang. Total variation-based image noise reduction with generalized fidelity function. *IEEE Signal Processing Letters*, 14(11):832–835, November 2007.

[20] M. Nikolova and M. K. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM J. Sci. Comput.*, 27(3):937–966, October 2005.

[21] G. Obozinski, L. Jacob, and J. P. Vert. Group lasso with overlaps: the latent group lasso approach. Technical report, HAL-Inria-00628498, 2011.

[22] 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.

[23] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation based image restoration. *Multiscale Model. & Simul.*, 4(2):460–489, 2005.

[24] G. Peyre and J. Fadili. Group sparsity with overlapping partition functions. In *Proc. European Sig. Image Proc. Conf. (EUSIPCO)*, Aug. 29 - Sept. 2 2011.

[25] 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.

[26] P. Rodriguez and B. Wohlberg. Efficient minimization method for a generalized total variation functional. *IEEE Trans. Image Process.*, 18(2):322–332, February 2009.

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

[28] I. Selesnick. Total variation denoising (an MM algorithm). *Connexions Web site*, 2012. http://cnx.org/content/m44934/1.2/.

[29] J.-L. Starck, Y. Moudden, J. Bobina, M. Elad, and D.L. Donoho. Morphological component analysis. In *Proceedings of SPIE*, volume 5914 (Wavelets XI), 2005.

[30] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc., Ser. B*, 58(1):267–288, 1996.

[31] L. A. Vese and S. Osher. Image denoising and decomposition with total variation minimization and oscillatory functions. *J. Math. Imag. Vis.*, 20:7–18, 2004.

[32] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. on Imaging Sciences*, 1(3):248–272, 2008.

[33] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM J. Imag. Sci.*, 1(1):143–168, 2008.

[34] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, 68(1):49–67, February 2006.

[35] M. Zhu, S. J. Wright, and T. F. Chan. Duality-based algorithms for total-variation-regularized image restoration. *J. Comp. Optimization and Applications*, 2008.

```
function [x, cost] = gstvd(y, K, lam, Nit)
% [x, cost] = gstvd(y, K, lam, Nit)
% Group-Sparse Total Variation Denoising.
%
% INPUT
%   y - noisy signal
%   K - group size (small positive integer)
%   lam - regularization parameter (lam > 0)
%   Nit - number of iterations
%
% OUTPUT
%   x - denoised signal
%   cost - cost function history

% Ivan Selesnick, selesi@poly.edu, 2012

y = y(:);                                               % Convert to column vector
cost = zeros(1, Nit);                                   % Cost function history
N = length(y);

e = ones(N-1, 1);
DDT = spdiags([-e 2*e -e], [-1 0 1], N-1, N-1);      % D*D' (sparse matrix)
D = @(x) diff(x);                                       % D (operator)
DT = @(x) [-x(1); -diff(x); x(end)];                    % D' (operator)

h = ones(1,K);                                          % For convolution
x = y;                                                  % Initialization
Dx = D(x);
Dy = D(y);

for k = 1:Nit
    r = sqrt(conv(abs(Dx).^2, h));
    cost(k) = 0.5*sum(abs(x-y).^2) + lam * sum(r);  % cost function value

    v = conv(1./r, h, 'valid');
    F = (1/lam) * spdiags(1./v, 0, N-1, N-1) + DDT; % F : Sparse matrix structure
    % F = (1/lam) * diag(1./v) + DDT;               % Not stored as sparse matrix

    x = y - DT(F\Dy);                               % Solve banded linear system
    Dx = D(x);
end
```