GENERALIZED TOTAL VARIATION DENOISING VIA AUGMENTED LAGRANGIAN CYCLE SPINNING WITH HAAR WAVELETS

Ulugbek Kamilov, Emrah Bostan, and Michael Unser

Biomedical Imaging Group, EPFL, Lausanne, Switzerland.

ABSTRACT

We consider the denoising of signals and images using regularized least-squares method. In particular, we propose a simple minimization algorithm for regularizers that are functions of the discrete gradient. By exploiting the connection of the discrete gradient with the Haar-wavelet transform, the n-dimensional vector minimization can be decoupled into n scalar minimizations. The proposed method can efficiently solve total-variation (TV) denoising by iteratively shrinking shifted Haar-wavelet transforms. Furthermore, the decoupling naturally lends itself to extensions beyond ℓ_1 regularizers.

Index Terms— signal denoising, soft-thresholding, cycle spinning, TV denoising

1. INTRODUCTION

Two popular signal-denoising methods, wavelet shrinkage [1] and total-variation (TV) denoising [2], have been shown to be effective in reducing noise, while preserving important signal features such as edges. Both methods serve the same purpose and rely on the concept of sparsity, either in the wavelet or in the finite-differences domains. In fact, in the 1-D case and for Haar wavelets, TV denoising has been shown to be closely related to wavelet shrinkage on a single scale with cycle spinning [3, 4].

The contribution of the present paper to image denoising is twofold: Our first contribution is to uncover the connection between cycle-spinning Haar denoising and TV denoising, both for the 1-D and 2-D cases. The observation is that TV denoising can be reformulated as a constrained minimization problem in the Haar-wavelet domain by making the transform shift-invariant. We then solve the equivalent constrained optimization problem via the augmented Lagrangian method [5]. Our technique decouples the n-dimensional optimization problem into n scalar problems, each of which can be solved directly. Our second contribution is to extend our algorithm to general, possibly non-convex, potential functions. By utilizing the decoupling, we can solve the high-dimensional problem by precomputing the solution of the scalar subproblems.

This work was supported by the European Commission under Grant ERC-2010-AdG 267439-FUN-SP.

2. SIGNAL DENOISING

2.1. Conventional Cycle Spinning

Consider the standard problem of estimating an unknown discrete-signal $\mathbf{x} \in \mathbb{R}^n$ from measurements $\mathbf{y} = \mathbf{x} + \mathbf{e}$ where $\mathbf{e}$ is a Gaussian white noise. In the simple *wavelet shrinkage* [1], the signal $\mathbf{x}$ is represented as $\mathbf{x} = \mathbf{W}^{-1}\mathbf{w}$, where the matrix $\mathbf{W}$ represents an orthogonal wavelet transform. Then, the estimation is performed by shrinking the wavelet coefficients as

$$\hat{\mathbf{w}} = \eta\left(\mathbf{W}\mathbf{y}; \lambda\right), \tag{1}$$

where the soft-thresholding operator η is applied componentwise and defined as $\eta\left(x; \lambda\right) = \max\left(|x| - \lambda, 0\right)\mathrm{sgn}(x)$. The final signal estimate is then computed as $\hat{\mathbf{x}} = \mathbf{W}^{-1}\hat{\mathbf{w}}$. The scalar parameter (or threshold) λ is called the regularization parameter and has to be chosen suitably [1]. If the wavelet representation $\mathbf{w}$ has both coarse and detail coefficients, the shrinkage is performed on detail coefficients only.

Basic wavelet-shrinkage methods perform fairly well; however, they can be further improved by applying *cycle spinning* [4]. Consider an orthogonal transform $\mathbf{W}$ to be periodically shift-invariant with integer period K and define $\mathbf{W}_k = \mathbf{W}\mathbf{S}_k$ to be the transform of a shifted signal, where $\mathbf{S}_k$ represents shifting by $k \in [1 \dots K]$. Cycle spinning works by using periodic shift-invariance of the wavelet transform to minimize artifacts by denoising wavelet representations of the shifted versions of the signal. Wavelet regularization with cycle spinning is often used in denoising and inverse problems and consistently improves performance at no extra cost [6, 7]. Here, this type of cycle spinning will be referred to as *conventional cycle spinning* to distinguish it from our method.

2.2. Total Variation

The idea of TV regularization is to penalize the total variation of the signal, which in the discrete case results in the minimization problem

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^n}{\mathrm{argmin}} \left\{ \tfrac{1}{2}\|\mathbf{x} - \mathbf{y}\|_{\ell_2}^2 + \lambda \|\mathrm{L}\mathbf{x}\|_{\ell_1} \right\}, \tag{2}$$

where λ is a scalar parameter to tune and $\mathrm{L}\mathbf{x}$ is the vector containing the magnitude of the discrete gradient of $\mathbf{x} \in \mathbb{R}^n$

at all pixels, defined as

$$[\mathrm{L}\mathbf{x}]_i = \begin{cases} |x_{i+1} - x_i| & \text{1-D signals} \\ \sqrt{(x_{i+m} - x_i)^2 + (x_{i+1} - x_i)^2} & \text{2-D signals,} \end{cases}$$

where we assume that the image of size $(m \times l)$ has been stored in a column-major order. In this paper, we assume isotropic TV regularization, but the concept is easily applicable to the anisotropic case as well. Besides substantial noise reduction, another motivation to use TV denoising is the existence of fast algorithms such as FISTA [8] for solving (2).

2.3. Connection to the Haar-Wavelet

The *Haar-wavelet* is one of the lazy and most popular wavelet transforms. For a given signal, the 1-level Haar transform produces coefficients

$$u_i = \tfrac{1}{\sqrt{2}}\left(x_{2i-1} + x_{2i}\right),\ v_i = \tfrac{1}{\sqrt{2}}\left(x_{2i-1} - x_{2i}\right), \quad (3)$$

where $i \in [1 \dots n/2]$, assuming that n is even. The coefficients $\mathbf{u}$ and $\mathbf{v}$ are referred to as coarse and detail coefficients, respectively. The Haar-wavelet expansion is orthonormal. Hence, it satisfies *Parseval's theorem* $\|\mathbf{x}\|_{\ell_2}^2 = \|\mathbf{w}\|_{\ell_2}^2 = \|\mathbf{u}\|_{\ell_2}^2 + \|\mathbf{v}\|_{\ell_2}^2$.

By comparing equations for the discrete gradient of the signal $\mathbf{x}$ at some pixel i with the Haar-wavelet expansion (3), we can guess that TV denoising is related to wavelet shrinkage. In fact, it has been shown in [3] that, for 1-D signals, Haar-wavelet shrinkage with conventional cycle spinning is equivalent to a single iteration of TV denoising. In this work, we would like to take advantage of this connection to derive a new fast method for generalized TV denoising.

3. PROPOSED METHOD

3.1. Problem Formulation

We would like to find an efficient solution to determine

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \|\mathbf{x} - \mathbf{y}\|_{\ell_2}^2 + \lambda\phi\left(\mathrm{L}\mathbf{x}\right) \right\}, \quad (4)$$

where the role of the regularizer with the general potential function ϕ is to enforce some desired features in the final solution. The TV-regularized least-squares objective function in (2) is a special case of the function in (4) when $\phi\left(\cdot\right) = \|\cdot\|_{\ell_1}$. When ϕ is non-convex, it is still possible to derive an iterative optimization algorithm that converges, although it is not guaranteed to reach the global optimum anymore.

3.2. One-Dimensional Signals

Let us first solve the problem for 1-D signals, since exactly the same reasoning applies to 2-D. Consider two complementary shifted Haar-wavelet transforms (only one level is considered)

$$\mathbf{w}_k = \mathbf{W}_k\mathbf{x}, \quad (5)$$

where $k \in \{1, 2\}$. Then, by using Parseval's theorem, (4) is expressed as the equivalent constrained minimization problem

$$\begin{aligned} \min_{\mathbf{w}_1,\mathbf{w}_2\in\mathbb{R}^n} \quad & G\left(\mathbf{w}_1, \mathbf{w}_2\right) = F\left(\mathbf{w}_1\right) + F\left(\mathbf{w}_2\right) \\ \text{subject to} \quad & \mathbf{W}_1^{-1}\mathbf{w}_1 = \mathbf{W}_2^{-1}\mathbf{w}_2. \end{aligned} \quad (6)$$

There, the function F is defined as

$$F\left(\mathbf{w}_i\right) = \tfrac{1}{2}\|\mathbf{w}_i - \mathbf{W}_i\mathbf{y}\|_{\ell_2}^2 + \lambda\sum_j \phi\left(\sqrt{2}|w_{ij}|\right). \quad (7)$$

We have deliberately omitted the summation range for j. The reason for this is that the penalization is only applied to the detail (or highpass) Haar coefficients. Note that Problem (6) is equivalent to (4) since, in the feasible region $\left\{(\mathbf{w}_1, \mathbf{w}_2) : \mathbf{W}_1^{-1}\mathbf{w}_1 = \mathbf{W}_2^{-1}\mathbf{w}_2\right\}$, the cost functions coincide with one another.

Since a constrained minimization problem such as (6) is difficult to handle directly, we replace it with an equivalent unconstrained minimization problem. One traditional technique for doing this is the *augmented Lagrangian method* [5]. To apply it, we replace the objective function in (6) with the new penalty function

$$\begin{aligned} \mathcal{L}(\mathbf{w}_1, \mathbf{w}_2, \boldsymbol{\eta}, \theta) =& G\left(\mathbf{w}_1, \mathbf{w}_2\right) \\ &+ \tfrac{\theta}{2}\left\|\mathbf{W}_1^{-1}\mathbf{w}_1 - \mathbf{W}_2^{-1}\mathbf{w}_2\right\|_{\ell_2}^2 \\ &- \boldsymbol{\eta}^T\left(\mathbf{W}_1^{-1}\mathbf{w}_1 - \mathbf{W}_2^{-1}\mathbf{w}_2\right), \end{aligned} \quad (8)$$

where $\theta > 0$ is called the *penalty parameter* and $\boldsymbol{\eta} \in \mathbb{R}^n$ is the vector of Lagrange multipliers. The method works by minimizing $\mathcal{L}$ with respect to $(\mathbf{w}_1, \mathbf{w}_2)$ while keeping θ fixed and updating $\boldsymbol{\eta}$ according to the simple rule

$$\boldsymbol{\eta}^{t+1} = \boldsymbol{\eta}^t - \theta\left(\mathbf{W}_1^{-1}\mathbf{w}_1^{t+1} - \mathbf{W}_2^{-1}\mathbf{w}_2^{t+1}\right), \quad (9)$$

where $t > 0$ is the iteration number.

To solve the minimization of the functional in (8), we apply *alternating direction method of multipliers* (ADMM) [9] and alternate between solving the denoising problem in one wavelet basis with the other one fixed. By alternating the wavelet bases, the minimization will be reduced to n-scalar minimizations of the form

$$\mathcal{T}_\phi\left(w_d; \lambda_d\right) = \min_{w\in\mathbb{R}}\left\{\tfrac{1}{2}(w - w_d)^2 + \lambda_d\phi\left(\sqrt{2}\,|w|\right)\right\}, \quad (10)$$

where w_d is function of $\mathbf{y}, \boldsymbol{\eta}, \theta$, and $\mathbf{w}_k$(the basis we have fixed). Similarly, the regularization parameter rescales to $\lambda_d = \lambda/\left(\theta + 1\right)$. In the case of TV denoising, Problem (10) admits a closed-form solution $\eta(w_d; \sqrt{2}\lambda_d)$. For the general potential function ϕ, the solution can be precomputed and stored in a lookup table. We call the resulting algorithm *Augmented Lagrangian Cycle Spinning (ALCS)*.

As can be seen, the algorithm is conceptually similar to conventional cycle spinning: it denoises in each of the wavelet

Algorithm Augmented Lagrangian Cycle Spinning (ALCS)

input : $\mathbf{y}$
choose : $\lambda, \theta > 0; \boldsymbol{\eta}^0, \hat{\mathbf{x}}_2^0$
set : $t = 0, \mu = 1/(1+\theta)$
repeat
  $\hat{\mathbf{x}}_1^{t+1} = \mathbf{W}_1^{-1}\mathcal{T}_\phi\left(\mu\mathbf{W}_1(\mathbf{y} + \theta\hat{\mathbf{x}}_2^t + \boldsymbol{\eta}^t); \mu\lambda\right)$
  $\hat{\mathbf{x}}_2^{t+1} = \mathbf{W}_2^{-1}\mathcal{T}_\phi\left(\mu\mathbf{W}_2(\mathbf{y} + \theta\hat{\mathbf{x}}_1^{t+1} - \boldsymbol{\eta}^t); \mu\lambda\right)$
  $\boldsymbol{\eta}^{t+1} = \boldsymbol{\eta}^t - \theta\left(\hat{\mathbf{x}}_1^{t+1} - \hat{\mathbf{x}}_2^{t+1}\right)$
  $t = t + 1$
until stopping criterion
return $\hat{\mathbf{x}} = \hat{\mathbf{x}}_2^{t+1}$

Note : The scalar function $\mathcal{T}_\phi$ is applied component-wise only on the detail coefficients.

bases and combines the solutions. However, our algorithm avoids the oscillations that were observed for example in [7], due to changing wavelet bases, since we are constraining our solutions to be consistent across shifted bases. Furthermore, as we have mentioned previously, the algorithm decouples the minimization to n-scalar shrinkages, which results in a final algorithmic complexity of $O(n)$ per iteration.

3.3. Two-Dimensional Signals

For 2-D signals, we proceed in exactly the same fashion: (a) express discrete gradient in terms of Haar-coefficients; (b) by using Parseval's theorem, derive constrained minimization problem equivalent to (4); (c) using the augmented-Lagrangian method, change the constrained problem into an unconstrained one; (d) solve unconstrained problem by simple shrinkage rules, either in closed form, or stored as a lookup table.

To represent the discrete gradient $\mathrm{D}_i\mathbf{x}$ at every pixel i, we need four orthogonal 1-D Haar transforms (two for each dimension of the gradient). Note that a 1-D Haar representation of a 2-D signal can be computed by applying the transform either column-wise or row-wise. Let us denote these transforms as $\mathbf{w}_k = \mathbf{W}_k\mathbf{x}$ for $k = 1, 2, 3, 4$. To simplify our notation, let us reorganize all the wavelet coefficients into a single $\tilde{\mathbf{w}}$, such that, for the detail coefficients, we have

$$\mathrm{D}_i\mathbf{x} \equiv \begin{pmatrix} x_{i+m} - x_i \\ x_{i+1} - x_i \end{pmatrix} = \sqrt{2}\tilde{\mathbf{w}}_i. \tag{11}$$

We denote this transform as $\tilde{\mathbf{w}} = \tilde{\mathbf{W}}\mathbf{x}$. Then, by using Parseval's theorem, we write an equivalent constrained minimization problem as

$$\begin{aligned} \min_{\tilde{\mathbf{w}}} \quad & G\left(\tilde{\mathbf{w}}\right) \\ \text{subject to} \quad & \mathbf{W}_k^{-1}\mathbf{w}_k = \mathbf{W}_l^{-1}\mathbf{w}_l,\ \forall k, l. \end{aligned} \tag{12}$$

The function G is given by

$$G\left(\tilde{\mathbf{w}}\right) = \tfrac{1}{4}\left\|\tilde{\mathbf{w}} - \tilde{\mathbf{W}}\mathbf{y}\right\|_{\ell_2}^2 + \lambda\sum_i \phi\left(\sqrt{2}\left\|\tilde{\mathbf{w}}_i\right\|_{\ell_2}\right), \tag{13}$$

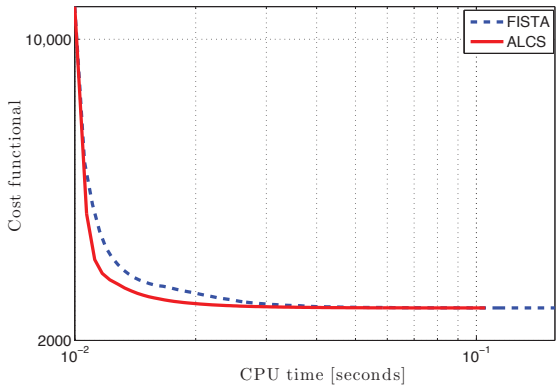

Fig. 1: Total-variation cost-functional evolution with respect to CPU time for FISTA and ALCS.

where we penalize only detail coefficients. Then, we can proceed as before and write out the corresponding augmented-Lagrangian objective function, which we omit due to space limitations. The augmented Lagrangian then can be minimized by alternating with respect to pair of bases, which will result in n-shrinkages over vectors in $\mathbb{R}^2$. For example, for TV denoising, the shrinkage admits the closed form

$$\mathcal{S}\left(\mathbf{x}; \lambda\right) = \max\left\{\|\mathbf{x}\|_{\ell_2} - \lambda, 0\right\}\frac{\mathbf{x}}{\|\mathbf{x}\|_{\ell_2}}, \tag{14}$$

where $\mathbf{x} \in \mathbb{R}^2$. As in the 1-D case, for a general ϕ we store the shrinkage into a lookup table. Note that, since the method amounts to recursively shrinking wavelet coefficients in different bases, its algorithmic complexity is $O(n)$ per iteration.

4. EXPERIMENTS

To investigate the performance of our method, we conducted three simple numerical experiments on a Mac Pro computer with 2.66 GHz Quad-Core Intel Xeon processor. All the schemes were implemented in MATLAB.

In the first experiment, we compare the performance of ALCS and the state-of-the-art FISTA, in terms of CPU time to reach the same value of the TV objective function. We consider a 1-D signal that follows the innovation model in [10] and generate a compound Poisson process (piecewise-constant signal) for which TV regularization is particularly well tailored. We show in Figure 1 the plots of the TV objective function for a signal of length 4000, generated with $\mathrm{Prob}(x = 0) = 0.9$ and a Gaussian amplitude distribution of variance 1. The noise-free signal is degraded by AWGN with zero mean and unit variance. The regularization parameters were optimized for both methods with the reference signal acting as the oracle. It was observed that the speed of ALCS is comparable to that of FISTA.

In Figure 2 , we consider the denoising of 1-D sparse Lévy processes with increments that are Cauchy distributed [11]. A MAP estimator for the process can be designed by using the potential function $\phi(\mathrm{L}\mathbf{x}) = \log((\mathrm{L}\mathbf{x})^2 + 1)$. ALCS equipped

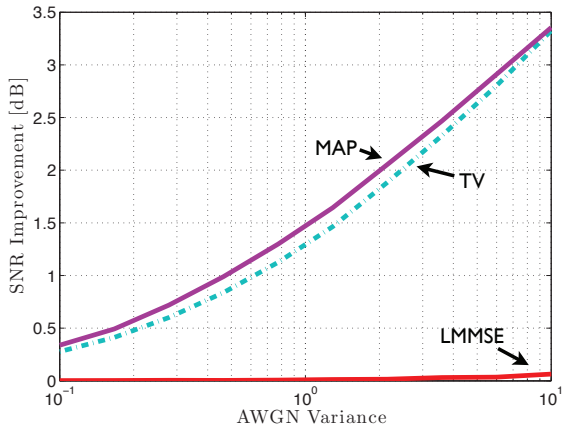

Fig. 2: Comparison of different estimators (noise variance vs. SNR improvement) for a Cauchy process. See text for a description of the experiment.

with a lookup table is used to solve this non-convex optimization problem. In the experiment, we compare three estimators: log-regularizer (MAP), TV, and LMMSE. For different noise levels, 100 realizations are made. For each realization, regularization parameters are optimized and SNR improvements are recorded. Algorithms are allowed to run for 200 iterations. As expected, the MAP estimator outperforms TV and Wiener filter reconstructions over the whole range of noise variances.

In Figure 3, we denoise the Shepp-Logan phantom, which has a sparse gradient. We use the log-potential function $\phi(\mathrm{L}\mathbf{x}) = \log((\mathrm{L}\mathbf{x})^2 + \epsilon)$ which is more sparsity-encouraging than the ℓ_1 norm [12]. The original image is corrupted by AWGN with zero mean and a variance of 0.01. We use ALCS to solve TV and log regularization problems where ϵ is set to 10^{-3}. Both methods are allowed to run for 500 iterations without any stopping criterion with optimal λ parameters. As expected, the log-regularizer achieves significant improvements over TV for this type of signal.

5. CONCLUSION

We presented an augmented-Lagrangian-based cycle-spinning algorithm for signal-denoising problems. We further showed that the method is able to perform TV denoising where we represented the discrete gradient operator by 1-level Haar wavelet transforms. The resulting algorithm involves the straightforward combination of Haar wavelet transforms and shrinkage steps so that it is computationally cheap if shrinkage is in closed form or stored in a lookup table. We also showed that the algorithm can deal with different potential functions in general and can be applied to a variety of convex and non-convex optimization problems.

6. REFERENCES

[1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

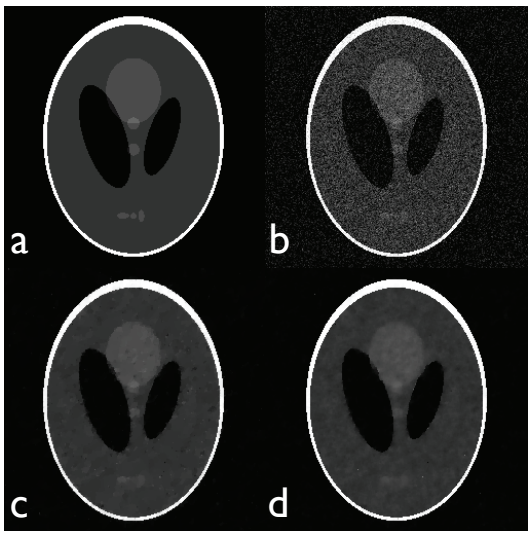

Fig. 3: Denoising of the noisy Shepp-Logan phantom: (a) original image; (b) noisy image, SNR = 7.85 dB; (c) TV regularization, SNR = 20.93 dB; (d) Log regularization, SNR = 22.59 dB

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[3] G. Steidl, J. Weickert, T. Brox, P. Mrazek, and M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs," *SIAM J. of Numerical Analysis*, vol. 42, no. 2, pp. 686–713, 2004.

[4] R. R. Coifman and D. L. Donoho, *Springer Lecture Notes in Statistics*, chapter Translation-invariant de-noising, pp. 125–150, Springer-Verlag, 1995.

[5] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 2^{nd} edition, 2006.

[6] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Signal Process.*, vol. 12, no. 8, pp. 906–916, 2003.

[7] M. Guerquin-Kern, M. Häberlin, K. P. Prüssmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Trans. Medical Imaging*, vol. 30, no. 9, pp. 1649–1660, 2011.

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

[9] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Prog.*, vol. 55, pp. 293–318, 1992.

[10] M Unser and P. D. Tafti, "Stochastic models for sparse and piecewise-smooth signals," *IEEE Trans. Signal Process.*, vol. 59, pp. 989–1006, 2011.

[11] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes–part I: Continuous-domain theory," *arXiv:1108.6150v1*.

[12] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, 2008, pp. 3869–3872.