SIAM J. IMAGING SCIENCES
Vol. 3, No. 3, pp. 253–276


# Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction*

Xiaoqun Zhang†, Martin Burger‡, Xavier Bresson†, and Stanley Osher†

**Abstract.** Bregman methods introduced in [S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *Multiscale Model. Simul.*, 4 (2005), pp. 460–489] to image processing are demonstrated to be an efficient optimization method for solving sparse reconstruction with convex functionals, such as the $\ell^1$ norm and total variation [W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *SIAM J. Imaging Sci.*, 1 (2008), pp. 143–168; T. Goldstein and S. Osher, *SIAM J. Imaging Sci.*, 2 (2009), pp. 323–343]. In particular, the efficiency of this method relies on the performance of inner solvers for the resulting subproblems. In this paper, we propose a general algorithm framework for inverse problem regularization with a single forward-backward operator splitting step [P. L. Combettes and V. R. Wajs, *Multiscale Model. Simul.*, 4 (2005), pp. 1168–1200], which is used to solve the subproblems of the Bregman iteration. We prove that the proposed algorithm, namely, Bregmanized operator splitting (BOS), converges without fully solving the subproblems. Furthermore, we apply the BOS algorithm and a preconditioned one for solving inverse problems with nonlocal functionals. Our numerical results on deconvolution and compressive sensing illustrate the performance of nonlocal total variation regularization under the proposed algorithm framework, compared to other regularization techniques such as the standard total variation method and the wavelet-based regularization method.

**Key words.** nonlocal regularization, Bregman iteration, primal dual method

**AMS subject classifications.** 47A52, 49N45, 65K10

**DOI.** 10.1137/090746379

## 1. Introduction.

We consider a general inverse problem formulation for image restoration. The objective is to find the unknown true image $u \in \mathbb{R}^n$ from an observed image (or measurements) $f \in \mathbb{R}^m$ defined by the forward model

$$f = Au + \epsilon,$$

where $\epsilon$ is a white Gaussian noise with variance $\sigma^2$ and $A$ is an $m \times n$ linear operator, typically a convolution operator in the deconvolution problem or a subsampling measurement operator in the compressive sensing problem.

*Received by the editors January 13, 2009; accepted for publication (in revised form) April 6, 2010; published electronically July 1, 2010.

http://www.siam.org/journals/siims/3-3/74637.html

†Department of Mathematics, UCLA, Box 951555, Los Angeles, CA 90095-1555 (xqzhang@math.ucla.edu, xbresson@math.ucla.edu, sjo@math.ucla.edu). The first author's work was supported by an ARO MURI subcontract from the University of South Carolina and NSF DMS-03-12222. The third author's work was supported by ONR N00014-03-1-0071 and an ONR MURI subcontract from Stanford University. The last author's work was supported by ONR N00014-07-1-0810, ONR N00014-08-1-1119, and NSF DMS-07-14087.

‡Institute for Computational and Applied Mathematics, Westfälische Wilhelms-Universität, Einsteinstr. 62, D48163 Münster, Germany (martin.burger@wwu.de). This author's work was supported by the German Research Foundation DFG via the project "Regularisierung mit singulären Energien" and the BMBF via the project "INVERS: Deconvolution with sparsity constraints."

Since inverse problems are typically ill-posed, it is standard to use a regularization technique to make them well-posed. Regularization methods assume some prior information about the unknown function $u$ such as sparsity, smoothness, or small total variation (TV). A well-known example of a regularized inverse problem is the Tikhonov regularization model, which consists of solving the following optimization problem:

$$\min_{u\in\mathbb{R}^n}\left(\frac{\mu}{2}||u||^2+\frac{1}{2}||Au-f||^2\right),$$

where $\mu>0$ is a scale parameter which balances the trade-off between the regularity of the restored image $u$ and the fidelity to the observed image $f$, and $||\cdot||$ denotes the $\ell^2$ norm. The notation $||\cdot||$ for the $\ell^2$ norm will be used throughout the paper.

Other examples of regularized inverse problems are image denoising problems, where $A$ is considered as the identity or an embedding operator. A successful edge preserving image denoising model is the Rudin–Osher–Fatemi (ROF) model proposed in [44]. This model uses the TV regularization functional since images are assumed to have bounded variation, which is the case for piecewise constant images. A discrete form of the ROF model can be defined as follows [6]:

$$\min_{u\in\mathbb{R}^n}\left(\mu|\nabla u|_1+\frac{1}{2}||Au-f||^2\right),$$

where $\nabla u$ is the weak gradient of $u$ and $|\nabla u|_1$ is the TV of $u$, in other words, the $\ell^1$ norm of the length vector of $\|\nabla u\|$ at each point.

Regularization based on sparsity properties with respect to a specified basis, such as frames or dictionaries, has become popular recently. Suppose that an image is formulated as a column vector (signal) of size $n$, and $D\in\mathbb{R}^{n\times m}$ is a given frame or a dictionary matrix; there are two different formulations for solving the problem: analysis-based or synthesis-based [22]. The analysis-based model is formulated as

$$u^*=\arg\min_u\left(\mu|D^*u|_1+\frac{1}{2}||Au-f||^2\right), \tag{1.1}$$

where $D^*$ denotes the conjugate transpose matrix. On the other hand, the synthesis-based method consists of solving the problem

$$\alpha^*=\arg\min_\alpha\left(\mu|\alpha|_1+\frac{1}{2}\|A(D\alpha)-f\|^2\right), \tag{1.2}$$

and the solution is $u^*=D\alpha^*$. If $D$ is an orthogonal basis, then the two models are equivalent. The analysis-based model (1.1) is largely used for inverse problems, such as the wavelet-vaguelette decomposition model defined in [18]. Usually, a scale-dependent shrinkage is employed to estimate the image wavelet coefficients. The synthesis-based model has been applied to wavelet-based deconvolution (for example, in [16] and [24]), and recently it has received a lot of research interest in the area of compressive sensing problems [10]. Several efficient algorithms, such as iterative soft thresholding (IST) [16], $l1\_ls$ [36], gradient projection for sparse reconstruction (GPSR) [25], fixed-point continuation (FPC) [31], and linearized Bregman [39, 7, 8, 47], are proposed for solving this formulation. Compressive sensing, also known as compressed sampling, originates from approximation theory and has recently received a lot of interest in different research areas. In a probabilistic setting, compressive sensing argues that if signals can be expressed with a small support in a proper basis, then they can be reconstructed from a number of measurements significantly below the Nyquist–Shannon limit by using convex optimization. See [11, 43] for an introduction. The crucial observation is that objects having a sparse representation in a certain basis must be spread out in the sensing domain, such as Fourier or Gaussian measurements. Therefore, many efforts are devoted to finding the best basis for natural signals/images to fit the theory of compressive sensing, such as curvelets [9], contourlets [17], and trained dictionaries [1]. The advantage of wavelet methods is that they can efficiently represent classes of signals containing singularities. However, results via shrinkage in the wavelet domain are usually unsatisfactory due to amplified noise and produce undesirable artifacts. Furthermore, it is difficult to choose a proper basis for different images.

In this paper, we make two main contributions. First, we propose a general algorithm framework for an equality constrained convex optimization formulation

$$\min_u J(u) \ \text{ subject to (s.t.) } \ Au = f, \tag{1.3}$$

where $J$ is a general convex functional. This problem (1.3) is shown to cover a wide range of signal and image processing tasks for various choices of the convex functionals $J$ and $A$, including $\ell^1$ basis pursuit [48] and image restoration by TV [38]. The algorithms proposed in this paper are based on the Bregman iteration introduced in [38] and the proximal forward-backward operator splitting method [26, 14, 31]. Note that if there is noise present in the measurements, we can use a discrepancy stopping criterion as in the original Bregman iteration [38], that is, $\|Au^k - f\| \leq \sigma$ with the same algorithm. The principle of our algorithms is to maximally decouple the minimization functionals. More specifically, the overall algorithms consist of two forward (explicit) gradient steps (one is the Bregman iteration step) and an implicit step equivalent to the ROF model [44], which we can often solve efficiently. The proposed algorithms can be also interpreted as inexact Uzawa methods used for linear saddle point problems [50, 28]. However, the convergence of our algorithm does not seem to be directly implied by classical convergence analysis. Therefore we will present a proof of the convergence in this paper. The second contribution of this paper is investigating the application of nonlocal total variation (NLTV) for compressive sensing and deconvolution. Our experiments show that the proposed NL regularization model can recover almost all the details of a textured image without explicitly choosing a basis compared to the above dictionary-based sparse representation algorithms. Our investigation demonstrates that the NLTV regularization itself sparsifies textured images and that the Bregman iteration is an effective method for sparse recovery.

The paper is organized as follows. In section 2, we briefly review some related optimization techniques: Bregman iteration, operator splitting, and linearized Bregman. We then present the general algorithm that we call Bregmanized operator splitting (BOS) with convergence analysis for solving the problem (1.3). In section 4 we present the NL regularization and the application of our proposed algorithm and a preconditioned one. An updating strategy for the weight function in the NLTV term is discussed. Finally, in section 5 we present the numerical results for deconvolution and compressive sensing reconstruction.

## 2. Related work.

**2.1. Bregman iteration.** In this section, we introduce the Bregman iteration method for solving the problem (1.3). It is well known that this problem is difficult to solve numerically when $J$ is nondifferentiable. An efficient method for solving this constrained minimization problem is to use the Bregman iteration, initially introduced to imaging in [38] to improve the ROF denoising models [44].

The Bregman iteration scheme is based on the Bregman distance. The Bregman distance of a convex functional $J(\cdot)$ between points $u$ and $v$ is defined as

$$D_J^p(u,v) = J(u) - J(v) - \langle p,\ u - v\rangle, \tag{2.1}$$

where $p \in \partial J$ is a subgradient of $J$ at the point $v$. Bregman distance is not a distance in the usual sense because it is generally not symmetric. However, it measures the closeness of two points since $D_J^p(u,v) \geq 0$ for any $u$ and $v$, and $D_J^p(u,v) \geq D_J^p(w,v)$ for all points $w$ on the line segment connecting $u$ and $v$. Using the Bregman distance (2.1), the original constrained minimization problem (1.3) can be solved by the following iterative scheme:

$$\begin{cases} u^{k+1} &= \min_u \left(\mu D_J^{p^k}(u,u^k) + \frac{1}{2}||Au - f||^2\right), \\ p^{k+1} &= p^k - \frac{1}{\mu}A^T(Au^{k+1} - f), \end{cases}$$

where $\mu > 0$. By a change of variable, we obtain a two-step Bregman iterative scheme [38]:

$$\begin{cases} u^{k+1} &= \min_u \left(\mu J(u) + \frac{1}{2}||Au - f^k||^2\right), \\ f^{k+1} &= f^k + f - Au^{k+1}. \end{cases} \tag{2.2}$$

It is shown in [38] that the sequence $u^k$ weakly converges to a solution of (1.3), and the residual $||Au^k - f||$ of the sequence generated by (2.2) converges to zero monotonically. Recently, Bregman iteration has been used successfully in sparse reconstruction problems due to its speed, simplicity, efficiency, and stability; see, for example, [32, 38, 30, 48].

**2.2. Forward-backward operator splitting.** Operator splitting methods have been extensively studied in the optimization community, e.g., [26, 46, 14, 20, 28]. They aim to minimize the sum of two convex functionals:

$$\min_u \Big(\mu J(u) + H(u)\Big), \tag{2.3}$$

where $\mu > 0$. In [14], Combettes and Wajs proposed using the forward-backward proximal point iteration technique for general signal recovery tasks. The proximal operator of a convex functional $J$ of a function $v$, which was originally introduced by Moreau in [37], is defined as

$$\mathrm{Prox}_J(v) := \arg\min_u \left(J(u) + \frac{1}{2}||u - v||^2\right). \tag{2.4}$$

By classical arguments of convex analysis, the solution of (2.3) satisfies the condition

$$0 \in \mu\partial J(u) + \partial H(u).$$

For any positive number $\delta$, we have

$$0 \in (u + \delta\mu\partial J(u)) - (u - \delta\partial H(u)).$$

This leads to a forward-backward splitting algorithm:

$$u^{k+1} = \mathrm{Prox}_{\delta\mu J}(u^k - \delta\partial H(u^k)). \tag{2.5}$$

Also in [14], a general convergence was established for the generic problem. More specifically, in the case of $H(u) = \frac{1}{2}||Au - f||^2$, the algorithm converges when $0 < \delta < \frac{2}{\|A^T A\|}$. The algorithm (2.5) for the solution of the minimization problem (2.3) can be reformulated as the following two-step algorithm:

$$\begin{cases} v^{k+1} &= u^k - \delta A^T(Au^k - f), \\ u^{k+1} &= \arg\min_u \left(\mu J(u) + \frac{1}{2\delta}||u - v^{k+1}||^2\right). \end{cases} \tag{2.6}$$

The main advantage of this algorithm is that the two functionals are decoupled. Furthermore, the proximal minimization (2.4) is strictly convex, and thus there exists a unique minimizer. In practice, the proximal operator solution (2.4) has well-known solutions for some models. For example, when the regularization functional $J$ is the $\ell^1$ norm of $u$, i.e., $J(u) = |u|_1$, then the solution is obtained by a soft shrinkage operator [14, 31, 7] as follows:

$$u = \mathrm{shrink}(v, \mu\delta) = \mathrm{sign}(v)\max\{|v| - \mu\delta, 0\}. \tag{2.7}$$

When the regularization functional $J$ is the TV norm of $u$, i.e., $J(u) = |\nabla u|_1$, then the solution can be determined, e.g., by Chambolle's projection method [12], the split Bregman method [30], or by graph cuts in the anisotropic case [32, 15, 29].

**2.3. Linearized Bregman.** The idea of the linearized Bregman iteration is to combine Bregman iteration and operator splitting to solve the constrained problem (1.3) for sparse reconstruction. The algorithm in a simple formulation is as follows:

$$\begin{cases} v^{k+1} &= v^k - \delta A^T(Au^k - f), \\ u^{k+1} &= \arg\min \left(\mu J(u) + \frac{1}{2\delta}||u - v^{k+1}||^2\right). \end{cases} \tag{2.8}$$

The difference between linearized Bregman and the operator splitting method (2.6) is in the way we update $v^{k+1}$. These methods solve different problems. In fact, Cai, Osher, and Shen proved the following propositions in [7].

Proposition 2.1. *If the sequence $u^k$ converges and $p^k$ is bounded, then the limit of $u^k$ is the unique solution of*

$$\min \left(\mu J(u) + \frac{1}{2\delta}||u||^2\right) \quad s.t. \quad Au = f. \tag{2.9}$$

In the case of $\ell^1$ sparse approximation, algorithm (2.8) can be written as follows:

$$\begin{cases} v^{k+1} &= v^k - \delta A^T(Au^k - f), \\ u^{k+1} &= \mathrm{shrink}(v^{k+1}, \mu\delta). \end{cases}$$

As $\mu \to \infty$, the solution of (2.9) tends to the solution of (1.3); even better, it was proved in [47] that, for $\mu$ large enough, the limit solution solves the original problem:

$$\min |u|_1 \quad \text{s.t.} \quad Au = f.$$

## 3. General algorithm framework.

**3.1. Bregmanized operator splitting (BOS).** In this section, we present the proposed algorithm. Our goal is to solve the general equality constrained minimization problem (1.3) by the Bregman iteration and operator splitting introduced in sections 2.1 and 2.2. First, the equality constraint in (1.3) is enforced with the Bregman iteration process:

$$\left\{ \begin{array}{rcl} u^{k+1} & = & \min_u \left( \mu J(u) + \frac{1}{2}||Au - f^k||^2 \right), \\ f^{k+1} & = & f^k + f - Au^{k+1}. \end{array} \right. \tag{3.1}$$

The first subproblem can sometimes be difficult and slow to solve directly, since it involves the inverse of the operator $A$ and the convex functional $J$. The forward-backward operator splitting technique is used to solve the unconstrained subproblem in (3.1) as follows: for $i \geq 0, u^{k+1,0} = u^k$,

$$\left\{ \begin{array}{rcl} v^{k+1,i+1} & = & u^{k,i} - \delta A^T(Au^{k+1,i} - f^k), \\ u^{k+1,i+1} & = & \min_u \left( \mu J(u) + \frac{1}{2\delta}||u - v^{k+1,i+1}||^2 \right) \end{array} \right. \tag{3.2}$$

for a positive number $0 < \delta < \frac{2}{||A^T A||}$. Ideally we need to run infinite inner iterations to obtain a convergent solution $u^{k+1}$ for the original subproblem. Nevertheless, the convergence and error bound with arbitrary finite steps is unclear. Therefore, we propose using only one inner iteration, which leads to Algorithm I.

Algorithm I (Bregmanized operator splitting).

$$\left\{ \begin{array}{rcl} v^{k+1} & = & u^k - \delta A^T(Au^k - f^k), \\ u^{k+1} & = & \arg\min_u \left( \mu J(u) + \frac{1}{2\delta}||u - v^{k+1}||^2 \right), \\ f^{k+1} & = & f^k + f - Au^{k+1}, \end{array} \right. \tag{3.3}$$

*which is equivalent to*

$$\left\{ \begin{array}{rcl} u^{k+1} & = & \arg\min_u \left( \mu J(u) + \frac{1}{2\delta}||u - ((1 - \delta A^T A)u^k + \delta A^T f^k)||^2 \right), \\ f^{k+1} & = & f^k + f - Au^{k+1}. \end{array} \right. \tag{3.4}$$

**3.2. Connections with existing methods.** The above algorithm can be interpreted as an inexact Uzawa method [50, 28] applied to the augmented Lagrangian [42] of the original problem as follows:

$$L(u, p) = \mu J(u) + \frac{1}{2}\|Au - f\|^2 - \langle Au - f, p - f \rangle, \tag{3.5}$$

where $p$ is a Lagrange multiplier of the original problem (1.3). Note that we use a change of variable for the Lagrange multiplier $p$ to get the same formulation as the BOS algorithm. If we apply an inexact Uzawa method [50] and Moreau–Yosida proximal point iteration [34] on this formulation, we get the following algorithm:

(3.6)
$$\begin{cases} \text{step 1:} & u^{k+1} & = & \min_u \left(\mu J(u) + \frac{1}{2}||Au - f||^2 - \langle Au - f, p^k - f\rangle + ||u - u^k||_D^2\right), \\ \text{step 2:} & p^{k+1} & = & p^k - (Au^{k+1} - f), \end{cases}$$

where $D$ is a positive definite matrix. The sequence $(u^k, p^k)$ generated by (3.6) gives

$$\begin{cases} \mu s^{k+1} + (D + A^T A)u^{k+1} & = & Du^k + A^T p^k, \\ p^{k+1} & = & p^k - (Au^{k+1} - f), \end{cases}$$

where $s^k \in \partial J(u^k)$. When $D = \frac{1}{\delta} - A^T A$ and when we change the variable $p^k$ to $f^k$, we get the BOS algorithm defined in (3.3).

Most analysis for inexact Uzawa methods is available for linear saddle point problems with strong convexity assumptions [50, 28]. Other available analysis based on the augmented Lagrangian methods [28, 42] is also different from ours due to the maximally decoupled structures. Note that the presented algorithm is different from the split Bregman algorithm [30] in the manner of splitting, for the latter can be recast as a Douglas–Rachford algorithm [19, 20, 45]. On the other hand, this algorithm can be generalized to a large range of convex minimization problems. A more detailed study of the BOS algorithm framework and theoretical connections to proximal point algorithms and augmented Lagrangian methods will be presented in a forthcoming paper.

**3.3. Convergence analysis.** In this section, we prove the convergence of the proposed BOS algorithm. In the following, we assume that the convex function $J$ of (1.3) is closed, proper, semicontinuous, and convex.

Theorem 3.1. *Let the sequence $(u^k, f^k)$ be generated by Algorithm I given in (3.4). If $0 < \delta < \frac{1}{\|A^T A\|}$, then every accumulation point of $u^k$ is a solution of (1.3).*

*Proof.* We first consider a Lagrangian formulation of the original constrained problem (1.3):

$$L(u, p) = \mu J(u) - \langle Au - f, p - f\rangle \quad \text{and} \quad Au = f.$$

Note that we use a change of variable for the Lagrangian multiplier, $p - f$ instead of $p$ as above.

We let $\overline{u}$ be an optimal solution of (1.3) and $\overline{p}$ be a Lagrangian multiplier, respectively, and we denote

$$\overline{s} = -\frac{1}{\mu} A^T (f - \overline{p}). \tag{3.7}$$

Then we can see that $\overline{s}$ is a subgradient of $J$ at $\overline{u}$ by the Lagrangian function. Therefore, the overall optimality conditions are as follows:

$$\begin{cases} \mu\overline{s} + A^T(f - \overline{p}) & = & 0, \\ A\overline{u} - f & = & 0. \end{cases} \tag{3.8}$$

We let $(u^k, f^k)$ be the sequence generated by (3.4) and $s^{k+1} \in \partial J(u^{k+1})$, and we have

$$\left\{\begin{array}{rcl} \mu s^{k+1} + \frac{1}{\delta} u^{k+1} & = & (\frac{1}{\delta} - A^T A) u^k + A^T f^k, \\ f^{k+1} & = & f^k + f - A u^{k+1}. \end{array}\right. \tag{3.9}$$

Let $L = (\frac{1}{\delta} - A^T A)$; then $L$ is positive definite since $0 < \delta < \frac{1}{\|A^T A\|}$. By rewriting the above sequence, we get

$$\left\{\begin{array}{rcl} \mu s^{k+1} + L u^{k+1} - A^T f^{k+1} & = & L u^k - A^T f, \\ f^{k+1} & = & f^k + f - A u^{k+1}. \end{array}\right. \tag{3.10}$$

On the other hand, we can rewrite the sequences in terms of error as follows:

$$\begin{aligned} \Delta s^{k+1} &= s^{k+1} - \overline{s}, \\ \Delta f^{k+1} &= f^{k+1} - \overline{p}, \\ \Delta u^{k+1} &= u^{k+1} - \overline{u}. \end{aligned}$$

Therefore (3.10) is rearranged in terms of the error differences as

$$\left\{\begin{array}{rcl} \mu(\Delta s^{k+1}) + L(\Delta u^{k+1}) - A^T(\Delta f^{k+1}) & = & L(\Delta u^k), \\ \Delta f^{k+1} + A \Delta u^{k+1} & = & \Delta f^k. \end{array}\right.$$

Denoting $||v||_L^2 := \langle Lv, v \rangle$, we obtain

$$\begin{aligned} &\|\Delta u^{k+1}\|_L^2 + \|\Delta f^{k+1}\|^2 + \|u^{k+1} - u^k\|_L^2 + \|f^{k+1} - f^k\|^2 - \|\Delta u^k\|_L^2 - \|\Delta f^k\|^2 \\ &= 2\langle L(u^{k+1} - u^k), \Delta u^{k+1} \rangle + 2\langle f^{k+1} - f^k, \Delta f^{k+1} \rangle \\ &= 2\langle A^T \Delta f^{k+1}, \Delta u^{k+1} \rangle - 2\mu \langle \Delta s^{k+1}, \Delta u^{k+1} \rangle + 2\langle f - A u^{k+1}, \Delta f^{k+1} \rangle \\ &= -2\mu \langle \Delta s^{k+1}, \Delta u^{k+1} \rangle. \end{aligned} \tag{3.11}$$

Recall that since $s^{k+1}$ is a subgradient of the convex functional $J(u)$ at $u^{k+1}$, we have

$$\langle \Delta s^{k+1}, \Delta u^{k+1} \rangle = \langle s^{k+1} - \overline{s}, u^{k+1} - \overline{u} \rangle = D_J^{\overline{s}}(u^{k+1}, \overline{u}) + D_J^{s^{k+1}}(\overline{u}, u^{k+1}) \geq 0 \quad \forall k. \tag{3.12}$$

This yields the inequality

$$\|\Delta u^{k+1}\|_L^2 + \|\Delta f^{k+1}\|^2 \leq \|\Delta u^0\|_L^2 + \|\Delta f^0\|^2.$$

Since $L$ is positive definite, the sequences $u^k$ and $f^k$ are bounded and there exists a convergent subsequence of $(u^k, f^k)$. Second, by summing the equality (3.11), we obtain

$$\sum_{k=0}^{\infty} \|u^{k+1} - u^k\|_L^2 + \sum_{k=0}^{\infty} \|f^{k+1} - f^k\|^2 + 2\mu \sum_{k=0}^{\infty} \langle \Delta s^{k+1}, \Delta u^{k+1} \rangle \leq \|\Delta u^0\|_L^2 + \|\Delta f^0\|^2 < \infty.$$

Thus

$$\|u^{k+1} - u^k\|_L^2 \to 0, \quad \|f^{k+1} - f^k\|^2 \to 0, \quad \langle \Delta s^{k+1}, \Delta u^{k+1} \rangle \to 0.$$

The first formula implies that $\|u^{k+1} - u^k\| \to 0$ since $L$ is positive definite. The second yields

$$\lim_{k\to\infty} \|Au^{k+1} - f\|^2 = \lim_{k\to\infty} \|f^{k+1} - f^k\| = 0.$$

Finally, the third formula together with (3.12) implies that the nonnegative Bregman distance satisfies

$$\lim_{k\to\infty} D_J^{\overline{s}}(u^{k+1}, \overline{u}) = \lim_{k\to\infty} \Big( J(u^{k+1}) - J(\overline{u}) - \langle \overline{s}, u^{k+1} - \overline{u} \rangle \Big) = 0.$$

Using (3.7) and $Au^{k+1} \to f = A\overline{u}$, we have

$$0 = \lim_{k\to\infty} \Big( \mu J(u^{k+1}) - \mu J(\overline{u}) + \langle f - \overline{p}, A(u^{k+1} - \overline{u}) \rangle \Big) = \lim_{k\to\infty} \mu J(u^{k+1}) - \mu J(\overline{u}). \tag{3.13}$$

Thus $J(u^{k+1}) \to J(\overline{u})$.

Hence, for any accumulation point $u^\infty$, we have $Au^\infty = f$ and $J(u^\infty) = J(\overline{u})$ by the semicontinuity of $J$. We conclude directly that $u^\infty$ is a solution of (1.3). ∎

**4. Nonlocal regularization.** In this section, we first present some notation of the NL regularization introduced in [27], and then discuss applications for solving inverse problems.

**4.1. Background.** In [21], Efros and Leung used similarities in natural images to synthesize textures and fill in holes in images. The basic idea of texture synthesis is to search for similar image patches in the image and determine the value of the hole using found patches. Texture synthesis also influences the image denoising task. Buades, Coll, and Morel introduced in [4] an efficient denoising model called nonlocal means (NL-means). The model consists of denoising a pixel by averaging the other pixels with structures (patches) similar to that of the current one. More precisely, given a reference image $f$, we define the NL-means solution $NLM_f$ of the function $u$ at point $x$ as

$$NLM_f(u)(x) := \frac{1}{C(x)} \int_\Omega w(f, h_0)(x, y) u(y) dy,$$

where

$$\begin{aligned} w(f, h_0)(x, y) &= \exp\left\{ -\frac{G_a * (||f(x + \cdot) - f(y + \cdot)||^2)(0)}{2h_0^2} \right\}, \\ C(x) &= \int_\Omega \exp\left\{ -\frac{G_a * (||f(x + \cdot) - f(y + \cdot)||^2)(0)}{2h_0^2} \right\} dy, \end{aligned} \tag{4.1}$$

where $G_a$ is the Gaussian kernel with standard deviation $a$, $C(x)$ is the normalizing factor, and $h_0$ is a filtering parameter. When the reference image $f$ is known, the NL-means filter is a linear operator. In the case where the reference image $f$ is chosen to be $u$, the operator is nonlinear and is the NL-means filter presented by Buades, Coll, and Morel in [4]. The definition of the weight function (4.1) shows that this function is significant only if the patch around $y$ has a structure similar to that of the corresponding patch around $x$. This filter is very efficient in reducing noise while preserving textures and contrast of natural images. It is generally preferred to choose a reference image as close as possible to the true image in order to include relevant information.

In a discrete formulation, if the images are represented by a column vector $u$ of $N$ elements, the operator $NLM_f(u)$ can be written as matrix multiplications such as

$$NLM_f(u) = D_f^{-1} W_f u,$$

where $W_f$ is the $N \times N$ weight matrix defined in (4.1), and $D_f(i,i) = C(i)$ is an $N \times N$ diagonal matrix.

The application of the NL-means filter for inverse problems such as image deblurring is not trivial since the observed image and the original image generally do not have the same distribution and structures. Based on the hypothesis that the deblurred image must maintain the same coherence as the blurry image, Buades, Coll, and Morel proposed in [5] an NL-means regularization energy for image deblurring defined as follows:

$$J_{NLM}(u) := ||u - NLM_f(u)||^2, \tag{4.2}$$

where $NLM_f := D_f^{-1} W_f$ is the NL-means filter defined above and $W_f$ is the weight computed from the blurry and noisy image $f$.

An alternative nonlocal model for texture restoration is introduced in [3]. The authors propose minimizing the functional:

$$J_{NLM}(u) := ||u - NLM_u(f)||^2. \tag{4.3}$$

This is a nonlinear model since the weight function depends on the unknown image $u$. The solution of (4.3) is approximated by an iterated scheme:

$$u^{k+1} = NLM_{u^k}(f).$$

This model updates the denoising weight function at each iteration step and keeps averaging on the original image. The convergence property of this iterative process has not been yet established.

In order to formulate the NL-means filter in a variational framework, Kindermann, Osher, and Jones in [33] started to investigate the use of regularization functionals with NL correlation terms for general inverse problems. Also, inspired from the graph Laplacian in [13], Gilboa and Osher defined variational framework–based NL operators in [27]. Note that Zhou and Schölkopf in [49] and Elmoataz, Lezoray, and Bougleux in [23] also used the graph Laplacian in the discrete setting for image denoising. Finally, the connection between the filtering methods and spectral bases of the NL graph Laplacian operator is discussed in [40] by Peyré.

In the following, we give the definitions of the NL functionals introduced in [27]. Let $\Omega \subset \mathbb{R}^2$, let $x \in \Omega$, and let $u(x)$ be a real function $\Omega \to R$. Assume $w : \Omega \times \Omega \to \mathbb{R}$ is a nonnegative symmetric weight function defined in (4.1) from a reference image; then the NL gradient $\nabla_w u(x)$ is defined as the vector of all partial differences $\nabla_w u(x, \cdot)$ at $x$ such that

$$\nabla_w u(x,y) := (u(y) - u(x))\sqrt{w(x,y)} \quad \forall y \in \Omega. \tag{4.4}$$

A graph divergence of a vector $\vec{p} : \Omega \times \Omega \to R$ can be defined by the standard adjoint relation with the gradient operator as follows:

$$\langle \nabla_w u, p \rangle := -\langle u, \mathrm{div}_w p \rangle \quad \forall u : \Omega \to \mathbb{R},\ \forall p : \Omega \times \Omega \to \mathbb{R}, \tag{4.5}$$

which leads to the definition of the graph divergence $\mathrm{div}_w$ of $p : \Omega \times \Omega \to \mathbb{R}$ such that

$$\mathrm{div}_w p(x) = \int_\Omega (p(x,y) - p(y,x))\sqrt{w(x,y)}dy. \tag{4.6}$$

The graph Laplacian is defined by

$$\Delta_w u(x) := \frac{1}{2}\mathrm{div}_w(\nabla_w u(x)) = \int_\Omega (u(y) - u(x))w(x,y)dy. \tag{4.7}$$

Note that a factor $\frac{1}{2}$ is used to get the related standard Laplacian definition.

These operators possess several properties. For example, the Laplacian operator is self-adjoint, i.e.,

$$\langle \Delta_w u, u\rangle = \langle u, \Delta_w u\rangle,$$

and negative semidefinite, i.e.,

$$\langle \Delta_w u, u\rangle = -\langle \nabla_w u, \nabla_w u\rangle \leq 0.$$

The nonlocal $H^1$ and TV norms are defined to be the $L^2$ and isotropic $L^1$ norms, respectively, of the weighted graph gradient $\nabla_w u(x)$:

$$J_{NL/H^1,w}(u) := \frac{1}{4}\int |\nabla_w u(x)|^2 dx, \tag{4.8}$$

$$J_{NL/TV,w}(u) := \int_\Omega |\nabla_w u(x)| dx. \tag{4.9}$$

The corresponding Euler–Lagrange equations of (4.8) and (4.9) are then written as

$$-\int_\Omega (u(y) - u(x))w(x,y)dy = 0 \tag{4.10}$$

and

$$-\int_\Omega (u(y) - u(x))w(x,y)\left[\frac{1}{|\nabla_w u(x)|} + \frac{1}{|\nabla_w u(y)|}\right] dy = 0. \tag{4.11}$$

Note that once the weight function $w$ is fixed, the Euler–Lagrange equation for the NLH1 is linear and can be solved by a gradient descent method. However, analogous to the classical TV, the functional (4.9) is not differentiable when $|\nabla_w u| = 0$. For this case, a dual method or a regularized version $\sqrt{|\nabla_w u|^2 + \epsilon}$ can be used to avoid a zero denominator. Finally, if the function $w(x,y)$ in (4.11) is chosen to be the NL weight function defined in (4.1), then the NL-means filter is generalized to a variational framework. Nevertheless, the minimization of the NLTV functional remains as a difficult optimization problem due to the computation complexity and the nondifferentiability.

### 4.2. Nonlocal regularization for inverse problems.

**4.2.1. Weight fixed.** The NL regularization for inverse problems is based on the following constrained formulation:

$$\min_u J_w(u) \quad \text{s.t. } Au = f, \tag{4.12}$$

with $J_w$ being an NL regularization term, such as the NLTV or the NLH1 with a given weight function $w$, and with $A$ being a convolution operator or a compressive sensing matrix. By applying Algorithm 2, we obtain the first algorithm proposed in this paper:

$$\begin{cases} v^{k+1} &= u^k - \delta A^T(Au^k - f^k), \\ u^{k+1} &= \arg\min_u \left(\mu J_w(u) + \frac{1}{2\delta}||u - v^{k+1}||^2\right), \\ f^{k+1} &= f^k + f - Au^{k+1}. \end{cases} \tag{4.13}$$

We can see that the key computation of this algorithm relies on the computation of products of vectors by $A$ and $A^T$ and on the ROF-like denoising step. In section 4.2.4, we will present a fast method based on split Bregman iteration for TV minimization.

**4.2.2. Weight updating.** In the previous discussion of NL regularization methods, the weight function $w$ was fixed. In the denoising case, most image similarity information can be discovered by the given noisy image. Unfortunately, a good estimation of the weight $w_0 \approx w(u, h_0)$ given in (4.1) is not always available, especially in the case of inverse problems, where given data lie in a different space from the true image. In the case of compressive sensing, due to a low sample rate, a weight function from an initial guess is not good enough and the standard TV compressive sensing is also not capable of restoring complex textures. This is why it is necessary to update the weight function $w(u^k, h_0)$ (4.1) during the reconstruction of signals. In [41], the authors have proposed updating the graph weight to solve inverse problems using the forward-backward operator splitting technique [14] to solve the relaxed Lagrangian formulation. Like Peyré, Bougleux, and Cohen in [41], we consider a more appropriate problem:

$$\min_u J_w(u) \quad \text{s.t.} \quad Au = f \quad \text{and} \quad w = w(u, h_0). \tag{4.14}$$

However, a direct numerical solution of this problem is difficult to compute. Instead, the simplified algorithm based on Algorithm I (BOS) with weight updating is proposed:

$$\begin{cases} \text{step 1:} & v^{k+1} &= u^k - \delta A^T(Au^k - f^k), \\ \text{step 2:} & w^{k+1} &= w(v^{k+1}, h_0), \\ \text{step 3:} & u^{k+1} &= \min_u \left(\mu J_{w^{k+1}}(u) + \frac{\delta}{2}||u - v^{k+1}||^2\right), \\ \text{step 4:} & f^{k+1} &= f^k + f - Au^{k+1}. \end{cases} \tag{4.15}$$

Note that during the preparation of the final version of the current paper, we discovered that a variational framework with NL weight updating is given in [2] in the context of image inpainting. Although the connection with the entropy energy is demonstrated, a theoretical analysis is still under investigation.

**4.2.3. Preconditioned Bregmanized operator splitting (PBOS).** As we have mentioned, an important question in nonlocal regularization methods for inverse problems is how to estimate a correct weight function $w$. In [35], the authors estimate the weight function with the solution of the Tikhonov regularization problem:

$$v = \arg\min_v \left( \frac{1}{2}||Av - f||^2 + \frac{\epsilon}{2}||v||^2 \right), \tag{4.16}$$

where $\epsilon$ is a small positive number. The solution amounts to

$$v = (A^T A + \epsilon)^{-1} A^T f.$$

The operator $(A^T A + \epsilon)^{-1} A^T$ is a preconditioned generalized inverse of $A$ when $A$ is not invertible or ill-conditioned. In fact, we have

$$\lim_{\epsilon \to 0} (A^T A + \epsilon)^{-1} A^T = \lim_{\epsilon \to 0} A^T (AA^T + \epsilon)^{-1} = A^+,$$

where $A^+$ is the Moore–Penrose pseudoinverse of $A$ even if $(AA^T)^{-1}$ and/or $(A^T A)^{-1}$ do not exist. If the columns of $A$ are linearly independent, then $A^T A$ is invertible. In this case, an explicit formula is $A^+ = (A^T A)^{-1} A^T$. It follows that $A^+$ is a left inverse of $A$: $A^+ A = I$. Similarly, if the rows of $A$ are linearly independent, then $AA^T$ is invertible. In this case, an explicit formula is $A^+ = A^T (AA^T)^{-1}$. Furthermore, if $A$ has orthonormal columns ($A^T A = I$) or orthonormal rows ($AA^T = I$), then $A^+ = A^T$.

In [35], we show that the weight estimated from the preconditioned image gives a better result than the one from the blurry image, because the main edge information is kept in the preconditioned image even when the noise is amplified. Since NL methods are robust to noise, it is more important to preserve as much edge information as possible. For this reason, we consider a modified operator splitting algorithm analogous to the operator splitting algorithm (2.6):

$$\begin{cases} v^{k+1} &= u^k - \delta A^+ (Au^k - f), \\ u^{k+1} &= \arg\min_u \left( \mu J_w(u) + \frac{1}{2\delta}||u - v^{k+1}||^2 \right), \end{cases} \tag{4.17}$$

where $A^+$ is the pseudoinverse of $A$ and $\delta > 0$. This similar idea is also considered in [8] for frame-based image deblurring. The operator $A^+ A$ is an orthogonal projector onto the range space of $A^+$; thus it is positive semidefinite. In the following, we replace $A^+$ by $A^T (AA^T + \epsilon)^{-1}$, and then algorithm (4.17) solves the minimization problem

$$\min_u \left( \mu J_w(u) + \frac{1}{2}||Bu - b||^2 \right), \tag{4.18}$$

where $B = PA$, $b = Pf$, and $P = (AA^T + \epsilon)^{-\frac{1}{2}}$. In particular, we have the following:

- If $A$ is full row rank ($A^+ = A^T (AA^T)^{-1}$), then we set $\epsilon = 0$ and

$$B = (AA^T)^{-\frac{1}{2}} A, \quad b = (AA^T)^{-\frac{1}{2}} f.$$

- If $A^T A = I$, i.e., $A^+ = A^T$, then
$$B = A, \quad b = f.$$
Then the modified algorithm (4.17) is consistent with the classical operator splitting (2.6).
- If $A$ is diagonalizable in an orthonormal basis, i.e., $A = P^{-1}DP$, where $P$ is orthonormal, then we can easily verify that the left and right pseudoinverse approximations are equal; i.e.,
$$(A^T A + \epsilon)^{-1} A^T = A^T (AA^T + \epsilon)^{-1}.$$

Now we can consider a preconditioned constrained problem

$$\min_u J_w(u) \quad \text{s.t. } Bu = b. \tag{4.19}$$

We apply the general Bregmanized operator splitting algorithm (Algorithm I) on problem (4.19) and let $b = Pf$; then we get the algorithm

$$\begin{cases} v^{k+1} &= u^k - \delta A^T P^T (PAu^k - b^k), \\ u^{k+1} &= \arg\min_u \left(\mu J_w(u) + \frac{1}{2\delta}||u - v^{k+1}||^2\right), \\ b^{k+1} &= b^k + b - PAu^{k+1}. \end{cases} \tag{4.20}$$

This is equivalent to the following algorithm.

Algorithm II (preconditioned Bregmanized operator splitting).

$$\begin{cases} v^{k+1} &= u^k - \delta A^T (AA^T + \epsilon)^{-1} (Au^k - f^k), \\ u^{k+1} &= \arg\min_u \left(\mu J_w(u) + \frac{1}{2\delta}||u - v^{k+1}||^2\right), \\ f^{k+1} &= f^k + f - Au^{k+1}. \end{cases} \tag{4.21}$$

According to Theorem 3.1, the condition for the convergence of Algorithm II is $0 < \delta < \frac{1}{\|B^T B\|}$, that is,
$$0 < \delta < \frac{1}{\|A^T (AA^T + \epsilon)^{-1} A\|}.$$

In the following, we discuss the computation for $v^{k+1}$ in Algorithm II (PBOS), which is obtained by inverting the operator $(AA^T + \epsilon)$ based on two specific applications.

- *Compressive sensing with partial Fourier measurement.* In this case, the operator $A = R\mathcal{F}$, where $\mathcal{F}$ represents the Fourier transform matrix $(n \times n)$ and $R$ represents a "row-selector" matrix $(m \times n)$, which could be represented as a binary matrix. Then $A^T A = \mathcal{F}^{-1} R^T R \mathcal{F}$. And the pseudoinverse $A^+ = A^T (AA^T)^{-1}$ is equal to $A^T$. Thus when $\epsilon = 0$, the algorithm is equivalent to Algorithm I.
- *Deconvolution.* We assume that $A$ is an invariant circular convolution matrix, and therefore the matrix $A$ is diagonalizable in a Fourier basis as
$$A = \mathcal{F}^{-1} \text{diag}(H) \mathcal{F},$$

where $H(\omega)$ is the Fourier transform of a kernel function $h$ and $\mathrm{diag}(H)$ is the diagonal matrix with $H$ as the main diagonal vector. In general, the matrix $A$ is not full row rank. As we mentioned above, the left and right pseudoinverse approximations are equal, i.e.,

$$A^T(AA^T+\epsilon)^{-1} = (A^TA+\epsilon)^{-1}A^T,$$

and the latter is equivalent to solving a Tikhonov regularization:

$$v^{k+1} = \arg\min_v \left( ||Av - f^{k+1}||^2 + \frac{\delta}{2}||v-u^k||^2 \right).$$

Then the solution $v^{k+1}$ can be computed via the fast Fourier transform:

$$v^{k+1} = u^k - \delta\mathcal{F}^{-1}\left( \frac{H^*(\omega)\cdot(G^{k+1}(\omega) - H(\omega)\cdot U^{k+1}(\omega))}{|H(\omega)|^2+\frac{1}{\delta}} \right), \tag{4.22}$$

where $G^k(\omega)$ and $U^k(\omega)$ are discrete Fourier transform coefficients of $f^k$ and $u^k$ at frequency $\omega$, and $H^*(\omega)$ is the conjugate transpose of $H(\omega)$. Consequently, implementing (4.22) requires only $O(N^2\log N)$ operations for an $N\times N$ image.

When the operator $A$ is not diagonalizable, a general quadratic minimization algorithm, such as a preconditioned conjugate gradient, can be applied to solve efficiently for $v^{k+1}$.

**4.2.4. Split Bregman for nonlocal TV denoising.** We can see that the efficiency of the BOS and the PBOS algorithms depends on solvers for the ROF-like subproblem. Here we focus on fast algorithms to minimize the NLTV functional defined in (4.9) in the extended NL-ROF model [44]:

$$\min_u \left( \mu J_w(u) + \frac{1}{2}||u-v||^2 \right), \tag{4.23}$$

where $w$ is a fixed weight function and $\mu > 0$ for a given image $v$. Notice that the algorithms for the NL-ROF model are extended from the fast algorithms originally developed for solving classical TV-based regularization problems. In particular, we extend the split Bregman method proposed by Goldstein and Osher in [30] to the NL case.

The main idea of the split Bregman algorithm is to transform the TV minimization problem into an $\ell^1$ norm minimization by introducing an auxiliary variable for the gradient of $u$, and then an efficient thresholding algorithm can be applied [30]. Here, we extend the split Bregman algorithm to the NLTV regularization by considering the related discrete problem:

$$\min_u \left( \mu|\nabla_w u|_1 + \frac{1}{2}||u-v||^2 \right). \tag{4.24}$$

The idea is to reformulate the problem as

$$\min_{u,d} \left( \mu|d|_1 + \frac{1}{2}||u-v||^2 \right) \quad \text{s.t.} \quad d = \nabla_w u. \tag{4.25}$$

By enforcing the constraint with the Bregman iteration process, the extended NL split Bregman algorithm uses the NLTV norm instead of the standard TV norm, and the algorithm scheme is given by

$$(4.26)\qquad \begin{aligned}(u^{k+1}, d^{k+1}) &= \arg\min_{u,d}\left(\mu|d|_1 + \frac{1}{2}||u-v||^2 + \frac{\lambda}{2}||d-\nabla_w u - b^k||^2\right),\\ b^{k+1} &= b^k + \nabla_w u^{k+1} - d^{k+1}.\end{aligned}$$

The solution of (4.26) is obtained by performing an alternating minimization process:

$$(4.27)\qquad \begin{aligned}u^{k+1} &= \arg\min_{u}\left(\frac{1}{2}||u-v||^2 + \frac{\lambda}{2}||d^k-\nabla_w u - b^k||^2\right),\\ d^{k+1} &= \arg\min_{d}\left(\mu|d|_1 + \frac{\lambda}{2}||d-\nabla_w u^{k+1} - b^k||^2\right).\end{aligned}$$

Note the equivalence of the alternating split Bregman method and the classical Douglas–Rachford splitting method [19, 20] recently shown by Setzer in [45]; thus the convergence is clarified.

Now, the subproblem for $u^{k+1}$ consists in solving the linear system

$$(4.28)\qquad (u^{k+1}-v) - \lambda \mathrm{div}_w(\nabla_w u^{k+1} + b^k - d^k) = 0,$$

which provides

$$u^{k+1} = (1-2\lambda\Delta_w)^{-1}(v + \lambda \mathrm{div}_w(b^k - d^k)).$$

Since the graph Laplacian $\Delta_w$ is negative semidefinite and the operator $1-2\lambda\Delta_w$ is diagonally dominant with NL weight $w$, therefore we can solve $u^{k+1}$ by a Gauss–Seidel algorithm. Similarly to [30], the vector $d^{k+1}$ is obtained by applying the shrinkage operator (2.7) for the vector field at each point $j$:

$$d_j^{k+1} = \mathrm{shrink}\left((\nabla_w u^{k+1} + b^k)_j, \frac{\mu}{\lambda}\right),$$

where $\mathrm{shrink}(p,\tau) = \frac{p}{|p|}\max\{|p|;\tau\}$ for each vector $p$.

**4.3. Algorithms.** To conclude this section, we describe the split Bregman method for the NLTV-ROF model and the BOS and PBOS algorithms presented above.

**Algorithm 1.** Split Bregman Method for Nonlocal TV Denoising.

**Initialization:** : $u^0 = v^0 = 0$, $\mu, \lambda, K$.
**for** $k = 0$ to $K$ **do**
  Solve $u^{k+1} = (1-2\lambda\Delta_w)^{-1}(v + \lambda \mathrm{div}_w(b^k - d^k))$ by the Gauss–Seidel method.
  Solve $d_j^{k+1} = \mathrm{shrink}((\nabla_w u^{k+1} + b^k)_j, \frac{\mu}{\lambda})$.
  $b^{k+1} = b^k + \nabla_w u^{k+1} - d^{k+1}$.
**end for**

## 5. Experimental results.

We present two applications: compressive sensing with Fourier measurements and image deconvolution. We compare the NLH1 and the NLTV with standard TV regularization, and wavelet-based $\ell^1$ regularization with the GPSR[1] algorithm [25].

[1] See http://www.lx.it.pt/∼mtf/GPSR.

**Algorithm 2.** Bregmanized Nonlocal Regularization for Inverse Problems (Algorithm I/II).

**Initialization:** : $u^0 = v^0 = 0$, $f^0 = f$, $h_0$, $\mu$, $\delta$, $nOuter$, $nUpdate$, $nInner$, $btol$.
**while** $k < nOuter$ and $\|Au^k - f\| > btol$ **do**
  Compute $v^{k+1}$ according to the method:
  **if** type = 'BOS' **then**
    $v^{k+1} = u^k - \delta A^T(Au^k - f^k)$
  **else if** type = 'PBOS' **then**
    $v^{k+1} = u^k - A^T(AA^T + \epsilon)^{-1}(Au^k - f^k)$
  **end if**
  **if** ($nUpdate > 0$ and mod $(k, nUpdate) = 0$) **then**
    (Update weight) Update the NL weight $w^{(k)} = w(v^{k+1}, h_0)$ using formula (4.1)
  **end if**
  Inner denoising step: Performing $nInner$ steps of the NLTV denoising iteration with input $v^{k+1}, \mu\delta$.
  Update $f^{k+1} = f^k + f - Au^{k+1}$.
  Increase $k$.
**end while**

In order to improve computational time and storage efficiency, we compute only the "best" neighbors; that is, for each pixel $x$, we include only the $K = 10$ best neighbors in the semilocal searching window of $21 \times 21$ centered at $x$ and the 4 nearest neighbors in comparing $5 \times 5$ patches with formula (4.1). For the TV and the NL regularization, we apply the BOS (PBOS) algorithm. For the TV-ROF denoising step, we use the split Bregman denoising algorithm,[2] and we implement the adapted split Bregman algorithm above (Algorithm 1) for the NLTV regularization. Similarly, a Gauss–Seidel method is applied to solve the NLH1 regularization. A MATLAB and MEX implementation of the proposed algorithms is available online.[3] For all the experiments, the inner denoising steps for both NLTV and NLH1 are fixed as $nInner = 20$ steps with the parameter $\delta = 1$.

**5.1. Nonlocal TV deconvolution.** We test both the BOS and PBOS methods on the Cameraman image for image deconvolution problems. As in [35], a fixed weight ($nUpdate= 0$) computed from a Tikhonov-based deblurred image $u^0$ (see (4.16)) is used for all the NL methods. Using optimal $\lambda$ and the noise level $\sigma$, we can obtain $u^0$ very efficiently with an estimated noise level $\sigma_1$ [35]. We set $h_0 = 2\sigma_1$. In [35], a gradient descent algorithm was applied to solve the unconstrained Lagrangian formulation:

$$\min\left(|\nabla_{w_0} u|_1 + \frac{\lambda}{2}||Au - f||^2\right). \tag{5.1}$$

This algorithm is generally very slow. Instead, in this paper, we solve a constrained minimization problem,

$$\min |\nabla_{w_0} u|_1 \quad \text{s.t.} \quad ||Au - f||^2 \le \sigma^2,$$

[2]See http://www.math.ucla.edu/∼tagoldst/public_codes/splitBregmanROF_m ex.zip.
[3]See http://www.math.ucla.edu/∼xqzhang/html/code.html.

by using the BOS and the PBOS until the residual noise level is around $\sigma$. We set the stopping criterion $btol = 0.99\sigma$ and the maximum Bregman iteration $nOuter = 30$. For the wavelet-based restoration, we use the $daubcqf(4)$ wavelet[4] with maximum decomposition level and with the scale parameter $\tau = 0.2$ as inputs for the GPSR code.

Figure 1 compares different algorithms. For the PBOS algorithm, the regularization parameter $\epsilon = 0.1$. The images reconstructed by NLTV (NLTV+gradient descent, NLTV+BOS, NLTV+PBOS) present better contrasts and edges compared to wavelet-, TV-, and NLH1-based methods. Compared to the algorithm in [35], the reconstruction results are similar, while the computation speed is improved. Note that the weight function was computed in the whole searching window in [35], while here only the 10 best and 4 nearest neighbors are used for each pixel. Furthermore, the algorithms BOS and PBOS take fewer than $nOuter$ steps to meet the stopping criterion. Overall, the NLTV+BOS algorithm stops with 25 steps for 138 seconds, and the PBOS stops at 8 steps for 51 seconds including weight computation, compared to 280 seconds with 500 steps with the gradient descent algorithm for solving (5.1).

We also tested the weight updating scheme: it appears that there is no improvement compared to a fixed weight function. In fact, a simple weight updating scheme tends to recover a smoother image. One explanation is that the weight function computed from a pre-deblurred image is good enough to express structured information in the NLTV regularization, while weight updating degrades the image structures.

**5.2. Compressive sensing.** In this section, we focus on exploring the sparsity of natural images with NL regularization operators. The compressive sensing matrix we choose is $A = R\mathcal{F}$, where $R$ is a row-selector matrix, and $\mathcal{F}$ is a Fourier transform matrix. For an $N \times N$ image, we randomly choose $m$ coefficients; then $R$ is a sampling matrix of size $m \times (N^2)$ with $m = 0.3$. We consider only the BOS algorithm since $A^T = A^+$, as discussed in section 4.2.3.

Figures 2 and 3 present the results for the Barbara picture and a composed texture picture. The weight parameter $h_0$ is empirically chosen as $h_0 = 20$ for the Barbara example (see Figure 2) and $h_0 = 15$ for the patch example (Figure 3). For this application, an initial guess by setting unknowns to be zeros hardly reveals right structures of true images. Hence, the weight updating strategy is necessary for this application; in particular, we update the weight every $nUpdate = 20$ steps. Experimentally, the update of weight is stable. As expected, the standard TV regularization is not capable of recovering texture patterns presented in these images. The results based on the wavelet method are obtained by using a $daubqf(8)$ wavelet with maximum decomposition level and an empirically optimal thresholding parameter with GPSR code. Since there is no noise considered in these two examples, we solve the equality constrained problem by activating the continuation and the debias options in the GPSR code. The residual stopping tolerance $btol$ is set as $10^{-5}$ for all of the BOS-based algorithms. The maximal outer iteration $nOuter$ for TV is set as 100 for both examples since the algorithm attains a steady state. For NLH1 and NLTV with weight updating, it is harder to determine a good iteration number. In fact, the peak signal to noise ratio (PSNR) of NLH1 is decreasing after a certain number of iterations. Empirically we choose $nOuter = 100$ for NLH1 as the optimal result for both examples, $nOuter = 500$ for NLTV in Figure 2 (the PSNR is still significantly increasing after 100 steps), and $nOuter = 100$ for NLTV in

[4]See http://dsp.rice.edu/software/rice-wavelet-toolbox.


**Figure 1.** *Deconvolution example on* $256 \times 256$ *Cameraman degraded with the* $9 \times 9$ *box average kernel and Gaussian noise* $\sigma = 3$. *Weight fixed.*


**Figure 2.** *Compressive sensing example: Barbara (*$256 \times 256$*), 30% randomly chosen Fourier coefficients, noiseless. Weight updated.*

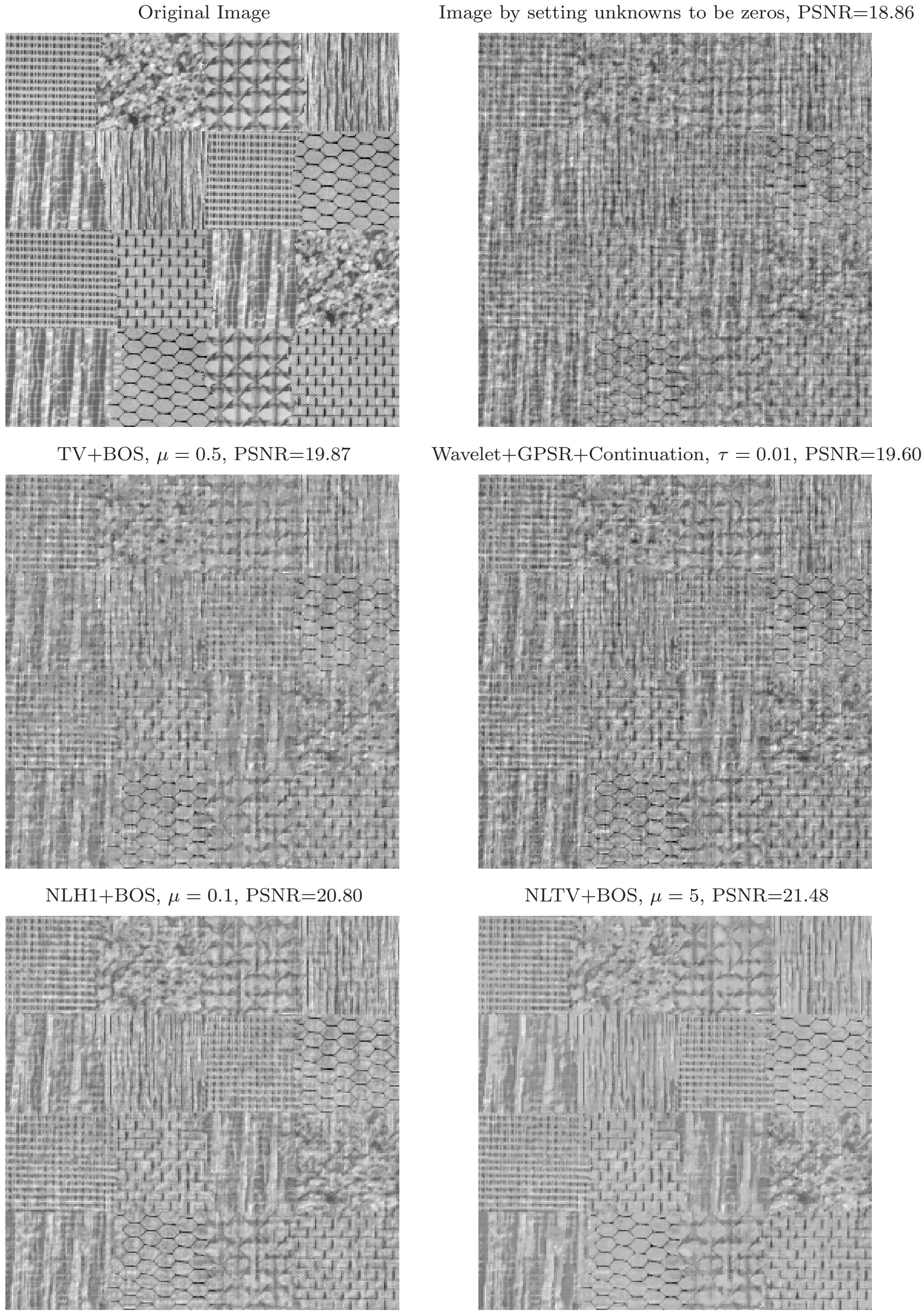


**Figure 3.** *Compressive sensing example: Textures (*$256 \times 256$*), 30% randomly chosen Fourier coefficients, noiseless. Weight updated.*

Figure 3, respectively. Surprisingly, with only a few measurements, the image textures are almost perfectly reconstructed by the NLTV regularization. This is because image structures are expressed implicitly in the NL weight function, and the NL regularization process with Bregman iteration provides an efficient way to recover textures without explicitly constructing a basis. Note that with fewer outer iterations for the NLTV, we can still obtain an improved result compared to other regularization methods, which leads to a faster reconstruction.

**6. Discussion.** In this paper, we propose a general algorithm framework for convex minimization problems with equality constraints. This simple algorithm framework overcomes the uncertainty and the efficiency of inner iterations involved in the Bregman iteration. In particular, we solve the compressive sensing problem for sparse reconstruction and the image deconvolution problem using the NLTV functional. Experiments show that the NLTV regularization is efficient in recovering natural images with few measurements without using a basis or dictionary learning. We also make the same observation as in [30]: the edges are quickly set after a small number of iterations. In the case of deconvolution, the algorithm converges very quickly using a small number of denoising steps and Bregman iteration. Finally, the proposed algorithms can in theory be applied for other inverse problems and regularization. We will investigate this question more carefully in the future. Furthermore, as mentioned in [41], it is also important to better understand the weight updating strategy in a theoretical framework.

**Acknowledgments.** We thank the authors of the codes used in this paper and the reviewers for their useful suggestions for improving the presentation of this paper. Martin Burger and Stanley Osher thank Fondazione CIME for a summer school in a stimulating atmosphere, which initiated a part of this project.

## REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein, *K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation*, IEEE Trans. Signal Process., 54 (2006), pp. 4311–4322.

[2] P. Arias, V. Caselles, and G. Sapiro, *A variational framework for non-local image inpainting*, in Proceedings of the 7th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Comput. Sci. 5681, D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, eds., Springer, Berlin, Heidelberg, 2009, pp. 345–358.

[3] T. Brox and D. Cremers, *Iterated nonlocal means for texture restoration*, in Scale Space and Variational Methods in Computer Vision, Lecture Notes in Comput. Sci. 4485, Springer, Berlin, Heidelberg, 2007, pp. 13–24.

[4] A. Buades, B. Coll, and J. M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4 (2005), pp. 490–530.

[5] A. Buades, B. Coll, and J. M. Morel, *Image Enhancement by Non-local Reverse Heat Equation*, Technical report 22, CMLA, ENS-Cachan, Cachan, France, 2006.

[6] M. Burger and S. Osher, *A guide to TV Zoo* II: *Computational methods*, in Level Set and PDE-Based Reconstruction Methods, Lecture Notes in Math., Springer, New York, to appear.

[7] J.-F. Cai, S. Osher, and Z. Shen, *Convergence of the linearized Bregman iteration for $\ell_1$-norm minimization*, Math. Comp., 78 (2009), pp. 2127–2136.

[8] J.-F. Cai, S. Osher, and Z. Shen, *Linearized Bregman iterations for frame-based image deblurring*, SIAM J. Imaging Sci., 2 (2009), pp. 226–252.

[9] E. J. Candès and D. L. Donoho, *Curvelets—a surprisingly effective nonadaptive representation for objects with edges*, in Curves and Surface Fitting: Saint-Malo 1999, Vanderbilt University Press, Nashville, TN, 2000, pp. 105–120.

[10] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[11] E. J. Candès and M. Wakin, *An introduction to compressive sampling*, IEEE Signal Processing Magazine, 25 (2) (2008), pp. 21–30.

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

[13] F. R. K. Chung, *Spectral Graph Theory*, AMS, Providence, RI, 1997.

[14] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[15] J. Darbon and M. Sigelle, *Exact optimization of discrete constrained total variation minimization problems*, in Combinatorial Image Analysis, Lecture Notes in Comput. Sci. 3322, Springer, Berlin, 2004, pp. 548–557.

[16] I. Daubechies, M. Defrise, and C. D. Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pure Appl. Math., 57 (2004), pp. 1413–1457.

[17] M. N. Do and M. Vetterli, *The contourlet transform: An efficient directional multiresolution image representation*, IEEE Trans. Image Process., 14 (2005), pp. 2091–2106.

[18] D. L. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 101–126.

[19] J. Douglas and H. Rachford, *On the numerical solutions of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 1 (1976), pp. 97–116.

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

[21] A. A. Efros and T. K. Leung, *Texture synthesis by non-parametric sampling*, in Proceedings of the International Conference on Computer Vision, Vol. 2, IEEE Computer Society, Washington, DC, 1999, pp. 1033–1038.

[22] M. Elad, P. Milanfar, and R. Rubinstein, *Analysis versus synthesis in signal priors*, Inverse Problems, 23 (2007), pp. 947–968.

[23] A. Elmoataz, O. Lezoray, and S. Bougleux, *Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing*, IEEE Trans. Image Process., 17 (2008), pp. 1047–1060.

[24] M. A. T. Figueiredo and R. D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process., 12 (2003), pp. 906–916.

[25] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE J. Sel. Top. Signal Process., 1 (2007), pp. 586–598.

[26] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Valued Problems, North-Holland, Amsterdam, 1983, pp. 299–331.

[27] G. Gilboa and S. Osher, *Nonlocal operators with applications to image processing*, Multiscale Model. Simul., 7 (2008), pp. 1005–1028.

[28] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator Splitting Methods in Nonlinear Mechanics*, SIAM Stud. Appl. Math. 9, SIAM, Philadelphia, 1989.

[29] D. Goldfarb and W. Yin, *Parametric Maximum Flow Algorithms for Fast Total Variation Minimization*, CAAM Technical report TR07-09, Rice University, Houston, TX, 2007.

[30] T. Goldstein and S. Osher, *The split Bregman method for $L1$-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.

[31] E. T. Hale, W. Yin, and Y. Zhang, *Fixed-point continuation for $L1$-regularized minimization: Methodology and convergence*, SIAM J. Optim., 19 (2008), pp. 1107–1130.

[32] D. S. Hochbaum, *An efficient algorithm for image segmentation, Markov random fields and related problems*, J. ACM, 48 (2001), pp. 686–701.

[33] S. Kindermann, S. Osher, and P. W. Jones, *Deblurring and denoising of images by nonlocal functionals*, Multiscale Model. Simul., 4 (2005), pp. 1091–1115.

[34] C. Lemaréchal and C. Sagastizábal, *Practical aspects of the Moreau–Yosida regularization: Theoretical preliminaries*, SIAM J. Optim., 7 (1997), pp. 367–385.

[35] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, *Image recovery via nonlocal operators*, J. Sci. Comput., 42 (2010), pp. 185–197.

[36] M. Lustig, S. Boyd, S.-J. Kim, K. Koh, and D. Gorinevsky, *An interior-point method for large-scale $L_1$-regularized least squares*, IEEE J. Sel. Top. Signal Process., 1 (2007), pp. 606–617.
[37] J.-J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C. R. Acad. Sci. Paris, 255 (1962), pp. 2897–2899.
[38] 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 (2005), pp. 460–489.
[39] S. Osher, Y. Mao, B. Dong, and W. Yin, *Fast linearized Bregman iteration for compressive sensing and sparse denoising*, Commun. Math. Sci., 8 (2010), pp. 93–111.
[40] G. Peyré, *Image processing with nonlocal spectral bases*, Multiscale Model. Simul., 7 (2008), pp. 703–730.
[41] G. Peyré, S. Bougleux, and L. Cohen, *Non-local regularization of inverse problems*, in ECCV 2008, Part III, Lecture Notes in Comput. Sci. 5304, Springer, Berlin, Heidelberg, 2008, pp. 57–68.
[42] R. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.
[43] J. Romberg, *Imaging via compressive sampling*, IEEE Signal Processing Magazine, 25 (2) (2008), pp. 14–20.
[44] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[45] S. Setzer, *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage*, in Scale Space and Variational Methods in Computer Vision, Lecture Notes in Comput. Sci. 5567, X.-C. Tai, K. Morten, M. Lysaker, and K.-A. Lie, eds., Springer, Berlin, 2009, pp. 464–476.
[46] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138.
[47] W. Yin, *Analysis and Generalizations of the Linearized Bregman Method*, CAAM report TR09-02, Rice University, Houston, TX, 2008.
[48] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for $\ell^1$-minimization with applications to compressed sensing*, SIAM J. Imaging Sci., 1 (2008), pp. 143–168.
[49] D. Zhou and B. Schölkopf, *Regularization on discrete spaces*, in Pattern Recognition, Proceedings of the 27th DAGM Symposium, Springer, Berlin, 2005, pp. 361–368.
[50] W. Zulehner, *Analysis of iterative methods for saddle point problems: A unified approach*, Math. Comput., 71 (2001), pp. 479–505.