

# TWO-STEP ALGORITHMS FOR LINEAR INVERSE PROBLEMS WITH NON-QUADRATIC REGULARIZATION

José M. Bioucas-Dias

Mário A. T. Figueiredo

Instituto de Telecomunicações,  
Instituto Superior Técnico,  
Lisboa, Portugal

Email: {jose.bioucas, mario.figueiredo}@lx.it.pt

## ABSTRACT

Iterative shrinkage/thresholding (IST) algorithms have been recently proposed to handle high-dimensional convex optimization problems arising in image inverse problems (namely deconvolution) under non-quadratic regularization (e.g., total variation or sparsity inducing regularizers on wavelet representations). The convergence speed of IST algorithms depends heavily on the nature of the direct operator, being very slow when this operator is severely ill-conditioned. In this paper, we introduce a two-step version of IST (termed 2IST, pronounced “twist”) showing much faster convergence for strongly ill-conditioned operators. We give theoretical results concerning the convergence behavior of 2IST and show its effectiveness for wavelet-based and total variation image deconvolution.

## 1. INTRODUCTION

### 1.1. Problem Formulation

In an inverse problem, the goal is to estimate an unknown image  $\mathbf{x}$  from a (possibly noisy) observation  $\mathbf{y}$ , produced by an operator  $\mathbf{K}$  applied to  $\mathbf{x}$  [2]. In a *linear inverse problem* (LIP),  $\mathbf{K}$  is a linear operator; if  $\mathbf{K}$  represents a convolution with some kernel, the LIP is called a *deconvolution problem*.

Most approaches to LIPs define a solution  $\hat{\mathbf{x}}$  as a minimizer of an objective function  $f : \mathcal{X} \rightarrow \bar{\mathbb{R}} = [-\infty, +\infty]$ ,

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \lambda \Phi(\mathbf{x}), \quad (1)$$

where  $\mathbf{K} : \mathcal{X} \rightarrow \mathcal{Y}$  is the (linear) *direct operator*,  $\mathcal{X}$  and  $\mathcal{Y}$  are real Hilbert spaces,  $\Phi : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  is the so-called regularizer function,  $\lambda \in [0, +\infty[$  is the *regularization parameter*, and  $\|\cdot\|$  denotes the norm in the Hilbert space of its argument (see [2], for a comprehensive text on inverse problems and regularization in imaging). The intuitive meaning of  $f$  is clear: its minimizers reach a compromise between lack of fitness to the observed data (given by  $\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2$ ) and degree of “undesirability” (as assessed by  $\Phi(\mathbf{x})$ ). The parameter  $\lambda$  controls the relative weight of these two terms.

State-of-the-art regularizers for LIPs in imaging (e.g., total-variation (TV) [5], [8], [22] and wavelet-based regularization [7], [13], [20]) are non-differentiable. This fact, together with the huge dimension of  $\mathbf{x}$ , place  $f$  beyond the reach of off-the-shelf optimization algorithms and has stimulated research in recent years [3, 5, 10, 12, 14, 17, 18, 21].

### 1.2. Previous Work and Our Contributions

Recently, *iterative shrinkage/thresholding* (IST) algorithms, tailored for objective functions with the form (1), were proposed independently by several authors [12, 14, 15, 17, 18, 21]. Convergence of IST was studied in [12]; more recently IST was shown to belong to a class of so-called *forward-backward* algorithms, whose convergence was studied in [10].

Another class of algorithms, termed IRS (for *iterative reweighted shrinkage*), was proposed in [3]. IRS algorithms were shown to be much faster than IST in cases where  $\mathbf{K}$  is very ill-conditioned [3]. Conversely, for mildly ill-conditioned  $\mathbf{K}$ , and strong noise, IST is faster than IRS [16].

This paper introduces a new algorithm bringing together the best of IRS and IST. It’s a *two-step IST* (2IST, pronounced “twist”), in the sense that the update equation uses the two previous estimates, rather than only the previous one. We present a theorem giving sufficient conditions for the convergence of 2IST. The effectiveness of 2IST is experimentally confirmed on a set of image deblurring problems, using wavelet-based and TV regularization.

The next section reviews several choices of  $\Phi$  and (old and new) results on existence and uniqueness of minimizers of  $f$ . The previous IST and IRS algorithms are described in Section 3, which also reviews previous results on the convergence of IST. The new 2IST algorithm is introduced in Section 4, which also contains some theoretical results concerning its convergence behavior. Finally, Section 5 reports experimental results.

## 2. REGULARIZERS

### 2.1. Convex Regularizers and Denoising Functions

We adopt the following standard assumptions about the regularizer  $\Phi$ : it is convex, lower semi-continuous (lsc), and proper (see [23, 25] for details on these and other concepts and results in convex analysis).

A LIP in which  $\mathbf{K}$  is the identity, i.e.,  $\mathbf{K}\mathbf{x} = \mathbf{x}$ , is termed a denoising problem. In this case, the objective function (1) simplifies to  $f_{\text{den}} = (1/2)d_{\mathcal{Y}}^2 + \lambda \Phi$ , where

$$d_{\mathcal{Y}}(\mathbf{x}) = \|\mathbf{x} - \mathbf{y}\| \quad (2)$$

is lsc, proper, and coercive (that is,  $\lim_{\|\mathbf{x}\| \rightarrow \infty} \|\mathbf{y} - \mathbf{x}\|^2 = \infty$ ). Consequently,  $f_{\text{den}}$  is also lsc, proper, and coercive, thus its set of minimizers is not empty [10, 23]. The strict convexity of  $d_{\mathcal{Y}}^2$  implies strict convexity of  $f_{\text{den}}$ ; hence, it has a unique minimizer which

allows defining the *denoising function*  $\Psi_\lambda : \mathcal{X} \rightarrow \mathcal{X}$  as

$$\Psi_\lambda(\mathbf{y}) = \arg \min_{\mathbf{x}} \left\{ \frac{d_{\mathbf{y}}^2(\mathbf{x})}{2} + \lambda \Phi(\mathbf{x}) \right\}, \quad (3)$$

also known as the (Moreau) *proximal mapping* of  $\Phi$  [10, 23].

If  $\Phi(\alpha \mathbf{x}) = \alpha \Phi(\mathbf{x})$ , for all  $\alpha \geq 0$  and  $\mathbf{x} \in \mathcal{X}$ ,  $\Phi$  is said *positively homogeneous of degree 1* (phd-1) and the following result applies (see [5, 10, 19]):

**Theorem 1** *If  $\Phi$  is convex, lsc, proper, and phd-1, then*

$$\Psi_\lambda(\mathbf{y}) = \mathbf{y} - \mathbf{P}_{\lambda C}(\mathbf{y}), \quad (4)$$

where  $C \subset \mathcal{X}$  is a closed convex set depending on the regularizer  $\Phi$ , and  $\mathbf{P}_A : \mathcal{X} \rightarrow \mathcal{X}$  denotes the orthogonal projection operator onto the convex set  $A \subset \mathcal{X}$ .

We next list common classes of (convex, lsc, proper) regularizers and the corresponding denoising functions.

**Weighted  $\ell^p$  Norms:** These norms ( $p \geq 1$ ), defined as

$$\Phi_{\ell_w^p}(\mathbf{x}) = \|\mathbf{x}\|_{p, \mathbf{w}} = \left( \sum_i w_i |x_i|^p \right)^{1/p}, \quad (5)$$

where  $\mathbf{w} = [w_1, \dots, w_i, \dots]$ , with  $w_i \geq 0$ , are convex, lsc, proper, and phd-1, thus Theorem 1 applies. This class of norms appears in the discrete version of Besov norm regularizers in wavelet-based image restoration (see [7]). The denoising function under a  $\Phi_{\ell_w^p}$  regularizer can't, in general, be obtained in closed form. An exception is  $p = 1$ , for which  $\Psi_\lambda$  is the well-known soft-thresholding function [13].

**$p$ -th Power of Weighted  $\ell^p$  Norms:** This class of regularizers, defined as  $\Phi_{\ell_w^p}(\mathbf{x}) = \|\mathbf{x}\|_{p, \mathbf{w}}^p$ , appears in many wavelet-based approaches (see [3, 12, 17, 18, 19] and reference therein). For  $p = 1$ ,  $\Phi_{\ell_w^1} = \Phi_{\ell_w^1}$ , and Theorem 1 applies. For  $p > 1$ ,  $\Phi_{\ell_w^p}$  is not phd-1, and the denoising operator doesn't have the form (4). In this case,  $\Psi_\lambda$  is given by the component-wise application of the function  $S_{\lambda w_i, p}$ , where  $S_{\tau, p}$  is the inverse function of

$$F_{\tau, p}(x) = x + \tau p \operatorname{sign}(x)|x|^{p-1}. \quad (6)$$

For  $p > 1$ ,  $F_{\tau, p}$  is bijective, thus  $S_{\tau, p} = F_{\tau, p}^{-1}$  (called the *shrinkage function*) is well defined. This function has simple closed forms for  $p = 4/3$ ,  $p = 3/2$ , or  $p = 2$  [10]. Key features of  $S_{\tau, p}$  (for  $p > 1$ ) are: it's strictly monotonic, differentiable, and its derivative is upper bounded by 1 [12].

**Total Variation:** TV regularizers, denoted  $\Phi_{TV}$  (either the continuous formulation [5, 8, 22], or its discrete versions [5, 6]) have been shown to satisfy the hypotheses of Theorem 1; thus, TV denoising corresponds to the residual of a projection onto a convex set [5, 10, 19]. Although in this case, there is no closed form for this projection (*i.e.*, for TV denoising), fast iterative methods have been recently introduced [5, 6, 11].

## 2.2. Existence and Uniqueness of Solutions

Sufficient conditions for the existence/uniqueness of minimizers of (1), under regularizers  $\Phi_{\ell_w^p}$ ,  $\Phi_{\ell_w^p}$ , and  $\Phi_{TV}$  can be derived from results in [10] (Propositions 3.1 and 5.3). In particular, (i) using  $\Phi_{\ell_w^p}$

or  $\Phi_{\ell_w^p}$ , with  $w_i > 0$ ,  $\forall_i$ , the set of minimizer of (1) is nonempty; (ii) using  $\Phi_{\ell_w^p}$ , with  $p > 1$  and  $w_i > 0$ ,  $\forall_i$ , the minimizer is unique; (iii) with  $\Phi_{\ell_w^p}$  or  $\Phi_{\ell_w^p}$ , with  $p \geq 1$ , if  $\mathbf{K}$  is injective, the minimizer is unique; (iv) in the finite-dimensional case, if  $\mathbf{K}$  is injective the minimizer is unique.

With  $\Phi_{TV}$  and a non-injective  $\mathbf{K}$ , the results in [10] can not be applied to guarantee existence of a minimizer. The following result (proved in [4]) gives a sufficient condition for the existence of minimizer(s) under a discrete TV regularizer:

**Theorem 2** *Let  $\mathcal{X} = \mathbb{R}^m$  and  $f$  be given by (1), with  $\Phi = \Phi_{TV}$ . Then, if the null-space of  $\mathbf{K}$  doesn't contain the constant image,  $f$  has at least one minimizer.*

## 3. PREVIOUS ALGORITHMS

From this point on, we consider only finite-dimensional spaces,  $\mathcal{X} = \mathbb{R}^m$ ,  $\mathcal{Y} = \mathbb{R}^n$ . The class of IST algorithms has the form

$$\mathbf{x}_{t+1} = (1 - \beta) \mathbf{x}_t + \beta \Psi_\lambda \left( \mathbf{x}_t + \mathbf{K}^T (\mathbf{y} - \mathbf{K} \mathbf{x}_t) \right), \quad (7)$$

where  $\beta > 0$ . Convergence of IST, with  $\beta = 1$ , was first shown in [12]; a more general result was shown in [10]. The following is a simplified (namely, for the finite-dimensional case) version of Theorem 5.5 from [10]:

**Theorem 3** *Let  $f$  be given by (1), where  $\Phi$  is convex, and  $\|\mathbf{K}\|_2^2 < 2$ , where  $\|\mathbf{K}\|_2$  is the matrix norm induced by the  $\ell^2$  norm. Let  $G$ , the set of minimizers of  $f$ , be non-empty. Fix some  $\mathbf{x}_1$ ; then, the sequence  $\{\mathbf{x}_t, t \in \mathbb{N}\}$  produced by (7), with  $\beta \in ]0, 1]$ , converges to a point  $\mathbf{x} \in G$ .*

The iterations of the IRS algorithm are given by

$$\mathbf{x}_{t+1} = \text{solution} \{ \mathbf{A}_t \mathbf{x} = \mathbf{b} \}, \quad (8)$$

with  $\mathbf{b} = \mathbf{K}^T \mathbf{y}$  and  $\mathbf{A}_t = \lambda \mathbf{D}_t + \mathbf{K}^T \mathbf{K}$ , where  $\mathbf{D}_t$  is a (non-negative) diagonal matrix which depends on  $\mathbf{x}_t$  and  $\Phi$ . The huge size of  $\mathbf{A}_t$  forces an iterative implementation of (8); in [3], a *two-step* (also called *second-order* [1]) *stationary iterative method* (2SIM) was adopted.

It was shown in [3] that, for strongly ill-conditioned systems, IRS is much faster than IST, due to the use of the 2SIM. On the other hand, when noise is the main factor, and the observation operator is not severely ill-conditioned, IST outperforms IRS due to its closed-form denoising step in each iteration [16]. In the extreme case of a pure denoising problem ( $\mathbf{K} = \mathbf{I}$ ), IST (with  $\beta = 1$  and initialized at  $\mathbf{x}_1 = \mathbf{0}$ ) converges in one step, while IRS does not. The 2IST method proposed in this paper keeps the good denoising performance of IST, but is able to handle severely ill-posed problems as efficiently as IRS.

## 4. TWO-STEP IST (2IST)

Inspired by the good performance of the 2SIM used in IRS, we propose a two-step version of IST (2IST), defined as

$$\mathbf{x}_1 = \Gamma_\lambda(\mathbf{x}_0) \quad (9)$$

$$\mathbf{x}_{t+1} = (1 - \alpha) \mathbf{x}_{t-1} + (\alpha - \beta) \mathbf{x}_t + \beta \Gamma_\lambda(\mathbf{x}_t), \quad (10)$$

for  $t \in \mathbb{N}$ , where  $\Gamma_\lambda : \mathbb{R}^m \rightarrow \mathbb{R}^m$ , is defined as

$$\Gamma_\lambda(\mathbf{x}) = \Psi_\lambda \left( \mathbf{x} + \mathbf{K}^T (\mathbf{y} - \mathbf{K} \mathbf{x}) \right). \quad (11)$$

The following theorem partially characterizes the convergence of the 2IST algorithm, when  $f$  has a unique minimizer.

**Theorem 4** Let  $f$  be given by (1), where  $\Phi$  is convex. Let  $\xi$  be a real number such that  $0 < \xi \leq \lambda_i(\mathbf{K}^T \mathbf{K}) < 1$ , where  $\lambda_i(\cdot)$  is the  $i$ -th eigenvalue of its matrix argument, and let  $\hat{\rho} = (1 - \sqrt{\xi})/(1 + \sqrt{\xi})$ . Let  $\hat{\mathbf{x}}$  be the unique (because  $\mathbf{K}$  is injective) minimizer of  $f$  and define the “error vector” as  $\mathbf{e}_t = \mathbf{x}_t - \hat{\mathbf{x}}$  and the “staked error vector” as

$$\mathbf{w}_t = \left[ (\mathbf{e}_{t+1})^T \sqrt{1 - \alpha} (\mathbf{e}_t)^T \right]^T. \quad (12)$$

- (i) There exist matrices  $\mathbf{Q}_t$  such that  $\mathbf{w}_{t+1} = \mathbf{Q}_t \mathbf{w}_t$ , for  $t \in \mathbb{N}$ ; moreover, if  $0 < \alpha < 2$  and  $0 < \beta < 2\alpha$ , then  $\rho(\mathbf{Q}_t) < 1$ , where  $\rho(\cdot)$  denotes the spectral radius, i.e., the largest absolute eigenvalue;
- (ii) setting  $\alpha = \hat{\rho}^2 + 1$  and  $\beta = \frac{2\alpha}{1+\xi}$  guarantees that  $\rho(\mathbf{Q}_t) = \hat{\rho}$ ;
- (iii) if  $0 < \alpha \leq 1$  and  $0 < \beta < 2\alpha$ , then  $\lim_{t \rightarrow \infty} \mathbf{w}_t = \mathbf{0}$ ;
- (iv) with  $\alpha = 1$ , 2IST becomes IST; taking  $\beta = \frac{2}{1+\xi}$  guarantees that  $\rho(\mathbf{Q}_t) \leq \frac{1-\xi}{1+\xi} \equiv \bar{\rho}$ .

Theorem 4 extends the results about the convergence of the linear 2SIM (see [1]) to the non-linear/non-differentiable case. While the proof in [1] uses linear algebra tools, the possible non-linear/non-differentiable nature of  $\Psi_\lambda$  demands non-smooth analysis techniques [9, 23]. The proof of Theorem 4, as well as a related result for the case where the minimizer is not unique, can be found in [4].

If  $\mathbf{Q}_t = \mathbf{Q}$ , the condition  $\rho(\mathbf{Q}) < 1$  would be sufficient for convergence to zero of  $\mathbf{w}_t$ . However, in 2IST,  $\mathbf{Q}_t$  is in general not constant, thus  $\rho(\mathbf{Q}_t) < 1$ ,  $\forall t$ , is not a sufficient condition for convergence. Convergence of a non-stationary linear iteration  $\mathbf{w}_{t+1} = \mathbf{Q}_t \mathbf{w}_t$ , where  $\mathbf{Q}_t \in \mathcal{Q}$ , depends on the *joint spectral radius* (JSR) of the matrix set  $\mathcal{Q}$  [24]. Computing (or bounding) the JSR of (even very small) matrix sets is a hard problem, currently under active research (see [24] and references therein). The convergence stated in (iii) results from the fact that for  $\alpha \leq 1$ ,  $\mathbf{Q}_t$  is symmetric, thus  $\rho(\mathbf{Q}_t) = \|\mathbf{Q}_t\|_2 < 1$ , which guarantees convergence.

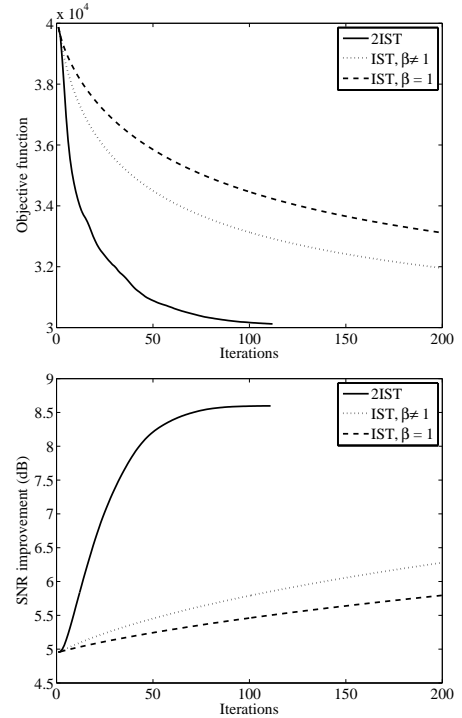
Although, when  $\alpha > 1$ , Theorem 4 does not guarantee convergence, we have observed, in a large number of image deconvolution experiments, that the algorithm always converges with the setting given in (ii). Although  $\hat{\rho} < 1$  and  $\bar{\rho} < 1$  do not guarantee convergence, we have experimentally verified that these values are good indicators of the relative speed of 2IST and IST. Treating the algorithms as linear stationary, we could make the following observation. The quantities  $-1/\log_{10} \hat{\rho}$  and  $-1/\log_{10} \bar{\rho}$ , are approximately the numbers of iterations needed to reduce the error norm by a factor of 10. For example, with  $\xi = 10^{-4}$  (not uncommon in image restoration problems),  $-1/\log_{10} \hat{\rho} \simeq 10^2$ , while  $-1/\log_{10} \bar{\rho} \simeq 10^4$ ; i.e., in this case 2IST is roughly two orders of magnitude faster than IST.

## 5. EXPERIMENTAL RESULTS

This section reports experiments comparing the convergence speeds of 2IST and IST. We stress that the goal of these experiments is not to assess the performance (e.g., in terms of SNR improvement) of image deconvolution criteria of the form (1). Such an assessment has been carried out (in comparison with other state-of-the-art techniques) in several previous publications [3, 17, 18].

Our first experiment uses a benchmark problem which was studied in [3, 17, 18] (and in many other deconvolution papers): the observed image is obtained by convolving the well-known “Cameraman” image with a  $9 \times 9$  uniform blur and then adding noise with variance 40dB below that of the blurred image. We deconvolve the observed image by minimizing (1), with  $\Phi$  being the isotropic

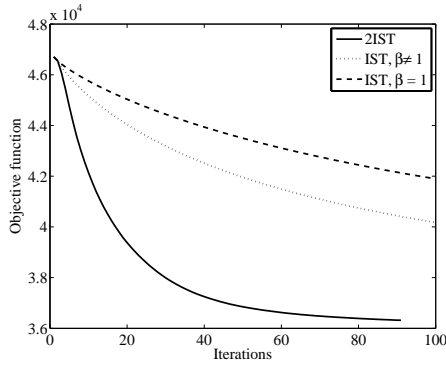
discrete version of the TV regularizer [5], and  $\lambda$  hand-tuned for optimal performance. In this case,  $\Psi_\lambda$  is a TV denoising function, implemented by running a few steps (e.g., 5) of the algorithm proposed in [5]. In all the experiments, the algorithm is initialized with a Wiener filter estimate, as in [3]. Figure 1 shows the evolution of the objective function  $f(\mathbf{x}_t)$  and of the SNR improvement, along the iterations of 2IST, IST with  $\beta = 1$  (the version introduced in [12], [21]), and IST with  $\beta \neq 1$ . The 2IST algorithm was stopped when  $|f(\mathbf{x}_t) - f(\mathbf{x}_{t-1})|/f(\mathbf{x}_{t-1}) < 10^{-4}$  and the other two algorithms were run until they reached the same value of  $f(\mathbf{x}_t)$ , which happened (not shown) after  $\sim 2100$  iterations, for  $\beta \neq 1$ , and  $\sim 4000$  iterations, for  $\beta = 1$ .



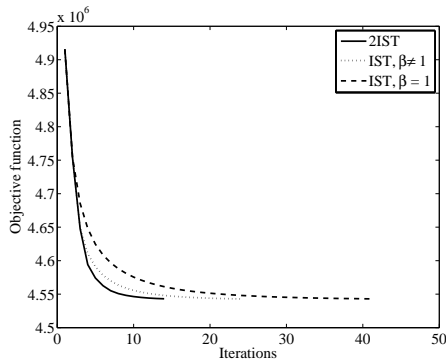
**Fig. 1.** TV deblurring example: evolution of the objective function and the SNR improvement, for 2IST, and IST with  $\beta = 1$  and  $\beta \neq 1$ .

In the second experiment, we apply wavelet-based deconvolution to the same observed image; in this case,  $\mathbf{K} = \mathbf{H}\mathbf{W}$ , where  $\mathbf{H}$  is the matrix representing the blur,  $\mathbf{W}$  is the inverse DWT transform, and  $\mathbf{x}$  is the set of wavelet coefficients of the unknown image  $\mathbf{W}\mathbf{x}$  [3, 17, 18]. We use Haar wavelets and take  $\Phi$  as the  $\ell^1$  norm, thus  $\Psi_\lambda$  is the soft-threshold function. The results of this experiment are reported in Figure 2. The qualitative behavior of the algorithms is very similar to the first experiment.

The third and last experiment uses the following setup (also studied in [3, 17, 18]): the observed image is obtained by convolving the well-known “Lena” image with a separable blur with kernel  $[1, 4, 6, 4, 1]^T [1, 4, 6, 4, 1]/256$  and then adding noise with standard deviation equal to 7. The blur is much less severe than the uniform  $9 \times 9$  considered in the previous experiments, thus 2IST is not expected to show such a clear superiority over IST as it did above. The evolution of  $f(\mathbf{x}_t)$  along the iterations of algorithms, with a TV regularizer, is shown in Figure 3; the plot shows that 2IST is still faster than both versions of IST, but by a smaller margin than above.



**Fig. 2.** Wavelet-based deblurring example: evolution of objective function for 2IST, and IST with  $\beta = 1$  and  $\beta \neq 1$ .



**Fig. 3.** TV deblurring example, under mild blur: evolution of objective function for 2IST, and IST with  $\beta = 1$  and  $\beta \neq 1$ .

## 6. CONCLUDING REMARKS

We have introduced a two-step version of the recent iterative shrinkage/thresholding (IST) algorithm, termed 2IST, for a class of convex objective functions, appearing namely in total-variation and wavelet-based image restoration. We have presented theoretical results concerning the convergence behavior of the 2IST algorithm. In a typical benchmark image deblurring problem (with strong blur), 2IST converges between 1 and 2 orders of magnitude faster than the original IST. Current work includes extending Theorem 4, namely showing convergence for  $\alpha > 1$ .

## 7. REFERENCES

- [1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, 1996.
- [2] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, IOP Publishing, Bristol, UK, 1998.
- [3] J. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Trans. Image Proc.*, vol. 15, pp. 937–951, 2006.
- [4] J. Bioucas-Dias and M. Figueiredo, "Two-step iterative thresholding algorithms for image restoration," submitted, 2007.
- [5] A. Chambolle, "An algorithm for total variation minimization and applications," *Jour. Math. Imaging and Vision*, vol. 20, pp. 89–97, 2004.
- [6] A. Chambolle, "Total-variation minimization and a class of binary MRF models", *Intern. Workshop on Energy Minim. Methods in Comp. Vis. and Patt. Rec.*, pp. 136–152, Springer, 2005.
- [7] A. Chambolle, R. De Vore, N. Lee, and B. Lucier, "Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Proc.*, vol. 7, pp. 319–335, 1998.
- [8] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *Mathematical Models of Computer Vision*, Springer Verlag, 2005.
- [9] F. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, 1983.
- [10] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling & Simulation*, vol. 4, pp. 1168–1200, 2005.
- [11] J. Darbon and M. Sigelle, "A fast and exact algorithm for total variation minimization", *Iberian Conference on Patt. Rec. and Image Anal.*, LNCS vol. 3522, pp. 351–359, Springer, 2005.
- [12] I. Daubechies, M. Defriese, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint", *Comm. Pure and Applied Math.*, vol. LVII, pp. 1413–1457, 2004.
- [13] D. Donoho, "De-noising by soft thresholding", *IEEE Trans. Information Theory*, vol. 41, pp. 613–627, 1995.
- [14] M. Elad, "Why simple shrinkage is still relevant for redundant representations?", submitted, 2006.
- [15] M. Elad, B. Matalon, and M. Zibulevsky, "Image denoising with shrinkage and redundant representations", *Proc. IEEE CVPR'2006*, New York, 2006.
- [16] M. Figueiredo, J. Bioucas-Dias, and R. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration" submitted, 2007.
- [17] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Proc.*, vol. 12, pp. 906–916, 2003.
- [18] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," *IEEE ICIP'05*, Genoa, Italy, 2005.
- [19] D. Lorenz, *Wavelet Shrinkage in Signal and Image Processing: An Investigation of Relations and Equivalences*, Phd Thesis, Universität Bremen, 2004. Available at [www.math.uni-bremen.de/~dlorenz](http://www.math.uni-bremen.de/~dlorenz)
- [20] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian and complexity priors," *IEEE Trans. Info. Theory*, vol. 45, pp. 909–919, 1999.
- [21] R. Nowak and M. Figueiredo, "Fast wavelet-based image deconvolution using the EM algorithm", *Proc. Asilomar Conf. Signals, Systems, and Computers*, vol. 1, pp. 371–375, 2001.
- [22] S. Osher, L. Rudin, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
- [23] R. T. Rockafellar and R. Wets, *Variational Analysis*, Springer Verlag, Berlin, 1998.
- [24] J. Theys, *Joint Spectral Radius: Theory and Approximations*, Phd Thesis, Univ. Catholique Louvain, 2005. Available at [www.inma.ucl.ac.be/~blondel/05thesetheys.pdf](http://www.inma.ucl.ac.be/~blondel/05thesetheys.pdf)
- [25] C. Zalinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, Singapore, 2002.