# Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction[*]

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

### Abstract

Bregman methods introduced in [35] to image processing are demonstrated to be an efficient optimization method for solving sparse reconstruction with convex functionals, such as $l_1$ and $TV$ norm [45, 28]. 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 problems regularization with a single forward-backward operator splitting step [12], 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 efficiency 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. For example, the sparse reconstruction of Barbara using nonlocal TV provides a PSNR of 20.37, whereas standard TV method and wavelet-based method give respectively a PSNR of 16.41 and 16.21. This shows that the nonlocal TV regularization itself can sparsify textured images and the Bregman iteration method is an efficient method for sparse signal recovery.

## 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 a $m \times n$ linear operator, typically a convolution operator in the deconvolution problem or a sub-sampling measurement operator in the compressive sensing problem.

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. A well-known example of regularized inverse problems 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 finally $|| \cdot ||$ denotes the $l^2$ norm. The notation $|| \cdot ||$ for the $l^2$ 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 ROF model proposed in [40]. This model uses the TV regularization functional since images are assumed to have bounded variation, which is the case for piecewise constant images. The generalized ROF model is defined by the following unconstrained minimization problem [5]:

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

where $\nabla u$ is the weak gradient of $u$, $| \cdot |_1$ denotes the $l^1$ norm, and $|\nabla u|_1$ is the total variation of $u$.

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 $n$, $D \in \mathbb{R}^{m \times n}$ is a given frame or a dictionary matrix, there are two different formulations for solving the problem: analysis based or synthesis based[21]. The analysis based model is formulated as

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

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

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

and the solution is $u^* = D\alpha^*$. If $D$ is an orthogonal basis, then the two models are equivalent. The analysis based model (1) is largely used for inverse problems, such as the wavelet-vaguelette decomposition model defined in [13]. Usually, a scale-dependent shrinkage is employed to estimate the image wavelet coefficients. The synthesis based model has been developing in the domain of compressive sensing problems [9]. Several efficient algorithms, such as the Iterative Soft Shrinkage proposed (IST)[16], $l1\_ls$[41] Gradient Projection for Sparse Reconstruction (GPSR) [23], Fixed-Point Continuation Method(FPC) [29], and linearized Bregman[36, 6, 7, 44], 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. Compressive sensing relies on two important principles to reconstruct signals: sparsity, which restricts the signal of interest, and incoherence, which is usually revealed by irregularly sampled measurements. 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 find the best basis for natural signals/images to fit the theory of compressive sensing, such as curvelets[8], 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 with amplified noise and produce undesirable artifacts. Furthermore, it is difficult to choose a proper basis for different images.

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

$$\min_u J(u) \ \text{s.t.} \ Au = f. \tag{3}$$

where $J$ is a general convex functional. This problem (4) is shown to cover a wide range of signal and image processing task for various choices of the convex functionals $J$ and $A$, including $l_1$ compressive sensing[45] and image restoration by total variation [35]. The algorithms proposed in this paper are based on the Bregman iteration introduced in [35] and the proximal forward backward operator splitting method [24, 12, 29]. Note that if there is noise present in the measurements, we can use a discrepancy stopping criterion as in the original Bregman iteration[35], that is $\|Au^k - f\| \leq \sigma$ with the same algorithm. The principle of our algorithms is to maximally decouple the minimization functionals. More specially, the overall algorithms consist of a two forward (explicit) gradient steps (one is the Bregman iteration step) and an implicit step equivalent to the ROF model [40], which we can often solve efficiently. The proposed algorithms can be also interpreted as inexact Uzawa methods used for linear saddle point problems [47, 26]. However, the convergence of our algorithm does not seem to be directly implied by the 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 for compressive sensing and deconvolution. Our experiments show that the proposed nonlocal regularization model can recover almost all the details of a textured image without explicitly choosing a basis compared to above dictionary based sparse representation algorithms. Our investigation demonstrates that the nonlocal TV regularization itself sparsifies textured image and the Bregman iteration is an efficient 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 (3). In Section 4 we present the nonlocal regularization and the application of proposed algorithm and a preconditioned one. An updating strategy for the weight function in the nonlocal TV term is discussed. Finally, 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 and some notations. We consider a general minimization problem as:

$$\min_u J(u) \ \text{ s.t. } \ H(u) = 0, \tag{4}$$

where $J$ and $H$ are both convex functionals defined over $\mathbb{R}^n \to \mathbb{R}^+$. It is well-known that this problem is difficult to be solved numerically when $J$ is non-differentiable. An efficient method to solve this constrained minimization problem is to use the Bregman iteration, initially introduced to imaging in [35] to improve the ROF denoising models [40].

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:

$$D_J^p(u, v) = J(u) - J(v) - \langle p, \ u - v \rangle, \tag{5}$$

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 (5), the original constrained minimization problem (4) can be solved by the following iterative scheme:

$$\begin{cases} u^{k+1} &= \min_u \left( \mu D_J^{p^k}(u, u^k) + H(u) \right) \\ p^{k+1} &= p^k - \partial H(u^{k+1}) \end{cases}$$

where $\mu > 0$ and $\partial H(u^{k+1})$ denotes a subgradient of $H$ at $u^{k+1}$. To solve the constrained minimization problem: (3), we choose $H(u) = \frac{1}{2}||Au - f||^2$ and we obtain a two-step Bregman iterative scheme [35]:

$$\begin{cases} u^{k+1} &= \min_u \left( \mu J(u) + \frac{1}{2}||Au - f^k||^2 \right) \\ f^{k+1} &= f^k + f - Au^{k+1} \end{cases} \tag{6}$$

It is shown in [35] that the sequence $u^k$ weakly converges to a solution of (3), and the residual $||Au^k - f||$ of the sequence generated by (6) converges to zero monotonically. Bregman iteration was successfully used in sparse reconstruction problems recently due to its speed, simplicity, efficiency and stability, see for example [30, 35, 14, 28, 45].

## 2.2 Forward-backward Operator Splitting

Operator splitting methods have been extensively studied in the optimization community, e.g. [24, 43, 12, 19, 26]. They aim to minimize the sum of two convex functionals:

$$\min_u \left( \mu J(u) + H(u) \right), \tag{7}$$

where $\mu > 0$. In [12], Combettes and Wajs introduced the forward-backward technique based on the proximal operator for general signal recovery tasks. The proximal operator of a convex functional $J$ of a function $v$, which was originally introduced by Moreau in [34], is defined as :

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

By classical arguments of convex analysis, the solution of (7) 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 and backward splitting algorithm:

$$u^{k+1} = \text{Prox}_{\delta \mu J}(u^k - \delta \partial H(u^k)), \tag{8}$$

where the proximal operator $\text{Prox}_{\delta \mu J}(v)$ is defined as:

$$\arg \min_u \left( \mu J(u) + \frac{1}{2\delta}||u - v||^2 \right). \tag{9}$$

Also in [12], a general convergence is established for the generic problem. More specially, in the case of $H(u) = \frac{1}{2}||Au - f||^2$, the algorithm converges when $0 < \delta < \frac{2}{||A^T A||}$. And the solution of the minimization problem (7) can be simplified as the following two-step algorithm:

$$\begin{cases} v^{k+1} &= u^k - \delta A^T(Au^k - f) \\ u^{k+1} &= \arg \min_u \left( \mu J(u) + \frac{1}{2\delta}||u - v^{k+1}||^2 \right) \end{cases} \tag{10}$$

The main advantage of this algorithm is that the two functionals are decoupled. Furthermore, the proximal minimization (9) is strictly convex, and then there exists a unique minimizer. In practice, the proximal operator solution (9) has well known solutions for some models. For example, when the regularization functional $J$ is the $l^1$ norm of $u$, i.e. $J(u) = |u|_1$, then the solution is obtained by a soft shrinkage operator [12, 29, 6] as follows:

$$u = \text{shrink}(v, \mu\delta) = \text{sign}(v) \max\{|v| - \mu\delta, 0\}. \tag{11}$$

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 [10], the split Bregman method [28] or by graph cuts in the anisotropic case[30, 15, 14, 27].

## 2.3 Linearized Bregman

The idea of the linearized Bregman iteration [14, 45] is to combine Bregman iteration and operator splitting to solve the constrained problem (3) for $l^1$ sparse reconstruction. Given $u^0 = 0 = p^0$, the iterative algorithm of linearized Bregman is defined for $k \geq 0$ as:

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

where $D_J^{p^k}(u, u^k)$ is the Bregman distance if $u$ and $u^k$ with relative to $p^k$ defined in (5). This minimization can be rewritten in a simpler formulation as follows:

$$\begin{cases} v^{k+1} & = & v^k - \delta A^T(Au^k - f) \\ u^{k+1} & = & \arg\min\left(\mu J(u) + \frac{1}{2\delta}||u - v^{k+1}||^2\right) \end{cases} \quad (12)$$

The difference between linearized Bregman and the operator splitting method (10) is in the way of we update $v^{k+1}$. They solve different problems. In fact, Cai *et al.* proved the following propositions in [6]:

**Proposition 1.** *If the sequence $u^k$ converges and $p^k$ is bounded, then the limit of $u^k$ is the unique solution of*

$$\min\left(\mu J(u) + \frac{1}{2\delta}||u||^2\right) \quad s.t. \ \ Au = f. \quad (13)$$

In the case of $l^1$ sparse approximation, algorithm (12) can be written as follows:

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

As $\mu \to \infty$, the solution of (13) tends to the solution of (3); even better, it was proved in [44] that, for $\mu$ large enough, the limit solution solves the original problem:

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

## 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 (3) by the Bregman iteration and operator splitting introduced in Section 2.1 and Section 2.2. First of all, the equality constraint in (3) is enforced with the Bregman iteration process:

$$\begin{cases} u^{k+1} & = & \min_u\left(\mu J(u) + \frac{1}{2}||Au - f^k||^2\right) \\ f^{k+1} & = & f^k + f - Au^{k+1} \end{cases} \quad (14)$$

The first subproblem could be sometimes 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 (14) as follows: for $i \geq 0, u^{k+1,0} = u^k$,

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

for a positive number $0 < \delta < \dfrac{2}{||A^T A||}$. 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 to use only one inner iteration, which leads to the algorithm I:

**Algorithm I (Bregmanized Operator Splitting):**

$$\begin{cases} v^{k+1} &= u^k - \delta A^T(Au^k - f^k) \\ u^{k+1} &= \arg\min_u \left( \mu J(u) + \frac{1}{2\delta}||u - v^{k+1}||^2 \right) \\ f^{k+1} &= f^k + f - Au^{k+1} \end{cases} \quad (15)$$

and it is equivalent to:

$$\begin{cases} u^{k+1} &= \arg\min_u \left( \mu J(u) + \frac{1}{2\delta}||u - ((1 - \delta A^T A)u^k + \delta A^T f^k)||^2 \right) \\ f^{k+1} &= f^k + f - Au^{k+1} \end{cases} \quad (16)$$

## 3.2 Connections with existing methods

The above algorithm can be interpreted as an inexact Uzawa method [47, 26] applied to the augmented Lagrangian [39] of the original problem as follows:

$$L(u,p) = \mu J(u) + \frac{1}{2}||Au - f||^2 - (Au - f)^T(p - f). \quad (17)$$

where $p$ is a Lagrange multiplier of the original problem (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 [47] and Moreau-Yosida proximal point iteration [32] on this formulation, we get the following algorithm:

$$\begin{cases} \text{step 1:} & u^{k+1} = \min_u \left( \mu J(u) + \frac{1}{2}||Au - f||^2 - \langle Au - f, p^k - f \rangle + ||u - u^k||_D^2 \right) \\ \text{step 2:} & p^{k+1} = p^k - (Au^{k+1} - f) \end{cases} \quad (18)$$

where $D$ is a positive-definite matrix. The sequence $(u^k, p^k)$ generated by (18) provides us:

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

where $s^k \in \partial J(u^k)$. When $D = \frac{1}{\delta} - A^T A$ and change the variable $p^k$ as $f^k$, we get the BOS algorithm defined in (38).

Most analysis for inexact Uzawa methods is available for linear saddle point problems with strong convexity assumptions [47, 26]. Other available analysis based on the augmented Lagrangian methods [26, 39] is also different from ours due to the maximally decoupled structures. Note that the presented algorithm is different from the split Bregman algorithm [28] in the manner of splitting, for the latter can be recast as a Douglas-Rachford algorithm [18, 19, 42]. On the other side, 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 with proximal point algorithms, 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 (3) is closed, proper, semi-continuous and convex.

7

**Theorem 1.** *If* $0 < \delta < \frac{1}{\|A^T A\|}$, *let the sequence* $(u^k, p^k)$ *be generated by **Algorithm I** given in* (16). *Then every accumulation point of* $u^k$ *is a solution of* (3).

**Proof:** We first consider a Lagrangian formulation of the original constrained problem (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. If we denote

$$\overline{s} = -\frac{1}{\mu} A^T (f - \overline{p}), \tag{19}$$

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

$$\begin{cases} \mu \overline{s} + A^T(f - \overline{p}) &= 0 \\ A\overline{u} - f &= 0. \end{cases} \tag{20}$$

where $\overline{u}$ is an optimal solution and $\overline{p}$ is a Lagrangian multiplier respectively.

Start from $f^0 = 0$, let $(u^k, f^k)$ be the sequence (16) generated by **Algorithm I**, and as above write $s^{k+1} = -\frac{1}{\mu} A^T(f - f^{k+1})$, the sequences satisfy:

$$\begin{cases} \mu s^{k+1} + \frac{1}{\delta} u^{k+1} &= (\frac{1}{\delta} - A^T A) u^k + A^T f^k \\ f^{k+1} &= f^k + f - A u^{k+1} \end{cases}. \tag{21}$$

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:

$$\begin{cases} \mu s^{k+1} + L u^{k+1} - A^T f^{k+1} &= L u^k - A^T f \\ f^{k+1} &= f^k + f - A u^{k+1} \end{cases}. \tag{22}$$

On the other side, 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 the equations (22) is rearranged in terms of the error differences as:

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

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

$$\begin{aligned} &\|\Delta u^{k+1}\|_L^2 + \|\Delta f^{k+1}\|^2 + \|u^{k+1} - u^k\|_L^2 + \|f^{k+1} - f^k\|^2 - \|\Delta u^k\|_L^2 - \|\Delta f^k\|^2 \\ =\ & 2\langle L(u^{k+1} - u^k), \Delta u^{k+1} \rangle + 2\langle f^{k+1} - f^k, \Delta f^{k+1} \rangle \\ =\ & 2\langle A^T \Delta f^{k+1}, \Delta u^{k+1} \rangle - 2\mu\langle \Delta s^{k+1}, \Delta u^{k+1} \rangle + 2\langle f - A u^{k+1}, \Delta f^{k+1} \rangle \\ =\ & -2\mu\langle \Delta s^{k+1}, \Delta u^{k+1} \rangle. \end{aligned} \tag{23}$$

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

$$\langle \Delta s^{k+1}, \Delta u^{k+1} \rangle = \langle s^{k+1} - \overline{s}, u^{k+1} - u^k \rangle = D_J^{\overline{s}}(u^{k+1}, \overline{u}) + D_J^{s^{k+1}}(\overline{u}, u^{k+1}) \geq 0, \quad \forall k. \quad (24)$$

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 sequence $u^k$ and $f^k$ are bounded and there exists a convergent subsequence of $(u^k, f^k)$. Secondly, by summing the equality (23), we obtain

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

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 one 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 (24) implies that the non-negative Bregman distance satisfy

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

Using (19) and $Au^{k+1} \to f = A\overline{u}$, we have

$$0 = \lim_{k \to \infty} \left( \mu J(u^{k+1}) - \mu J(\overline{u}) + \langle f - \overline{p}, A(u^{k+1} - \overline{u}) \rangle \right) = \lim_{k \to \infty} \mu J(u^{k+1}) - \mu J(\overline{u}) \quad (25)$$

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 semi-continuity of $J$. We conclude directly that $u^{\infty}$ is a solution of (3). $\square$

# 4  Nonlocal regularization

In this section, we first present some notations of the nonlocal regularization introduced in[25], and then discuss applications for solving inverse problems.

## 4.1  Background

In [20], 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 *et al.* introduced in [3] an efficient denoising model called nonlocal means (NL-means). The model consists in denoising a pixel by averaging the other

pixels with similar structures (patches) to 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

$$
\begin{aligned}
w(f, h_0)(x, y) &= \exp\{-\frac{G_a * (||f(x + \cdot) - f(y + \cdot)||^2)(0)}{2h_0^2}\}, & (26) \\
C(x) &= \int_\Omega \exp\{-\frac{G_a * (||f(x + \cdot) - f(y + \cdot)||^2)(0)}{2h_0^2}\} dy.
\end{aligned}
$$

and $G_a$ 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 non-local means filter is a linear operator. In the case where the reference image $f$ is chosen to be $u$, the operator is non-linear and it is the nonlocal means filter presented by Buades *et al.* in [3]. The definition of the weight function (26) shows that this function is significant only if the patch around $y$ has similar structure as the corresponding patch around $x$. This filter is very efficient to reduce noise while preserving textures and contrast of natural images. It is generally better to choose a reference image as close as possible to the true image to introduce in the weight function relevant information regarding image structures.

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 (26), and $D_f(i, i) = C(i)$ is a $N \times N$ diagonal matrix.

The application of the nonlocal 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 similar distribution and structures. Based on the hypothesis that the deblurred image must maintain the same coherence as the blurry image, Buades *e*t al. proposed in [4] a NL-means regularization energy for image deblurring defined as follows:

$$J_{NLM}(u) := ||u - NLM_f(u)||^2 \tag{27}$$

where $NLM_f := D_f^{-1} W_f$ is the nonlocal 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 [2]. The authors propose to minimize the functional:

$$J_{NLM}(u) := ||u - NLM_u(f)||^2. \tag{28}$$

It is a nonlinear model since the weight function depends on the unknown image $u$. The solution of (28) 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 nonlocal means filter in a variational framework, Kindermann *et al.* in [31] started to investigate the use of regularization functionals with nonlocal correlation terms for general inverse problems. Also, inspired from the graph Laplacian in [11], Gilboa and Osher defined a variational framework based nonlocal operators in [25]. Note that Zhou and Schölkopf in [46] and Elmoataz *et al.* in [22] also used graph Laplacian in the discrete setting for image denoising. Finally, the connection between the filtering methods, and spectral bases of the nonlocal graph Laplacian operator are discussed in [37] by Peyré.

In the following, we give the definitions of the nonlocal functionals introduced in [25]. Let $\Omega \subset \mathbb{R}^2, x \in \Omega$, and $u(x)$ be a real function $\Omega \to R$. Assume $w : \Omega \times \Omega \to \mathbb{R}$ is a nonnegative symmetric weight function defined in (26) from a reference image, then the nonlocal gradient $\nabla_w u(x)$ is defined as the vector of all partial differences $\nabla_w u(x, \cdot)$ at $x$ such that:

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

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:

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

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

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

The graph Laplacian is defined by:

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

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:

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

and negative semi-definite:

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

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

$$J_{NL/H1,w}(u) \quad := \quad \frac{1}{4}\int |\nabla_w u(x)|^2 dx \tag{29}$$

$$J_{NL/TV,w}(u) \quad := \quad \int_\Omega |\nabla_w u(x)| dx \tag{30}$$

The corresponding Euler-Lagrange equation of (29) and (30) are then written as:

$$-\int_\Omega (u(y) - u(x)w(x,y)dy = -\delta_w u(x) = 0 \tag{31}$$

and

$$-\int_\Omega (u(y) - u(x))w(x,y)\left[\frac{1}{|\nabla_w u(x)|} + \frac{1}{|\nabla_w u(y)|}\right]dy = 0. \tag{32}$$

Note that once the weight function $w$ is fixed, the Euler-Lagrange equation for the nonlocal $H^1$ is linear and can be solved by a gradient descent method. However, analogous to the classical total variation, the functional (30) 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 zero denominator. Finally, if the function $w(x,y)$ in (32) is chosen to be the nonlocal weight function defined in (26), then the nonlocal means filter is generalized to a variational framework. Nevertheless, the minimization of the nonlocal TV functional remains as a difficult optimization problem due to the computation complexity and the non-differentiability.

## 4.2 Nonlocal Regularization for Inverse Problems

### 4.2.1 Weight fixed

The nonlocal regularization for inverse problems is based on the following constrained formulation:

$$\min_u J_w(u) \quad \text{subject to } Au = f, \tag{33}$$

with $J_w$ being nonlocal regularization term, such as the nonlocal TV or the nonlocal $H^1$ with a given weight function $w$, and $A$ is a deconvolution or compressive sensing matrix. By applying the algorithm (2), we obtain the first algorithm proposed in this paper:

$$\begin{cases} v^{k+1} & = & u^k - \delta A^T(Au^k - f^k) \\ u^{k+1} & = & \arg\min_u \left(\mu J_w(u) + \frac{1}{2\delta}||u - v^{k+1}||^2\right) \\ f^{k+1} & = & f^k + f - Au^{k+1} \end{cases} \tag{34}$$

We can see that the key computation of this algorithm relies on the computation of $A$ and $A^T$, and the ROF like denoising step. In Section 4.2.4, we will present a fast method based on split bregman iteration for total variation minimization.

### 4.2.2 Weight Updating

In the previous discussion of nonlocal regularization methods, the weight function $w$ was fixed. In the denoising case, most of 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 (26) 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 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 $w(u^k, h_0)$ (26) during the reconstruction of signals. In [38], the authors have

also proposed to update the graph weight to solve inverse problems using the forward-backward operator splitting technique [12] to solve the relaxed Lagrangian formulation. Similar to [38], we consider a more appropriate problem:

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

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:

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

In practice, we actually do not need to update the weight function at every step. Instead, we update every $M$ steps, see Section 4.3.

### 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 [33], we estimate the weight function with the solution of the Tikhonov regularization problem:

$$v = \arg\min_v \left( \frac{1}{2}||Av - f||^2 + \frac{\epsilon}{2}||v||^2 \right), \tag{35}$$

where $\epsilon$ is a small positive number. It 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 [33], 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 nonlocal 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 (10):

$$\begin{cases} v^{k+1} = u^k - \delta A^+(Au^k - f) \\ u^{k+1} = \arg\min_u \left( \mu J_w(u) + \frac{1}{2\delta}||u - v^{k+1}||^2 \right) \end{cases}, \tag{36}$$

13

where $A^+$ is the pseudo-inverse of $A$ and $\delta > 0$. This similar idea is also considered in [7] for frame based image deblurring. The operator $A^+A$ is an orthogonal projector onto the range space of $A^+$, thus it is positive semi-definite. In the following, we replace $A^+$ by $A^T(AA^T + \epsilon)^{-1}$, then algorithm (36) solves the minimization problem

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

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

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

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

- If $A^TA = I$, i.e. $A^+ = A^T$, then

$$B = A, \quad b = f.$$

  Then the modified algorithm (36) is consistent with the classical operator splitting (10). This is the case when $A$ is a row selector of an orthogonal transformation.

- If $A$ is a square matrix diagonalizable in an orthonormal basis, i.e. $A = P^*DP$ where $P$ is orthogonal, then

$$(A^TA + \epsilon)^{-1}A^T = (P^*D^*DP + \epsilon)^{-1}P^*D^*P = P^*(\frac{D}{|D|^2 + \epsilon})P = A^T(AA^T + \epsilon)^{-1}.$$

Now we can consider a preconditioned constrained problem:

$$\min_u J_w(u) \quad \text{subject to } Bu = b, \tag{37}$$

By applying the general Bregmanized Operator splitting algorithm presented in Section 3, we get: $b^0 = Pf$

$$\begin{cases} v^{k+1} &= u^k - \delta A^T P^T(PAu^k - b^k) \\ u^{k+1} &= \arg\min_u \left( \mu J_w(u) + \frac{1}{2\delta}||u - v^{k+1}||^2 \right) \\ b^{k+1} &= b^k + b - PAu^{k+1} \end{cases} \tag{38}$$

This is equivalent to the algorithm:

**Algorithm II (Preconditioned Bregmanized Operator Splitting):**

$$\begin{cases} v^{k+1} &= u^k - \delta A^T(AA^T + \epsilon)^{-1}(Au^k - f^k) \\ u^{k+1} &= \arg\min_u \left( \mu J_w(u) + \frac{1}{2\delta}||u - v^{k+1}||^2 \right) \\ f^{k+1} &= f^k + f - Au^{k+1} \end{cases} \tag{39}$$

According to Theorem 1, the condition for the convergence of **Algorithm II** is $0 < \delta < \frac{1}{||B^TB||}$, 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 pseudo-inverse $A^+ = A^T (AA^T)^{-1}$ 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, therefore the matrix $A$ is a 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 $h$. In general, the matrix $A$ is not full row rank. As we mentioned above, the left and right pseudo inverse approximation are equal:

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

and the latter is equivalent to solve 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(FFT):

$$v^{k+1} = u^k - \delta \mathcal{F}^{-1} \left( \frac{H^*(\omega) \cdot (G^{k+1}(\omega) - H(\omega) \cdot U^{k+1}(\omega))}{|H(\omega)|^2 + \frac{1}{\delta}} \right), \qquad (40)$$

where $G^k(\omega), U^k(\omega)$ are discrete Fourier transform coefficients of $f^k$ and $u^k$ at frequency $\omega$. Consequently, implementing (40) 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 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 nonlocal total variation functional defined in (30) in the extended nonlocal ROF model [40]:

$$\min_u \left( \mu J_w(u) + \frac{1}{2} ||u - v||^2 \right), \qquad (41)$$

where $w$ is a fixed weight function and $\mu > 0$ for a given image $v$. Notice that the algorithms for the nonlocal ROF model are extended from the fast algorithms originally developed for solving classical TV based regularization problems. In particular, we are interesting in the extension of the split Bregman method proposed by Goldstein and Osher in [28].

The main idea of the split Bregman algorithm is to transform the total variation minimization problem onto a $l_1$ norm minimization by introducing an auxiliary variable for the gradient of

$u$, and then an efficient thresholding algorithm can be applied[28]. Here, we extend the split Bregman algorithm to the nonlocal TV regularization by considering the related problem:

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

The idea is to reformulate the problem as:

$$\min_{u,d} \left( \mu |d|_1 + \frac{1}{2} ||u - v||^2 \right) \quad \text{subject to} \quad d = \nabla_w u,$$

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

$$
\begin{aligned}
(u^{k+1}, d^{k+1}) &= \arg\min_{u,d} \left( \mu |d|_1 + \frac{1}{2} ||u - v||^2 + \frac{\lambda}{2} ||d - \nabla_w u - b^k||^2 \right) \quad (42) \\
b^{k+1} &= b^k + \nabla_w u^{k+1} - d^{k+1}.
\end{aligned}
$$

The solution of (42) is obtained by performing an alternating minimization process:

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

Note that the equivalence of the split Bregman method with classical Douglas-Rachford splitting method [18, 19] recently shown by Setzer in [42], thus the convergence is clarified.

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

$$(u^{k+1} - v) - \lambda \text{div}_w (\nabla_w u^{k+1} + b^k - d^k) = 0, \quad (43)$$

which provides

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

Since the graph Laplacian $\Delta_w$ is negative semi definite, the operator $1 - \Delta_w$ is diagonally dominant. Therefore we can solve $u^{k+1}$ by a Gauss-Seidel algorithm. Similarly to [28], the vector $d^{k+1}$ is obtained by applying the shrinkage operator (11):

$$d^{k+1} = \text{shrink}(\nabla_w u^{k+1} + b^k, \frac{\mu}{\lambda}).$$

## 4.3 Algorithms

To conclude this section, we describe the Split Bregman method for 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 $u^{k+1} = (1 - \Delta_w)^{-1}(v + \lambda \text{div}_w(b^k - d^k))$ by Gauss-Seidel method.
      Solve $d_{k+1} = \text{shrink}(\nabla_w u^{k+1} + b^k, \frac{\mu}{\lambda})$
      $b^{k+1} = b^k + \nabla_w u^{k+1} - d^{k+1}$
   **end for**
---

---
**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 nonlocal weight $w^{(k)} = w(v^{k+1}, h_0)$ // using the formula (26)
      **end if**
      Inner denoising step: Performing $nInner$ steps of the nonlocal TV denoising iteration with input $v^{k+1}, \mu\delta$.
      Update $f^{k+1} = f^k + f - Au^{k+1}$.
      Increase $k$.
   **end while**
---

## 5  Experimental Results

We present two applications: compressive sensing with Fourier measurements and image deconvolution. We compare the nonlocal H1 and TV regularization with the standard TV regularization and wavelet based $l^1$ regularization with the GPSR[1] algorithm [23].

In order to improve computational time and storage efficiency, we only compute the "best" neighbors, that is, for each pixel $x$, we only include the $K = 10$ best neighbors in the semi-local searching window of $21 \times 21$ centered at $x$ and the 4 nearest neighbors on comparing $5 \times 5$ patches with the formula (26). For the TV and the nonlocal 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 nonlocal TV regularization. Similarly, a Gauss-Seidel method is applied to solve the nonlocal $H^1$ regularization. A matlab and mex implementation of proposed algorithms is available online[3]. For all the experiments, the inner denoising steps for both NLTV and NLH1 are fixed as

---

[1]http://www.lx.it.pt/∼mtf/GPSR
[2]http://www.math.ucla.edu/∼tagoldst/public_codes/splitBregmanROF_mex.zip
[3]http://www.math.ucla.edu/∼xqzhang/html/code.html

$nInner = 20$ steps and the parameter $\delta = 1$.

## 5.1  Nonlocal TV Deconvolution

We implement both the BOS and PBOS methods on the Cameraman image for image deconvolution problem. As in [33], a fixed weight ($nUpdate = 0$) computed from a Tikhonov-based deblurred image $u^0$, see (35), is used for all the nonlocal methods. Using optimal $\lambda$ and the noise level $\sigma$, we can obtain $u^0$ very efficiently with an estimated noise level $\sigma_1$[33]. We set $h_0 = 2\sigma_1$. In [33], a gradient descent algorithm was applied to solve the unconstrained Lagrangian formulation:

$$\min \left( |\nabla_{w_0} u|_1 + \frac{\lambda}{2} ||Au - f||^2 \right). \tag{44}$$

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

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

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 [33], the reconstruction results are similar, while the computation speed is improved. Note that the weight function was computed in the whole searching window in [33], while here only 10 best and 4 nearest neighbors are used for each pixel. Furthermore, the algorithm BOS and PBOS takes less than $nOuter$ steps to meet the stopping criterion. Overall, the NLTV+BOS algorithm stops with 25 steps for 138 seconds, and the PBOS stopped at 8 steps for 51 seconds including weight computation, compared to 280 seconds with 500 steps with the gradient descent algorithm for solving (44).

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 an smoother image. One explanation is that the weight function computed from a pre-deblurred image is good enough to express structured information in the nonlocal TV regularization, while weigh updating degrades the image structures.

## 5.2  Compressive Sensing

In this section, we focus on exploring the sparsity of natural images with non-local regularization operators. The compressive sensing matrix we choose is $A = R\mathcal{F}$, where $R$ is a row-selector matrix, and $\mathcal{F}$ is 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 only consider 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 $h_0$ are empirically chosen as $h_0 = 20$ for the Barbara example (see Figure

---

[4]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.

2) and $h_0 = 15$ for the patch example respectively (Figure 3). For this application, an initial guess by setting unknown 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. As expected, the standard TV regularization is not capable of recovering texture patterns presented in these images. The results based on wavelet 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 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 iterations number. In fact, the PSNR of NLH1 is decreasing after a certain number of iterations. Empirically we choose $nOuter = 100$ for NLH1 as optimal results 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 Figure 3, respectively. Surprisingly, with only few measurements, the image textures are almost perfectly reconstructed by the nonlocal TV regularization. This is because image structures are expressed implicitly in the nonlocal weight function, and the nonlocal regularization process with Bregman iteration provides an efficient way to recover textures without explicitly construct a basis. Note that with a fewer outer iteration for the nonlocal TV, 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 nonlocal TV functional. Experiments show the nonlocal TV regularization is efficient to recover natural images with few measurements without using a basis or dictionary learning. We also have the similar observation as in [28], 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 in theory can be applied for other inverse problems and regularization. We will investigate more carefully this question in the future. Furthermore, as mentioned in [38], it is also important to better understand the weight updating strategy in a theoretical framework.

## References

[1] M. Aharon, M. Elad, and A. Bruckstein. $rmk$-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, Nov. 2006. 3

[2] T. Brox and D. Cremers. Iterated nonlocal means for texture restoration. In Springer-Link, editor, *Scale Space and Variational Methods in Computer Vision*, volume 4485/2007, pages 13–24. Springer Berlin / Heidelberg, 2007. 10

[3] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.*, 4(2), 2005. 9, 10

[4] A. Buades, B. Coll, and J. M. Morel. Image enhancement by non-local reverse heat equation. Technical Report 22, CMLA, 2006. 10

[5] M. Burger and S. Osher. A guide to the TV zoo. Springer, 2009, to appear. 2

[6] J.-F. Cai, S. Osher, and Z. Shen. Convergence of the linearized Bregman iteration for $l_1$-norm minimization. CAM Report 08-52, UCLA, 2008. 2, 5, 6

[7] J.-F. Cai, S. Osher, and Z. Shen. Linearized Bregman iterations for Frame-based Image Deblurring. CAM Report 08-57, UCLA, May 2008. 2, 14

[8] E. J. Candes and D. L. Donoho. Curvelets - a surprisingly effective nonadaptive representation for objects with edges, 2000. 3

[9] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006. 2

[10] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004. 5

[11] F. R. K. Chung. *Spectral graph theory*, pages 1–212. American Mathematical Society, 1997. 11

[12] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4), 2005. 1, 3, 4, 5, 13

[13] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harm. Anal.*, 2:101–126, 1995. 2

[14] J. Darbon and S. Osher. Fast discrete optimization for sparse approximations and deconvolutions. preprint, UCLA, 2007. 4, 5

[15] J. Darbon and M. Sigelle. Exact optimization of discrete constrained total variation minimization problems. volume 3322 of *LNCS*, pages 540–549. IWCIA 2004, 2004. 5

[16] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004. 2

[17] M. N. Do and M. Vetterli. The Contourlet Transform: An Efficient Directional Multiresolution Image Representation. *IEEE Transactions on Image Processing*, 14(12):2091–2106, 2005. 3

[18] J. Douglas and H.H. Rachford. On the numerical solutions of heat conduction problems in two and tree space variables. *Transactions of the American mathematical society*, Vol 1:97–116, 1976. 7, 16

[19] J. Eckstein and D. P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992. 4, 7, 16

[20] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proc. Int. Conf. Computer Vision*, volume 2, pages 1022–1038, 1999. 9

[21] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968, 2007. 2

[22] A. Elmoataz, O. Lezoray, and S. Bougleux. Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing, 2008. 11

[23] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):586–597, 2007. 2, 17

[24] D. Gabay. *Augmented Lagrangian methods: applications to the solution of boundary-value problems*, chapter Applications of the method of multipliers to variational inequalities. North-Holland, Amsterdam, 1983. 3, 4

[25] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. CAM Report 07-23, UCLA, July 2007. 9, 11

[26] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and operator splitting methods in nonlinear mechanics*. SIAM, 1989. 3, 4, 7

[27] D. Goldfarb and W. Yin. Parametric maximum flow algorithms for fast total variation minimization. Technical Report TR07-09, Rice University CAAM Technical Report, 2007. 5

[28] T. Goldstein and S. Osher. The split bregman method for $l1$ regularized problems. CAM Report 08-29, UCLA, 2008. 1, 4, 5, 7, 15, 16, 20

[29] E. T. Hale, W. Yin, and Y. Zhang. A fixed-point continuation method for $l1$-regularized minimization with applications to compressed sensing. CAAM Technical Report TR07-07, Rice University, 2007. 2, 3, 5

[30] D. S. Hochbaum. An efficient algorithm for image segmentation, markov random fields and related problems. *J. ACM*, 48(4):686–701, 2001. 4, 5

[31] S. Kindermann, S. Osher, and P. W. Jones. Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling and Simulation*, 4(4):1091–1115, 2005. 11

[32] C. Lemarechal and C. Sagastizbal. Practical aspects of the moreau-yosida regularization: Theoretical preliminaries. *SIAM Journal on Optimization*, 7:367–385, 1997. 7

[33] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi. Image recovery via nonlocal operators. CAM Report 08-35, UCLA, May 2008. 13, 18, 19

[34] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci. Paris Ser. A. Math.*, 255:2897–2899, 1962. 5

[35] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *SIAM Multiscale Model. and Simu*, 4:460–489, 2005. 1, 3, 4

[36] S. Osher, Y. Mao, B. Dong, and W. Yin. Fast linearized bregman iteration for compressive sensing and sparse denoising. CAM Report 08-37, UCLA, June 2008. 2

[37] G. Peyré. Image processing with nonlocal spectral bases. *Multiscale Model. Simul.*, 7:703–730, 2008. 11

[38] G. Peyré, S. Bougleux, and L. Cohen. Non-local regularization of inverse problems. In *Proceedings of Proc. ECCV 2008*, 2008. 12, 13, 20

[39] R.T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of operations research*, 82:421–439, 1956. 7

[40] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. 2, 3, 4, 15

[41] M. Lustig S. Boyd S.-J. Kim, K. Koh and D. Gorinevsky. A method for large-scale l1-regularized least squares. *IEEE Journal on Selected Topics in Signal Processing*, 1(4):606–617, 2007. 2

[42] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. Technical report, 2008. 7, 16

[43] P. Tseng. Applications of splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim.*, 29(1):119–138, 1991. 4

[44] W. Yin. Analysis and generalizations of the linearized bregman method. CAAM Report TR09-02, Rice University, 2008. 2, 6

[45] 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.*, 1:143–168, 2008. 1, 3, 4, 5

[46] D. Zhou and B. Schölkopf. Regularization on discrete spaces. *Pattern Recognition*, 361:361–368, 2005. 11

[47] W. Zulehner. Analysis of iterative methods for saddle point problems: A unified approach. *Mathematics of Computation*, 71(238):479–505, 2001. 3, 7

Figure 2: Compressive sensing example: Barbara ($256 \times 256$), 30% randomly chosen Fourier coefficients, noiseless. Weight updating.
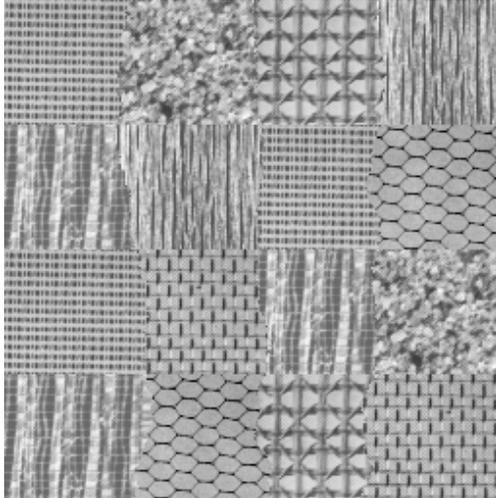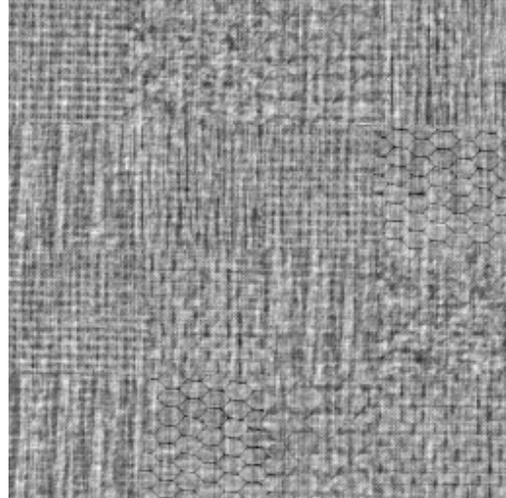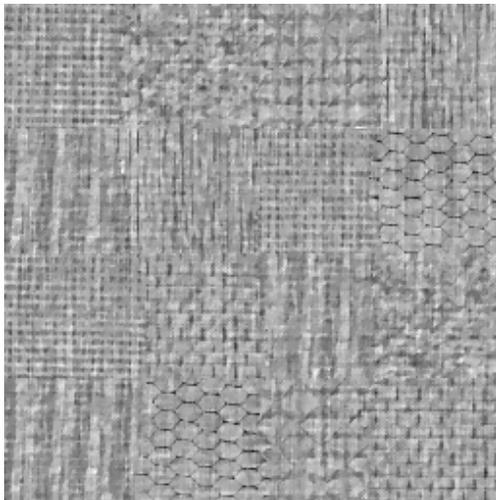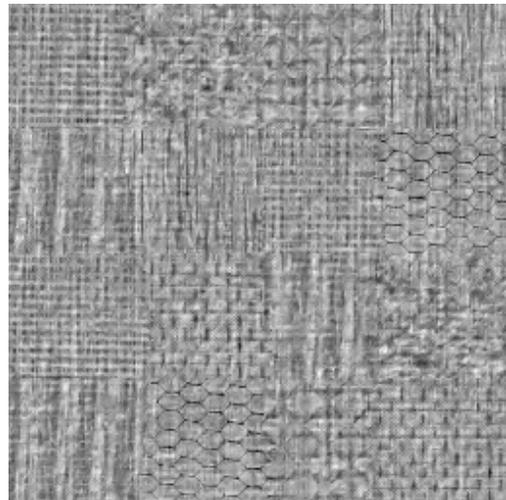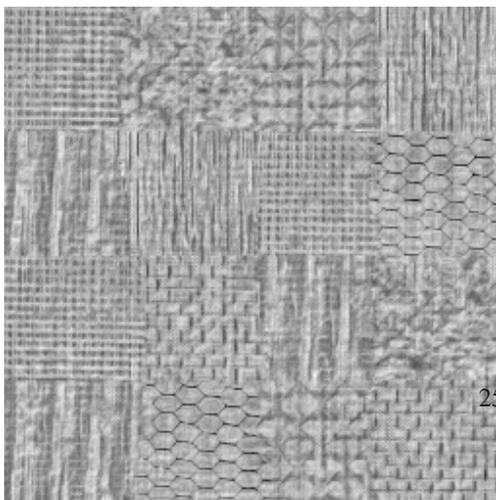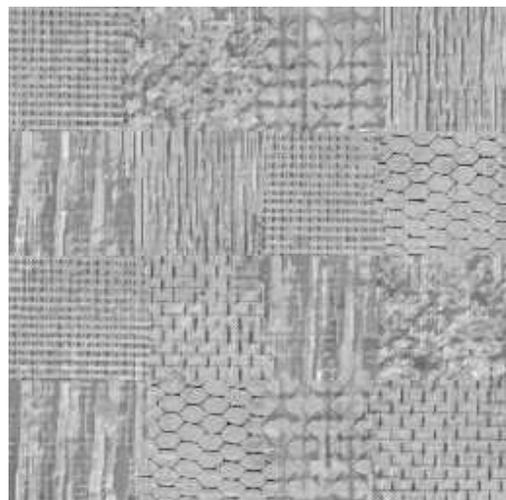
Figure 3: Compressive sensing example: Textures ($256 \times 256$), 30% randomly chosen Fourier coefficients, noiseless. Weight updating.