Image Denoising with Compound Regularization Using a Bregman Iterative Algorithm

Manya V. Afonso*, José M. Bioucas-Dias, Mário A. T. Figueiredo
Instituto de Telecomunicações, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001, Lisboa, PORTUGAL
Phone: +351 218418467, Fax: + 351 218418472, E-mail: {mafonso,bioucas,mario.figueiredo}@lx.it.pt

Abstract

Some imaging inverse problems may require the solution to simultaneously exhibit properties that are not enforceable by a single regularizer. In this paper, we use regularization with linear combinations of simple regularizers, to encourage the solution to simultaneously exhibit the characteristics enforced by each. We apply the split Bregman iterative method to deal with the optimization problem resulting from addressing a linear inverse problem with compound regularization. The resulting algorithm only requires the ability to efficiently compute the denoising operator associated to each involved regularizer. Convergence is guaranteed by the theory behind the Bregman iterative approach to solving constrained optimization problems.

1 INTRODUCTION

Linear inverse problems involve estimating an unknown signal/image with certain characteristics (such as sparseness or piece-wise smoothness) enforced by a suitable regularizer. In several problems such as image denoising, image restoration [1, 2], image reconstruction, and compressed sensing [3, 4], the solution is defined as the minimizer of an objective function, leading to an optimization problem of the form

$$\hat{\mathbf{u}} \;=\; \arg\min_{\mathbf{u}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{f}\|_2^2 \;+\; \tau\Phi(\mathbf{u}), \qquad (1)$$

where $\mathbf{A}:\mathbb{R}^n\to\mathbb{R}^m$ is the linear observation (direct) operator, $\mathbf{f}\in\mathbb{R}^m$ is the observed data, $\Phi:\mathbb{R}^n\to\mathbb{R}$ is the regularizer function, and $\tau\in[0,+\infty[$ is the regularization parameter. If the operator $\mathbf{A}$ is the identity, (1) is a denoising problem, the solution of which is unique (if Φ is convex) and called the *Moreau proximal mapping* (MPM) of Φ [5]. For some choices of Φ, the MPM has a simple close form (*e.g.*, the well-known soft-threshold [6], if $\Phi(\mathbf{u}) = \|\mathbf{u}\|_1 = \sum_i |u_i|$) or can be efficiently computed. For non-diagonal operators, (1) has to be solved using an iterative algorithm, such as *iterative shrinkage-thresholding* (IST) [7], also known as forward-backward splitting [5], or the faster *two-step IST* (TwIST) [8], in which the MPM is iteratively applied.

In certain problems, it may be desirable to favor solutions that simultaneously exhibit properties that are enforced by two (or more) different regularizers. For example, *total variation* (TV) regularization [9] encourages piecewise smooth solutions, while an l_1 (or l_p, with $p\leq 1$) regularizer favors sparse solutions; however, there is no "simple" regularizer that favors both these characteristics simultaneously, as may be important in certain problems. To achieve this, compound regularizers (*i.e.*, linear combinations of "simple" regularizers [10]) must be used, leading to

$$\hat{\mathbf{u}} \;=\; \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{f}\|_2^2 \;+\; \tau_1\Phi_1(\mathbf{u}) \;+\; \tau_2\Phi_2(\mathbf{u}), \quad (2)$$

where $\Phi_1:\mathbb{R}^n\to\mathbb{R}$ and $\Phi_2:\mathbb{R}^n\to\mathbb{R}$ are the regularizers, with respective regularization parameters $\tau_1>0$ and $\tau_2>0$. An iterative algorithm for solving (2) (easily generalizable to more than two regularizers) has been recently proposed [10]; that approach involves a constrained optimization formulation of (2) followed by minimization of the associated Lagrangian using a block-coordinate descent algorithm. A similar formulation, specifically tailored for regularizers which can be written as ℓ_1 norms (such as the ℓ_1 norm itself and TV) was also very recently proposed [11]; in that work, the constrained problem is attacked using a so-called split Bregman method.

In this paper, we propose an approach for solving problems of the form (2) involving any regularizers for which the MPMs are known (not just ℓ_1 norms). As in [10], the approach involves a constrained optimization formulation of (2), which is then directly addressed using a Bregman iterative method [12]. In [13], we have illustrated this approach in the problem of deconvolving an image which is known to have a few white blobs on a black background; such an image is characterized by having a low ℓ_1 norm (it's mostly black, *i.e.*, sparse) and a low TV norm (it's piecewise flat). Using a combination of ℓ_1 and TV regularizers, we have shown the ability of the algorithm to solve the resulting problem and also that the resulting estimates have lower MSE than what can be achieved using each of the two regularizers alone. In that problem, the ℓ_1 regularizer was used so as to minimize the ℓ_1 norm of the image pixel values. In this paper, we demonstrate this approach in the problem of denoising an image which admits a sparse rep-

*The work of Manya Afonso is supported by the EU by a Marie-Curie Fellowship (EST-SIGNAL program : http://est-signal.i3s.unice.fr) under contract No MEST-CT-2005-021175.

resentation (such as a wavelet decomposition), and show that the estimate obtained using a combination of the ℓ_1 (applied to the wavelet coefficients) and TV regularizers again has a lower MSE than what can be achieved using each of the two regularizers alone.

2 PROPOSED METHOD

2.1 Bregman Iterations

We begin by very briefly reviewing the Bregman iterative approach for solving constrained problems of the form

$$\begin{aligned} \min_{\mathbf{x}} \quad & E(\mathbf{x}) \\ \text{subject to} \quad & H(\mathbf{x}) = 0, \end{aligned} \tag{3}$$

with E and H convex, H differentiable, and $\min_{\mathbf{x}} H(\mathbf{x}) = 0$ (see [11, 12], for more details). The Bregman divergence associated with E is defined as

$$D_E^{\mathbf{p}}(\mathbf{x}, \mathbf{y}) \equiv E(\mathbf{x}) - E(\mathbf{y}) - \langle \mathbf{p}, \mathbf{x} - \mathbf{y} \rangle, \tag{4}$$

where $\mathbf{p}$ belongs to the subgradient of E at $\mathbf{y}$ (*i.e.*, $\mathbf{p} \in \partial E(\mathbf{y})$). The Bregman iteration is given by

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg\min_{\mathbf{x}} D_E^{\mathbf{p}}(\mathbf{x}, \mathbf{x}^k) + \mu H(\mathbf{x}) \\ &= \arg\min_{\mathbf{x}} E(\mathbf{x}) - \langle \mathbf{p}^k, \mathbf{x} - \mathbf{x}^k \rangle + \mu H(\mathbf{x}), \end{aligned} \tag{5}$$

where $\mathbf{p}^k \in \partial E(\mathbf{x}^k)$; it has been shown that, for any $\mu > 0$, this procedure converges to a solution of (3) [11, 12]. Moreover, it can be shown that $\mathbf{p}^{k+1}$ should be chosen as

$$\mathbf{p}^{k+1} = \mathbf{p}^k - \mu \nabla H(\mathbf{x}^{k+1}). \tag{6}$$

2.2 Constrained Formulation

A constrained optimization problem equivalent to the unconstrained problem (2) is

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v}} \quad & \frac{1}{2}\|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2 \; + \; \tau_1 \Phi_1(\mathbf{u}) \; + \; \tau_2 \Phi_2(\mathbf{v}) \\ \text{subject to} \quad & \|\mathbf{u} - \mathbf{v}\|_2^2 = 0. \end{aligned} \tag{7}$$

One approach to handling this constrained problem is to consider its Lagrangian and minimize it using a block-descent algorithm [10]. However, an extremely large value of the Lagrange multiplier is required for the minimizer of the Lagrangian to closely approximate that of (7), causing numerical difficulties. The alternative herein proposed is to use a Bregman iterative method to directly solve (7).

Letting $\mathbf{x} = [\mathbf{u}^T \mathbf{v}^T]^T$, we can write (7) in the form $\min_{\mathbf{x}} E(\mathbf{x})$, subject to $H(\mathbf{x}) = 0$, where $H(\mathbf{x}) = \|\mathbf{u} - \mathbf{v}\|^2$ and $E(\mathbf{x})$ is the objective function in (7).

The split Bregman formulation for l_1-regularized problems, proposed in [11], separates the l_1 and l_2 portions of the energy in the problem

$$\min_{\mathbf{u}} \|\Phi(\mathbf{u})\|_1 + H(\mathbf{u}) \tag{8}$$

where $H(.)$ and $\Phi(.)$ are convex functionals and $\Phi(.)$ is differentiable, by introducing an additional variable $\mathbf{d} \in \mathbb{R}^n$ and the constraint $\mathbf{d} = \Phi(\mathbf{u})$. The constrained problem is formulated as

$$\begin{aligned} \min_{\mathbf{u},\mathbf{d}} \quad & \|\mathbf{d}\|_1 + H(\mathbf{u}) \\ \text{subject to} \quad & \|\mathbf{d} - \Phi(\mathbf{u})\|_2^2 \; = \; 0. \end{aligned} \tag{9}$$

Applying the Bregman iteration, it can be shown that this problem is solved by the two-phase algorithm

$$\begin{aligned} (\mathbf{u}^{k+1}, \mathbf{d}^{k+1}) &= \min_{\mathbf{u},\mathbf{d}} \|\mathbf{d}\|_1 + H(\mathbf{u}) + \\ & \quad + \frac{\lambda}{2}\|\mathbf{d} - \Phi(\mathbf{u}) - \mathbf{b}^k\|_2^2 \qquad (10) \\ \mathbf{b}^{k+1} &= \mathbf{b}^k + (\Phi(\mathbf{u}^{k+1}) - \mathbf{d}^{k+1}) \qquad (11) \end{aligned}$$

The problem (10) can be minimized efficiently by iteratively minimizing with respect to $\mathbf{u}$ and $\mathbf{d}$, in two steps

$$\begin{aligned} \mathbf{u}^{k+1} &= \min_{\mathbf{u}} H(\mathbf{u}) + \frac{\lambda}{2}\|\mathbf{d}^k - \Phi(\mathbf{u}) - \mathbf{b}^k\|_2^2, \qquad (12) \\ \mathbf{d}^{k+1} &= \min_{\mathbf{d}} \|\mathbf{d}\|_1 + \frac{\lambda}{2}\|\mathbf{d} - \Phi(\mathbf{u}^{k+1}) - \mathbf{b}^k\|_2^2. \qquad (13) \end{aligned}$$

The problem (8) is thus reduced to a sequence of unconstrained problems and Bregman updates. In (7), clubbing the data misfit term $\frac{1}{2}\|\mathbf{A}\mathbf{u} - \mathbf{f}\|^2$ and the regularizer term $\tau_1 \Phi_1(\mathbf{u})$ together, we can apply a similar approach, that is, iteratively minimizing with respect to $\mathbf{u}$ and $\mathbf{v}$, separately.

After some algebraic manipulations, we can show that the Bregman iteration for this problem has the form

$$\begin{aligned} \mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2 \; + \; \tau_1 \Phi_1(\mathbf{u}) \; + \\ & \quad + \frac{\mu}{2}\|\mathbf{u} - \mathbf{v}^k - \mathbf{b}^k\|_2^2 \\ &= \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{K}\mathbf{u} - \mathbf{g}\|_2^2 \; + \; \tau_1 \Phi_1(\mathbf{u}) \qquad (14) \\ \mathbf{v}^{k+1} &= \arg\min_{\mathbf{v}} \tau_2 \Phi_2(\mathbf{v}) \; + \; \frac{\mu}{2}\|\mathbf{u}^k - \mathbf{v} - \mathbf{b}^k\|_2^2 \\ &= \arg\min_{\mathbf{v}} \tau_2 \Phi_2(\mathbf{v}) \; + \; \frac{\mu}{2}\|\mathbf{u}^k - \mathbf{b}^k - \mathbf{v}\|_2^2 \qquad (15) \end{aligned}$$

$$\text{where } \mathbf{K}^T = \left[\mathbf{A}, \sqrt{\mu}\, \mathbf{I}_n\right], \; \mathbf{g} = \left[\mathbf{f}^T, \sqrt{\mu}(\mathbf{v}^k + \mathbf{b}^k)^T\right]^T,$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k - (\mathbf{u}^k - \mathbf{v}^k), \tag{16}$$

and the initial values are $\mathbf{u}^0 = \mathbf{0}$, $\mathbf{v}^0 = \mathbf{0}$, and $\mathbf{b}^0 = \mathbf{0}$.

Since each of the problems (14) and (15) involves only one regularizer, for which the MPM is known, they can be efficiently solved using, *e.g.*, the IST or TwIST algorithms [8]. As convergence is guaranteed for any value of $\mu > 0$, we can choose it so as to make these problems well-conditioned. The iterations can be terminated when the constraint term $\|\mathbf{u}^k - \mathbf{v}^k\|_2^2$ falls below some threshold and the relative change in the objective function in (7) goes below some tolerance level. The final value of either $\mathbf{u}^k$ or $\mathbf{v}^k$, after applying any inverse transform if applicable, is taken as the estimate of $\mathbf{u}$, $\hat{\mathbf{u}} = \mathbf{u}^{\text{final}}$

Figure 1: Original image

3 RESULTS

For the purpose of demonstration, we will consider a denoising problem (*i.e.*, $\mathbf{A}$, the matrix representation of a convolution, is simply an identity matrix), with the cameraman image shown in Fig. 1. This image has mostly piecewise smooth regions except for a bit of texture, and can admit a sparse representation through a wavelet decomposition. We therefore use a combination of the ℓ_1 and the TV regularizers. The ℓ_1 regularizer seeks to minimize the ℓ_1 norm of the wavelet coefficients. The MPM (denoising operator) for the TV regularizer is implemented as in [14], and the MPM for the ℓ_1 regularizer is the well-known soft threshold [6].

Figure 2: Image from Figure 1, contaminated with Gaussian noise (SNR = 20dB).

We use a four level Haar wavelet redundant decomposition. The optimal values of the regularization parameters (which yielded the lowest mean square error) were found to be $\tau_{\ell_1} = 0.5$ and $\tau_{\text{TV}} = 0.424\sigma^2$, where σ^2 is the variance of the noise. The best experimental value of μ, which led to convergence in as few as 2 or 3 Bregman iterations, was found to be 0.15.

The image with added zero mean Gaussian noise is shown in Fig. 2. The SNR in this case is 20 dB. The estimate obtained by the proposed method is shown in Fig. 3 and has MSE = 49.9. This corresponds to an ISNR = $10\log_{10}(\frac{\|\mathbf{u}-\mathbf{f}\|^2}{\|\mathbf{u}-\hat{\mathbf{u}}\|^2}) = 5.56$ dB, which is better than the best result using each of ℓ_1- (ISNR = 1.15 dB) and TV (ISNR = 5.47 dB) alone. The estimated images for ℓ_1- and TV regularization (for their respective best possible regularization paramter values), are respectively, shown in Fig. 5 and Fig. 4.

Figure 3: Image estimated using the compound regularizer (ℓ_1+TV), with MSE = 49.9 (ISNR = 5.56 dB)

.

Figure 4: Image estimated using the TV regularizer, with MSE = 51 (ISNR = 5.47 dB)

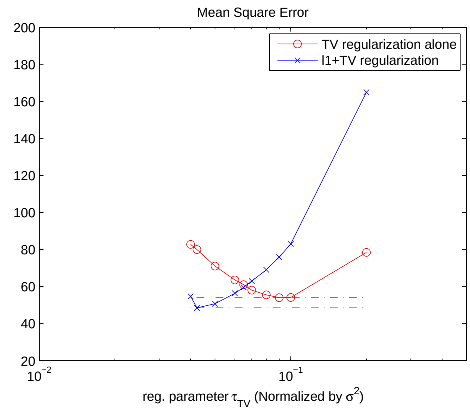
.

The plot of the MSE obtained with the compound regularizer (ℓ_1+TV) and with only TV regularization, for different values of the regularization parameter τ_{TV}, is shown in Fig. 6(a). For each value of τ_{TV}, the value of τ_{ℓ_1} that was used was found by hand-tuning to obtain the lowest MSE. It can be observed that the minimum MSE that can be obtained using ℓ_1+TV regularization is lower than the minimum obtained using TV alone. Figure 6(b) shows a similar MSE comparison between using only ℓ_1 regularization and ℓ_1+TV regularization. The value of τ_{TV} used was $0.424\sigma^2$, which was the optimal for all values of τ_{ℓ_1}. For any value of τ_{ℓ_1}, the MSE obtained with the ℓ_1+TV regularizer is lower than that obtained with ℓ_1 alone.

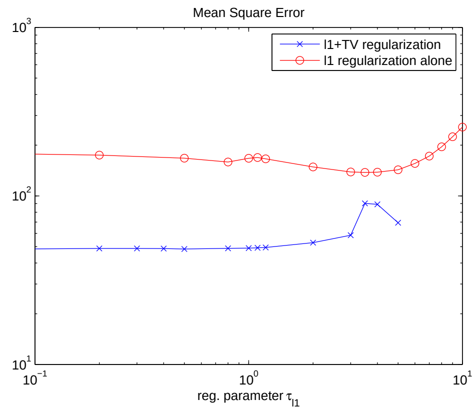
Figure 5: Image estimated using the ℓ_1 regularizer, with MSE = 138 (ISNR = 1.15 dB).

4 CONCLUDING REMARKS

We have introduced a new algorithm for solving the optimization problems resulting from using more than one regularizer in imaging inverse problems. The algorithm only requires the ability to efficiently compute the denoising operator associated to each involved regularizer. It was illustrated on a problem of image denoising, with encouraging results. The lowest MSE obtained using the ℓ_1+TV regularizer was lower than that obtained with TV or ℓ_1 regularization alone.



(a)



(b)

Figure 6: (a): MSE obtained with ℓ_1+TV and only TV regularizers, for different values of the regularization parameter τ_{TV}, (b): MSE obtained with ℓ_1+TV and only ℓ_1 regularizers, for different values of the regularization parameter τ_{ℓ_1}.

References

[1] M. Figueiredo and R.Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, pp. 906–916, 2003.

[2] J. Bioucas-Dias and M. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, pp. 2292–3004, 2007.

[3] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[4] J. Haupt and R.Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, pp. 4036–4048, 2006.

[5] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling & Simulation*, vol. 4, pp. 1168–1200, 2005.

[6] D. Donoho, "De-noising by soft thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[7] M. Figueiredo, J. Bioucas-Dias, and M. Figueiredo, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 16, pp. 2980–2991, 2007.

[8] J. Bioucas-Dias and M. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. on Image Processing*, vol. 16, no. 12, pp. 2992–3004, December 2007.

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

[10] J. Bioucas-Dias and M. Figueiredo, "An iterative algorithm for linear inverse problems with compound regularizers," in *IEEE International Conf. on Image Processing – ICIP 2008*, San Diego, CA, USA, October 2008.

[11] T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," *UCLA CAAM Report 08-29*, April 2008.

[12] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences (SIIMS)*, vol. 1, no. 1, pp. 143–168, December 2006.

[13] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "A Bregman iterative approach to linear inverse problems with compound regularizers," in *Signal Processing with Adaptive Sparse Structured Representations (SPARS'09)*, 2009, (Accepted).

[14] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vis.*, vol. 20, no. 1-2, pp. 89–97, 2004.