Total Variation Denoising with Overlapping Group Sparsity

Ivan W. Selesnick and Po-Yu Chen

Polytechnic Institute of New York University

Brooklyn, New York

selesi@poly.edu

POLYTECHNIC INSTITUTE OF NYU

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 rarely 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.

Total variation denoising

Data model: signal plus noise

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \in \mathbb{R}^N$$

$\mathbf{x}$: derivative of $\mathbf{x}$ is sparse
$\mathbf{w}$: white Gaussian noise

Definition of total variation (TV) denoising:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \sum_n |y(n) - x(n)|^2 + \lambda \sum_n |x(n) - x(n-1)| \tag{1}$$

Total variation denoising (TVD) is suitable for piecewise-constant signals (i.e. signals with a sparse derivative function).

For 1-D TV denoising, the exact solution can be obtained by a direct algorithm[1].

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

Applications of total variation[12] (TV)

1. denoising[2,3,4]
2. deconvolution[5,6,7]
3. reconstruction[8]
4. nonlinear decomposition[9,10]
5. compressed sensing[11].

---

[2] A. Chambolle. "An algorithm for total variation minimization and applications". In: *J. of Math. Imaging and Vision* 20 (2004), pp. 89–97.

[3] L. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D* 60 (1992), pp. 259–268.

[4] R. Chartrand and V. Staneva. "Total variation regularisation of images corrupted by non-Gaussian noise using a quasi-Newton method". In: *Image Processing, IET* 2.6 (Dec. 2008), pp. 295–303. ISSN: 1751-9659. DOI: 10.1049/iet-ipr:20080017.

[5] J. Oliveira, J. Bioucas-Dias, and M. A. T. Figueiredo. "Adaptive total variation image deblurring: A majorization-minimization approach". In: *Signal Processing* 89.9 (Sept. 2009), pp. 1683–1693. DOI: doi:10.1016/j.sigpro.2009.03.018.

[6] S. Osher et al. "An Iterative Regularization Method for Total Variation Based Image Restoration". In: *Multiscale Model. & Simul.* 4.2 (2005), pp. 460–489.

[7] J. Bect et al. "A l^1-Unified Variational Framework for Image Restoration". In: *European Conference on Computer Vision, Lecture Notes in Computer Sciences*. Ed. by T. Pajdla and J. Matas. Vol. 3024. 2004, pp. 1–13.

[8] Y. Wang et al. "A new alternating minimization algorithm for total variation image reconstruction". In: *SIAM J. on Imaging Sciences* 1.3 (2008), pp. 248–272.

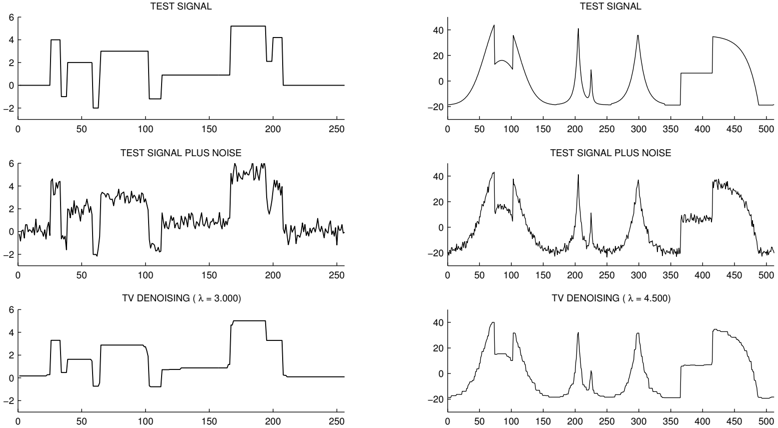
[9] L. A. Vese and S. Osher. "Image denoising and decomposition with total variation minimization and oscillatory functions". In: *J. Math. Imag. Vis.* 20 (2004), pp. 7–18.

[10] J.-L. Starck et al. "Morphological component analysis". In: *Proceedings of SPIE*. Vol. 5914 (Wavelets XI). 2005.

[11] W. Yin et al. "Bregman Iterative Algorithms for ℓ_1-Minimization with Applications to Compressed Sensing". In: *SIAM J. Imag. Sci.* 1.1 (2008), pp. 143–168. DOI: 10.1137/070703983. URL: http://link.aip.org/link/?SII/1/143/1.

[12] L. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D* 60 (1992), pp. 259–268.

Staircase artifacts



Staircase artifacts

TV denoising works well for piecewise constant signals.

But for piecewise smooth signals, TV denoising produces stair-case artifacts.

Several generalizations of TVD have been developed to address the staircase artifact[13][14][15][16][17].

[13] P. Rodriguez and B. Wohlberg. "Efficient Minimization Method for a Generalized Total Variation Functional". In: *IEEE Trans. Image Process.* 18.2 (Feb. 2009), pp. 322–332. ISSN: 1057-7149. DOI: 10.1109/TIP.2008.2008420.

[14] Y. Hu and M. Jacob. "Higher Degree Total Variation (HDTV) Regularization for Image Recovery". In: *IEEE Trans. Image Process.* 21.5 (May 2012), pp. 2559–2571. ISSN: 1057-7149. DOI: 10.1109/TIP.2012.2183143.

[15] K. Bredies, K. Kunisch, and T. Pock. "Total generalized variation". In: *SIAM J. Imag. Sci.* 3.3 (2010), pp. 492–526. DOI: 10.1137/090769521. eprint: http://epubs.siam.org/doi/pdf/10.1137/090769521. URL: http://epubs.siam.org/doi/abs/10.1137/090769521.

[16] F. I. Karahanoglu, I. Bayram, and D. Van De Ville. "A Signal Processing Approach to Generalized 1-D Total Variation". In: *IEEE Trans. Signal Process.* 59.11 (Nov. 2011), pp. 5265–5274. ISSN: 1053-587X. DOI: 10.1109/TSP.2011.2164399.

[17] S.-H. Lee and M. G. Kang. "Total Variation-Based Image Noise Reduction With Generalized Fidelity Function". In: *IEEE Signal Processing Letters* 14.11 (Nov. 2007), pp. 832–835. ISSN: 1070-9908. DOI: 10.1109/LSP.2007.901697.

Introduction

We propose *group sparse* total variation: the signal derivative exhibits *group* sparsity.

Signal model: derivative of $x(t)$ is sparse and and large values of the derivative are rarely isolated (i.e., large values usually arise near, or adjacent to, other large values).

1. Problem formulated via convex sparse optimization
2. Group/clustering of the signal derivative promoted by suitable penalty function
3. Group locations unknown
4. Translation-invariant denoising

Algorithm:

1. Majorization-minimization (MM) optimization method.[18]
2. Fast algorithm for groups sparse-TVD uses fast solvers for banded systems.
3. No algorithm parameters

---

[18] M. Figueiredo, J. Bioucas-Dias, and R. Nowak. "Majorization-minimization algorithms for wavelet-based image restoration". In: *IEEE Trans. Image Process.* 16.12 (Dec. 2007), pp. 2980–2991.

Notation

An N-point signal $x(n)$ is denoted

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

The first-order difference of an N-point signal $\mathbf{x}$ is given by $\mathbf{Dx}$ where

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & & \ddots & \\ & & & -1 & 1 \end{bmatrix}. \tag{2}$$

$\mathbf{D}$ is $(N-1) \times N$

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

$$\mathbf{v}_{n,K} = [v(n), \, \ldots, \, v(n+K-1)] \in \mathbb{R}^K. \tag{3}$$

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

Group-sparse total variation (GS-TV) denoising

Signal plus noise

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \in \mathbb{R}^N$$

$\mathbf{x}$: derivative of $\mathbf{x}$ is group sparse
$\mathbf{w}$: white Gaussian noise

$$\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathbb{R}^N} \left\{ F(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\, \phi(\mathbf{D}\mathbf{x}) \right\} \tag{4}$$

$\mathbf{v} = \mathbf{D}\mathbf{x} \in \mathbb{R}^{(N-1)}$,

To promote group sparsity, define $\phi : \mathbb{R}^{(N-1)} \to \mathbb{R}$,

$$\phi(\mathbf{v}) = \sum_n \left[\sum_{k=0}^{K-1} |v(n+k)|^2 \right]^{1/2}. \tag{5}$$

K : group size.

If $K = 1$, then $\phi(\mathbf{v}) = \|\mathbf{v}\|_1$ and problem (4) is the standard 1D total variation.

We refer to problem (4) as group-sparse total variation (GS-TV) denoising.

Majorization-minimization (MM) algorithm

We use MM to derive a computationally efficient, fast converging, algorithm to minimize $F(\mathbf{x})$.

Using (3), penalty $\phi(\mathbf{v})$ is

$$\phi(\mathbf{v}) = \sum_n \|\mathbf{v}_{n,K}\|_2. \tag{6}$$

To find a majorizor of $F(\mathbf{x})$ defined in (4), we first find a majorizor of $\phi(\mathbf{v})$.

First, note

$$\frac{1}{2\|\mathbf{u}\|_2}\|\mathbf{v}\|_2^2 + \frac{1}{2}\|\mathbf{u}\|_2 \geq \|\mathbf{v}\|_2, \quad \forall \mathbf{u} \neq \mathbf{0}. \tag{7}$$

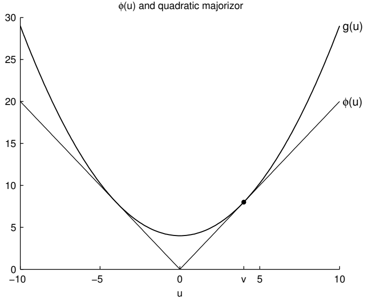
Using (7) for each group, a majorizor of $\phi(\mathbf{v})$ is given by

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

with

$$g(\mathbf{v}, \mathbf{u}) \geq \phi(\mathbf{v}), \qquad g(\mathbf{u}, \mathbf{u}) = \phi(\mathbf{u}) \tag{8}$$

provided $\|\mathbf{u}_{n,K}\|_2 \neq 0$ for all n.



φ(u) and quadratic majorizor
g(u)
φ(u)
30
25
20
15
10
5
0
–10
–5
0
v
5
10
u

Note that $g(\mathbf{v}, \mathbf{u})$ is quadratic in $\mathbf{v}$. It can be written as

$$g(\mathbf{v}, \mathbf{u}) = \frac{1}{2} \mathbf{v}^T \mathbf{\Lambda}(\mathbf{u}) \, \mathbf{v} + C \tag{9}$$

C : does not depend on $\mathbf{v}$
$\mathbf{\Lambda}(\mathbf{u})$: diagonal matrix. After some manipulations,

$$[\mathbf{\Lambda}(\mathbf{u})]_{n,n} = \sum_{j=0}^{K-1} \left[\sum_{k=0}^{K-1} |u(n-j+k)|^2 \right]^{-1/2}. \tag{10}$$

The entries of $\mathbf{\Lambda}$ are easily computed.

Using (9), a majorizor of $F(\mathbf{x})$ is given by

$$G(\mathbf{x}, \mathbf{u}) = \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \, g(\mathbf{D}\mathbf{x}, \mathbf{D}\mathbf{u}) \tag{11}$$

$$= \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{\lambda}{2} \mathbf{x}^T \mathbf{D}^T \mathbf{\Lambda}(\mathbf{D}\mathbf{u}) \, \mathbf{D}\mathbf{x} + \lambda C, \tag{12}$$

i.e.,

$$G(\mathbf{x}, \mathbf{u}) \geq F(\mathbf{x}), \qquad G(\mathbf{u}, \mathbf{u}) = F(\mathbf{u}). \tag{13}$$

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

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

where i is the iteration index.

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

$$\mathbf{x}^{(i+1)} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \, \mathbf{x}^T \mathbf{D}^T \mathbf{\Lambda}(\mathbf{D}\mathbf{x}^{(i)}) \, \mathbf{D}\mathbf{x}, \tag{14}$$

which has the solution

$$\boxed{\mathbf{x}^{(i+1)} = \Big(\mathbf{I} + \lambda \mathbf{D}^T \mathbf{\Lambda}(\mathbf{D}\mathbf{x}^{(i)}) \, \mathbf{D}\Big)^{-1} \mathbf{y}} \tag{15}$$

where the diagonal matrix $\mathbf{\Lambda}(\mathbf{D}\mathbf{x}^{(i)})$ depends on $\mathbf{D}\mathbf{x}^{(i)}$ per (10).

Problem: some diagonal entries of $\mathbf{\Lambda}(\mathbf{D}\mathbf{x}^{(i)})$ go to infinity as $\mathbf{D}\mathbf{x}^{(i)}$ becomes sparse (divide by zero).

Solution[19]: use the matrix inverse lemma (MIL)

$$\left(\mathbf{I} + \lambda \mathbf{D}^T \mathbf{\Lambda}(\mathbf{D}\mathbf{x}^{(i)})\, \mathbf{D}\right)^{-1} = \mathbf{I} - \mathbf{D}^T \left(\frac{1}{\lambda}\mathbf{\Lambda}^{-1}(\mathbf{D}\mathbf{x}^{(i)}) + \mathbf{D}\mathbf{D}^T\right)^{-1} \mathbf{D}. \qquad (16)$$

Using (16), update (15) becomes

$$\mathbf{x}^{(i+1)} = \mathbf{y} - \mathbf{D}^T \Big(\underbrace{\frac{1}{\lambda}\mathbf{\Lambda}^{-1}(\mathbf{D}\mathbf{x}^{(i)}) + \mathbf{D}\mathbf{D}^T}_{\text{banded}} \Big)^{-1} \mathbf{D}\mathbf{y}. \qquad (17)$$

Equation (17) constitutes an algorithm for GS-TV denoising (4).

The large system matrix in (17) is *banded* (in fact, tridiagonal). Therefore, fast solvers for banded systems[20] can be used.

The algorithm requires no user parameters (no step size parameters, etc.).

[19] M. Figueiredo et al. "On total-variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavelet denoising". In: *Proc. IEEE Int. Conf. Image Processing*. 2006.

[20] W. H. Press et al. *Numerical recipes in C: the art of scientific computing (2nd ed.)* Cambridge University Press, 1992. ISBN: 0-521-43108-5Sect 2.4.

Group-sparse total variation (GS-TV) denoising algorithm

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: $\mathbf{x}$

Example 1

TV and group-sparse TV denoising.

- Simple synthetic test signal + white Gaussian noise
- TV denoising: stair-case artifacts
- GS-TV denoising ($K = 3$): much less stair-case behavior, smaller root-mean-square-error (RMSE)

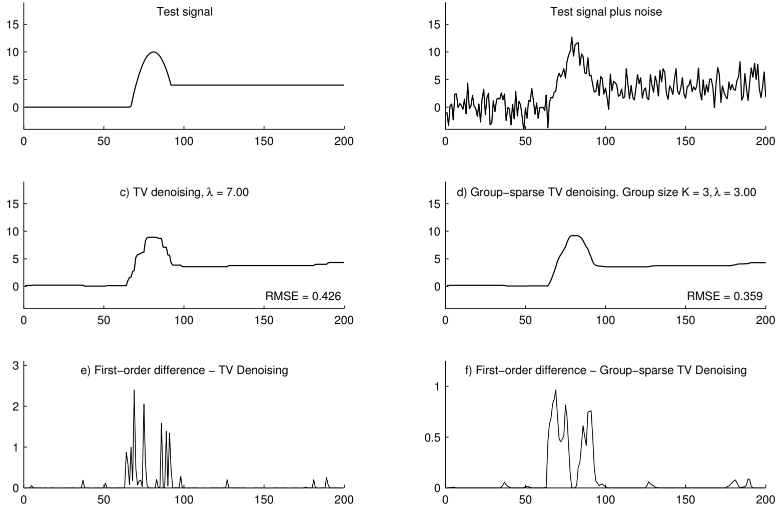
TV promotes sparsity of $|\mathbf{Dx}|$ but does not promote any grouping or clustering tendencies (see figure).

GS-TV promotes group sparsity of $|\mathbf{Dx}|$: large values are adjacent to other large values (see figure).

The group-sparse penalty function smooths the sparse derivative signal.

The algorithm converges rapidly. The cost function monotonically decreases.

Example 1

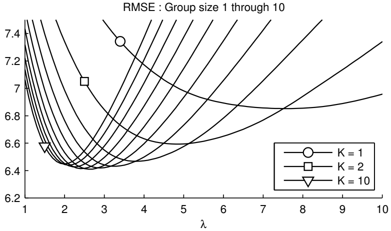


Test signal
Test signal plus noise
c) TV denoising, λ = 7.00
RMSE = 0.426
d) Group−sparse TV denoising. Group size K = 3, λ = 3.00
RMSE = 0.359
e) First−order difference − TV Denoising
f) First−order difference − Group−sparse TV Denoising

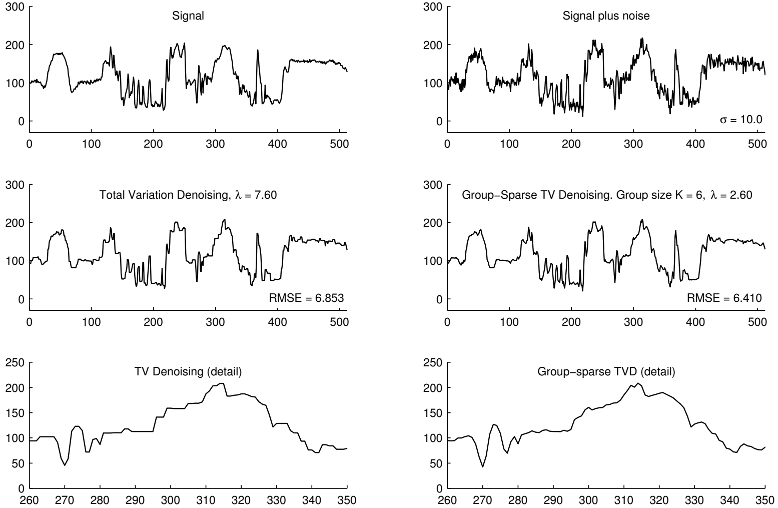
Example 2

TV and group-sparse TV denoising.

1. The signal is row 256 of the 'lena' image (512×512).
2. Group-sparse TV denoising gives less artificial blockiness (The signals around $n = 300$ is shown in detail.)
3. Group-sparse TV denoising has improved RMSE (6.41 compared with 6.85).
4. 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 minimal RMSE is obtained for group size $K = 6$ and $\lambda = 2.6$.



Example 2



Signal
Signal plus noise
σ = 10.0
Total Variation Denoising, λ = 7.60
RMSE = 6.853
Group−Sparse TV Denoising. Group size K = 6, λ = 2.60
RMSE = 6.410
TV Denoising (detail)
Group−sparse TVD (detail)

Conclusion

1. GS-TV extends TV denoising to signals wherein the first-order difference function is not only sparse, but also exhibits a basic form of structured sparsity: large values of the first-order difference function rarely occur in isolation.
2. It is intended that this approach alleviates the staircase (blocking) artifact often arising in total variation based solutions.
3. A convex cost function is proposed and a fast converging computationally efficient algorithm is derived. The algorithm harnesses fast solvers for banded systems.
4. How should a suitable parameters K and λ be chosen based on minimal knowledge of the signal characteristics?
5. Group-sparse TV denoising for images ...
6. Non-convex penalty functions for enhanced group-sparsity ...

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