Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction*

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

Abstract

We propose two algorithms based on Bregman iteration and operator splitting technique for nonlocal TV regularization problems. The convergence of the algorithms is analyzed and applications to deconvolution and sparse reconstruction are presented.

1 Introduction

Image restoration can be formulated as an inverse problem. 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 ϵ is a white Gaussian noise with variance σ^2, 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 [25]. This model uses the TV regularization

---

[*] Revised on January 9, 2009

[†] M. Burger is with the Institute for Computational and Applied Mathematics, Westfälische Wilhelms-Universität, Einsteinstr. 62, D48163 Münster, Germany. Email:martin.burger@wwu.de

[‡] X. Zhang, X. Bresson, and S. Osher are with the Department of Mathematics, UCLA , Box 951555, Los Angeles, CA 90095-1555, USA. Email: {xqzhang,xbresson,sjo}@math.ucla.edu

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 [4]:

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

where ∇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 wavelets or frames, has become popular recently. Given a frame $(\phi_m)_m$ of $\mathbb{R}^n$ (m is not necessarily equal to n), the regularized inverse problem is defined by the l^1 minimization problem:

$$\arg\min_u\Big(\mu\sum_m|<u,\phi_m>|_1+\frac{1}{2}||Au-f||^2\Big), \tag{1}$$

where $< u,\phi_m >$ denotes the inner product of the m-th basis function and the function u. The model (1) was used for the deconvolution problem using the wavelet-vaguelette decomposition model defined in [11]. Usually, a scale-dependent shrinkage is employed to estimate the image wavelet coefficients. 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.

The minimization problem (1) is widely used in compressive sensing problems [7]. 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 (with probability very close to one). 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.

In this paper, we propose two efficient algorithms to solve nonlocal TV regularization problems such as deconvolution or compressive sensing. Our algorithms combine the Bregman iteration and operator splitting techniques to solve nonlocal TV regularization problems. These two algorithms can also be interpreted as inexact Uzawa methods, which were traditionally used for saddle point problems [29]. Our experiments show that the proposed Bregmanized nonlocal regularization algorithms can recover almost all the details of a textured image without explicitly choosing a basis. Hence, nonlocal TV itself sparsifies and the Bregman iteration is efficient for sparse recovery.

The paper is organized as follows. We begin with a brief review of related sparse optimization techniques: Bregman iteration, operator splitting and linearized Bregman. We then present the proposed nonlocal sparse regularization framework and the corresponding numerical schemes. We define two algorithms that we call Bregmanized operator splitting

(BOS) and preconditioned Bregmanized operator splitting (PBOS). An updating strategy for the weight function in the nonlocal TV term is also presented. The convergence of the algorithms with fixed weight is discussed. Finally, we apply the two algorithms to the deconvolution problem and the compressive sensing reconstruction problem.

2 Bregman Iteration and Operator Splitting

2.1 Solving Equality Constrained Minimization Problems with Bregman Iteration

The following constrained minimization problem is considered:

$$\min_u J(u) \;\text{ s.t. }\; H(u) = 0, \tag{2}$$

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 in [22] to improve the ROF denoising models [25].

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:

$$D_J^p(u, v) = J(u) - J(v) - < p,\; u - v >, \tag{3}$$

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 (3), the original constrained minimization problem (2) 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}. In the particular case we want to solve the constrained minimization problem:

$$\min_u J(u) \;\text{ s.t. }\; Au = f. \tag{4}$$

Choosing $H(u) = \frac{1}{2}||Au - f||^2$, the previous minimization problem is solved by a two-step Bregman iterative scheme [22]:

$$\begin{cases} u^{k+1} &= \min_u \left(\mu J(u) + \frac{1}{2}||Au - f^k||^2\right) \\ f^{k+1} &= f^k + f - Au^{k+1} \end{cases} \tag{5}$$

It is also shown in [22] that if there is a function u^* such that $Au^* = f$, then u^* is a solution to the original constrained minimization problem (4). Moreover, the residual $||Au^k - f||$ of the sequence generated by (5) 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 [27, 22, 12, 17].

2.2 Solving Unconstrained Minimization Problems with Operator Splitting

The following unconstrained minimization problem is considered:

$$\min_u \Big(\mu J(u) + H(u)\Big), \tag{6}$$

where $\mu > 0$, J and H are convex functionals defined over $\mathbb{R}^n \to \mathbb{R}^+$. The authors in [10] developed a forward-backward technique to minimize the sum of two convex functionals based on the proximal operator introduced by Moreau in [21]. A proximal operator of a convex functional J of a function v is defined as :

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

By classical arguments of convex analysis, the solution of (6) satisfies the condition:

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

For any positive number δ, we have:

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

This leads to a forward and backward splitting algorithm:

$$u^{k+1} = \text{Prox}_{\delta\mu J}(u^k - \delta\partial H(u^k)), \tag{7}$$

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

$$\arg\min_u \Big(\mu J(u) + \frac{1}{2\delta}||u - v||^2\Big). \tag{8}$$

The following convergence theorem is a consequence of the main result of [10]:

Theorem 1. *Assume that:*

- *J and H are proper lower semi-continuous convex and coercive, i.e.*

$$\lim_{||x|| \to \infty} \mu J(x) + H(x) = \infty.$$

- *H is differentiable with β-Lipschitz continuous gradient for some $\beta \in (0, \infty)$, that is*

$$||\partial H(u) - \partial H(v)|| \leq \beta||u - v||.$$

Then, the problem (6) possess at least one solution, and for any initial guess u^0, the algorithm (7) converges to a solution of (6) when $\delta \leq \frac{2}{\beta}$.

In the case of $H(u) = \frac{1}{2}||Au - f||^2$, we can see that $\partial H(u) = A^T(Au - f)$ is β-Lipschitz, where $\beta = \lambda_{\max}(A^TA)$ is the maximum eigenvalue of A^TA. Therefore the solution of the minimization problem (6) can be computed by the following two-step algorithm:

$$\begin{cases} v^{k+1} & = & u^k - \delta A^T(Au^k - f) \\ u^{k+1} & = & \arg\min_u \left(\mu J(u) + \frac{1}{2\delta}||u - v^{k+1}||^2\right) \end{cases} \tag{9}$$

The proximal operator solution (8) 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 [10, 18, 5] as follows:

$$u = \text{shrink}(v, \mu\delta) = \text{sign}(v)\max\{|v| - \mu\delta, 0\}. \tag{10}$$

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 the Chambolle's projection method [8], the split Bregman method [17] or by graph cuts in the anisotropic case[12].

2.3 Linearized Bregman

The idea of the linearized Bregman iteration [12, 27] is to combine Bregman iteration and operator splitting to solve the constrained problem (4) for l^1 sparse reconstruction. Given $u^0 = 0 = p^0$, the iterative algorithm of linearized Bregman is defined for $k \geq 0$ as:

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

where $D_J^p(u, v)$ is the Bregman distance defined in (3). This minimization can be rewritten in a simpler formulation as follows:

$$\begin{cases} v^{k+1} & = & v^k - \delta A^T(Au^k - f) \\ u^{k+1} & = & \min \left(\mu J(u) + \dfrac{1}{2\delta}||u - v^{k+1}||^2\right) \end{cases} \tag{11}$$

The difference between linearized Bregman and the operator splitting method (9) is the way of updating v^{k+1}. Cai *et al.* proved the following propositions in [5]:

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

$$\min \left(\mu J(u) + \frac{1}{2\delta}||u||^2\right) \quad \textit{s.t.} \;\; Au = f. \tag{12}$$

The convergence of (11) requires further assumptions on the regularization functional $J(u)$:

Proposition 2. *Assume that A is subjective, the convex function $J(u)$ is continuously differentiable and there exists a positive constant β such that*

$$||\partial J(u) - \partial J(v)||^2 \leq \beta < \partial J(u) - \partial J(v), u - v > \quad \forall (u, v),$$

where $\partial J(u)$ denotes a subgradient of $J(u)$, then the sequences $\{u^k\}$ and $\{p^k\}$ generated by (11) converges. And the limit of u^k is the unique solution of (12).

As $\mu \to \infty$, the solution of (12) tends to the solution of (4).

In the case of l^1 sparse approximation, algorithm (11) can be written as follows:

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

Clearly, the l^1 norm is not differentiable, hence the convergence can not be guaranteed. It was shown in [5] that the linearized Bregman is equivalent to one step operator splitting in the inner iteration of a Bregmanized iteration. The following convergence result was also shown in [5]:

Proposition 3. *If* $0 < \delta < \frac{1}{||AA^T||}$, *then the sequence* u^k *converges to the solution*

$$u = \arg\min \left(\mu|u|_1 + \frac{1}{2\delta}||u||^2\right) \textit{ s.t. } \; Au = f.$$

Moreover, it was proved in [26] that, for μ large enough, the limit solution solves the original problem:

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

3 Nonlocal Methods

3.1 Nonlocal Means

In [13], 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 [2] an efficient denoising model called nonlocal means (NL-means). The model consists in denoising a pixel value by averaging the nearby pixel values with similar structures (patches).

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

$$\begin{aligned} w(f,h_0)(x,y) &= \exp\{-\frac{G_a * (||f(x+\cdot) - f(y+\cdot)||^2)(0)}{2h_0^2}\}, \\ C(x) &= \int_\Omega \exp\{-\frac{G_a * (||f(x+\cdot) - f(y+\cdot)||^2)(0)}{2h_0^2}\} dy. \end{aligned} \tag{13}$$

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 [2]. The definition of the weight function (13) 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, the operator $NLM_v(u)$ can be written as matrix multiplications such as

$$NLM_f(u) = D(f, h_0)^{-1} W(f, h_0) U$$

where $W(f, h_0)$ is the $N \times N$ weight matrix defined in (13), $D(f, h_0)(i, i) = C(i)$ is a $N \times N$ diagonal matrix and U is the image vector.

3.2 Nonlocal Variational Framework

3.2.1 NL-means Based Regularization

The application of the nonlocal means filter to image deblurring is not trivial since the observed image and the original image generally do not have the same similarities 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 [3] a NL-means regularization energy for image deblurring defined as follows:

$$J_{NLM}(u) := ||u - NLM_f(u)||^2 \tag{14}$$

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. The corresponding optimality condition for (14) is:

$$(Id - D_f^{-1} W_f)^T (Id - D_f^{-1} W_f) U = 0.$$

An alternative nonlocal model for texture restoration is introduced in [1]. The authors propose to minimize the functional:

$$E_{NLM}(u) := ||u - NLM_u(f)||^2. \tag{15}$$

It is a nonlinear model since the weight function depends on the unknown image u. The solution of (15) 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.

3.2.2 Nonlocal Total Variation Functional

In [19], Kindermann *et al.* investigated the use of regularization functionals with nonlocal correlation terms for the problems of image denoising and image deblurring. Inspired from the graph Laplacian in [9], Gilboa and Osher defined a variational framework based nonlocal operators in [16]. Note that Zhou and Schölkopf in [28] and Elmoataz *et al.* in [14] also used graph Laplacian in the discrete setting for image denoising. In [23], Peyré considered spectral bases of the graph Laplacian operator, which bridges the gap between the adaptive filtering methods and thresholding in orthogonal bases.

Let us now define the nonlocal TV functional. 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 (13) from a reference image, then the nonlocal gradient $\nabla_w u(x)$ is defined as the vector of all partial derivatives $\nabla_w u(x, \cdot)$ 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:

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

which leads to the definition of the graph divergence 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:

$$< \Delta_w u, u >=< u, \Delta_w u >,$$

and negative semi-definite:

$$< \Delta_w u, u >= - < \nabla_w u, \nabla_w u > \leq 0.$$

The nonlocal TV norm is the isotropic L^1 norm of the weighted graph gradient $\nabla_w u(x)$ defined as:

$$\begin{aligned} J_w(u) &:= \int_\Omega |\nabla_w u(x)| dx \qquad (16) \\ &= \int_\Omega \sqrt{\int_\Omega (u(x) - u(y))^2 w(x,y) dy} dx. \end{aligned}$$

The corresponding Euler-Lagrange equation of (16) is written as:

$$- \int_\Omega (u(y) - u(x))w(x,y) \left[\frac{1}{|\nabla_w u(x)|} + \frac{1}{|\nabla_w u(y)|} \right] dy = 0. \qquad (17)$$

Finally, if the function $w(x,y)$ in (17) is chosen to be the nonlocal weight function defined in (13), then the nonlocal means filter is generalized to a consistent variational framework.

3.3 Numerical Schemes for Nonlocal TV Denoising

In this section, we focus on fast algorithms to minimize the nonlocal total variation functional defined in (16) in the extended nonlocal ROF model [25]:

$$\min_u \Big(\mu J_w(u) + \frac{1}{2}||u - v||^2\Big), \tag{18}$$

where w is a fixed weight function and $\mu > 0$.

3.3.1 Chambolle's Projection Algorithm

Gilboa-Osher solved (18) in [16] by extending Chambolle's projection algorithm [8]. The problem (18) is equivalent to the min-max problem :

$$\min_u \max_{||p||\leq 1} \Big(\mu \int_\Omega < \nabla_w u, p > + \frac{1}{2}||u - v||^2\Big),$$

where $||p||$ is the usual l^2 norm of vectors. The solution u can be solved by projection:

$$u = v - \mu \text{div}_w p,$$

and the dual variable p is obtained by solving:

$$\max_{||p||\leq 1} \Big(\int_\Omega < \nabla_w u, p > + \frac{\mu}{2}||\text{div}_w p||^2\Big),$$

which is realized with the following semi-implicit fixed point iterated method:

$$p^{n+1} = \frac{p^n + \tau \nabla_w(\text{div}_w p^n - \frac{1}{\mu}v)}{1 + \tau|\nabla_w(\text{div}_w p^n - \frac{1}{\mu}v)|},$$

where $0 < \tau \leq \frac{1}{||\text{div}_w||^2}$ guarantees the convergence of the iteration scheme.

3.3.2 Split Bregman Algorithm

Another fast algorithm can be defined by extending the split Bregman algorithm originally introduced by Goldstein and Osher in [17] for solving classical TV based regularization problems :

$$\min_u \Big(\mu|\nabla u|_1 + \frac{1}{2}||u - v||^2\Big),$$

The idea is to reformulate the problem as:

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

and force the constraint with the Bregman iteration process as follows:

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

The extended nonlocal split Bregman algorithm uses the nonlocal TV norm instead of the standard TV norm:

$$\begin{aligned} (u^{k+1}, d^{k+1}) &= \arg\min_{u,d} \left(\mu|d|_1 + \frac{1}{2}||u-v||^2 + \frac{\lambda}{2}||d - \nabla_w u - b^k||^2 \right) \\ b^{k+1} &= b^k + \nabla_w u^{k+1} - d^{k+1}. \end{aligned} \tag{19}$$

The solution of (19) is obtained by performing an alternative minimization process:

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

The Euler-Lagrange equation for u^{k+1} is given by:

$$(u^{k+1} - v) - \lambda \mathrm{div}_w(\nabla_w u^{k+1} + b^k - d^k) = 0, \tag{20}$$

which provides

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

Since the graph Laplacian Δ_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 [17], the vector d^{k+1} is obtained by applying the shrinkage operator (10):

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

4 Nonlocal Regularization for Inverse Problems via Bregman Iteration and Operator Splitting

4.1 Nonlocal regularization with weight fixed

In the following we consider the constrained minimization formulation for inverse problems:

$$\min_u J_w(u) \quad \text{subject to } Au = f, \tag{21}$$

with J_w being nonlocal regularization term (s.a. nonlocal TV) with fixed weight function w.

4.1.1 Bregmanized Operator Splitting (BOS)

Our goal is to solve the equality constrained minimization problem (21) by the Bregman iteration and operator splitting introduced in Sections 2.1 and 2.2. First of all, the equality constraint in (21) is enforced with the Bregman iteration process:

$$\left\{ \begin{array}{lcl} u^{k+1} &=& \min_u \left(\mu J_w(u) + \frac{1}{2}||Au - f^k||_2^2 \right) \\ f^{k+1} &=& f^k + f - Au^{k+1} \end{array} \right. . \tag{22}$$

Then, the operator splitting technique is used to solve the unconstrained subproblem in (22) as follows: for $i \geq 0, u^{k+1,0} = u^k$,

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

for a positive number $0 < \delta < \dfrac{2}{\|A^TA\|}$. Ideally we need to run infinite inner iterations to obtain u^{k+1}. Here we propose to use only one inner iteration, which leads to the algorithm I:

Algorithm I (Bregmanized Operator Splitting):

$$\begin{cases} v^{k+1} & = & u^k - \delta A^T(Au^k - f^k) \\ u^{k+1} & = & \arg\min_u \left(\mu J_w(u) + \frac{1}{2\delta}\|u - v^{k+1}\|^2\right). \\ f^{k+1} & = & f^k + f - Au^{k+1} \end{cases} \tag{23}$$

which is equivalent to:

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

The solution u^{k+1} is given by regularization of v^{k+1}. In the case of nonlocal TV regularization, a fast regularization technique was introduced in Section 3.3.2.

4.1.2 Preconditioned Bregmanized Operator Splitting (PBOS)

The auxiliary variable v introduced in Algorithm I can be interpreted in another manner. If we consider a Moreau-Yosida regularization of (4) as follows:

$$\min_{u,v} \left(\mu J_w(u) + \frac{1}{2}\|Av - f\|^2 + \frac{1}{2\delta}\|u - v\|^2\right) \text{ s.t. } Au = f. \tag{24}$$

It is direct to see that if $\overline{u}$ is a solution of (4), then $(\overline{u}, \overline{u})$ is a solution pair for (24). This problem is solved using Bregman iteration scheme:

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

As above, we propose to use a one-step alternative algorithm:

Algorithm II (Preconditioned Bregmanized Operator Splitting):

$$\begin{cases} v^{k+1} & = & \min_v \left(\frac{1}{2\delta}\|u^k - v\|^2 + \frac{1}{2}\|Av - f^k\|^2\right) \\ u^{k+1} & = & \arg\min_u \left(\mu J_w(u) + \frac{1}{2\delta}\|u - v^{k+1}\|^2\right). \\ f^{k+1} & = & f^k + f - Au^{k+1} \end{cases} \tag{25}$$

We can also write this algorithm in the form:

$$\left\{\begin{array}{lll} v^{k+1} & = & u^k-(A^TA+\frac{1}{\delta})^{-1}A^T(Au^k-f^k) \\ u^{k+1} & = & \arg\min_u\left(\mu J_w(u)+\frac{1}{2\delta}||u-v^{k+1}||^2\right) \\ f^{k+1} & = & f^k+f-Au^{k+1} \end{array}\right. . \tag{26}$$

Using the optimality condition for the second line we reduce to:

$$\left\{\begin{array}{lll} \mu\delta s^{k+1}+u^{k+1} & = & u^k-(A^TA+\frac{1}{\delta})^{-1}A^T(Au^k-f^k) \\ f^{k+1} & = & f^k+f-Au^{k+1} \end{array}\right. , \tag{27}$$

where $s^{k+1}\in\partial J(u^{k+1})$.

4.1.3 Pseudo-Inverse Interpretation

As we have mentioned, an important question in nonlocal regularization methods for inverse problems is how to estimate a correct weight function w. In [20], we estimate the weight function with the solution of the Tikhonov regularization problem:

$$v=\arg\min_v\Big(\frac{1}{2}||Av-f||^2+\frac{1}{2\delta}||v||^2\Big),$$

where δ is a large positive number. It amounts to

$$v=(A^TA+\frac{1}{\delta})^{-1}A^Tf.$$

The operator is a preconditioned generalized inverse of A when A is not invertible or ill-conditioned. We have:

$$\lim_{\delta\to\infty}(A^TA+\frac{1}{\delta})^{-1}A^T=\lim_{\delta\to\infty}A^T(AA^T+\frac{1}{\delta})^{-1}=A^+,$$

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

In [20], 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.

Now, we consider a modified operator splitting algorithm analogous to (9):

$$\left\{\begin{array}{lll} v^{k+1} & = & u^k-\delta A^+(Au^k-f) \\ u^{k+1} & = & \arg\min_u\left(\mu J_w(u)+\frac{1}{2\delta}||u-v^{k+1}||^2\right) \end{array}\right. , \tag{28}$$

where A^+ is the pseudo-inverse of A and $\delta>0$. This similar idea is also considered in [6] 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. If we replace A^+ by $A^T(AA^T+\frac{1}{\delta})^{-1}$, then algorithm (28) solves the minimization problem

$$\min_u \Big(J_w(u)+\lambda||Bu-b||^2\Big),$$

where B and b are constructed as:

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

In particular,

- If A is full row rank ($A^+=A^T(AA^T)^{-1}$), then we set $\delta=\infty$ 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 (28) is consistent with the classical operator splitting (9). 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

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

4.1.4 Solving v^{k+1} in (26)

In the algorithm II (PBOS), we need to invert the operator $(A^TA+\frac{1}{\delta})$ to compute

$$v^{k+1}=u^k-(A^TA+\frac{1}{\delta})^{-1}(Au^k-f^k).$$

In the following, we illustrate how to obtain v^{k+1} for two 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 $A^TA=R^T\mathcal{F}^*R\mathcal{F}=R^TR$ when Id_m is $m\times m$ identity matrix (R^T is zero-padding operator). In this case, the pseudo-inverse $A^+=A^T(AA^T)^{-1}$ is equal to A^T. Thus when $\delta=\infty$, the algorithm is equivalent to Algorithm I. When $\delta\neq\infty$, the iterate v^{k+1} is given as :

  $$v^{k+1}=u^k-\delta\mathcal{F}^{-1}(R^TR+\frac{1}{\delta})^{-1}(Au^k-f^{k+1}).$$

  Note that $R^TR+\frac{1}{\delta}$ is a diagonal matrix.

- **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. The solution v^{k+1} can be computed via the fast Fourier transform(FFT):

  $$v^{k+1} = u^k - \delta\mathcal{F}^{-1}\left(\frac{H^*(\omega)\cdot(G^{k+1}(\omega) - H(\omega)\cdot U^{k+1}(\omega))}{|H(\omega)|^2 + \frac{1}{\delta}}\right), \tag{29}$$

  where $G^k(\omega)$, $U^k(\omega)$ are discrete Fourier transform coefficients of f^k and u^k at frequency ω. Consequently, implementing (29) requires only $O(N^2 \log N)$ operations for an $N \times N$ image.

  Note that this is also equivalent to a Tikhonov regularization:

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

  because

  $$(AA^T + \frac{1}{\delta})^{-1}A^T = A^T(AA^T + \frac{1}{\delta})^{-1}.$$

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 Nonlocal Regularization with 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 (13) 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 $w(u^k, h_0)$ (13) during the reconstruction of signals. In [24], the authors have also proposed to update the graph weight to solve inverse problems using the forward-backward operator splitting technique [10].

As a result, we consider a more appropriate problem:

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

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

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

Similarly, the algorithm based on the algorithm II (PBOS) with weight updating can be written as

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

In practice, we actually do not need to update the weight function at every step. Instead, we update every M steps, see Section (6).

5 Convergence Analysis

In this section, we prove the convergence of the two proposed algorithms. Note that the proof can be generalized for different regularization functionals, such as l^1 sparse reconstruction, which we will be discussed in another paper. In the following, we define

$$J(u) := |\nabla_w u|_1,$$

with fixed weight w.

Lemma 1. *Let $s \in \partial J(u), s' \in \partial J(u')$, then*

$$< s - s', u - u' > \geq 0.$$

Proof: Since $s \in J(u), s' \in \partial J(u')$, we consider the Bregman distances between u and u' :

$$D_J^{s'}(u, u') = J(u) - J(u') - < s',\ u - u' > .$$

and

$$D_J^{s}(u', u) = J(u') - J(u) - < s,\ u' - u > .$$

Since Bregman distances are nonnegative, we find by adding the above two formulas

$$< s - s', u - u' > \geq 0.$$

Note that this statement can be generalized to any convex functional.

Theorem 2. *If $0 < \delta < \frac{1}{\|A^T A\|}$, let the sequence (u^k, s^k, p^k) be generated by Algorithm I given in (23). Then the sequence converges along subsequences to solutions of (4).*

Proof: Consider the Lagrangian formulation of the original constrained problem (4):

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

Then the optimality condition is as follows:

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

where $\overline{u}$ and $\overline{p}$ are the optimal solutions and $\overline{s}$ is a subgradient of J at $\overline{u}$.

Let (u^k, s^k, f^k) be a sequence generated by the algorithm (I) or (II), and rewrite the sequences in terms of error as follows:

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

By the optimality conditions given in **Algorithm 2**, the sequence (u^k, s^k, f^k) satisfies:

$$\left\{ \begin{array}{rcl} \mu s^{k+1} + \frac{1}{\delta}u^{k+1} &=& (\frac{1}{\delta} - A^TA)u^k + A^Tf^k \\ f^{k+1} &=& f^k + f - Au^{k+1} \end{array} \right. .$$

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

$$\left\{ \begin{array}{rcl} \mu s^{k+1} + Lu^{k+1} - A^Tf^{k+1} &=& Lu^k - A^Tf \\ f^{k+1} &=& f^k + f - Au^{k+1} \end{array} \right. .$$

In terms of the error differences, we have:

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

Using $||v||_L^2 :=< Lv, v >$, we evaluate:

$$\begin{aligned} & \|\Delta u^{k+1}\|_L^2 + \|\Delta f^{k+1}\|^2 + \|u^{k+1} - u^k\|_L^2 + \|f^{k+1} - f^k\|^2 - \|\Delta u^k\|_L^2 - \|\Delta f^k\|^2 \\ = \; & 2 < u^{k+1} - u^k, \Delta u^{k+1} >_L + 2 < f^{k+1} - f^k, \Delta f^{k+1} > \\ = \; & 2 < A^T\Delta f^{k+1}, \Delta u^{k+1} > -2\mu < \Delta s^{k+1}, \Delta u^{k+1} > + 2 < f - Au^{k+1}, \Delta f^{k+1} > \\ = \; & -2\mu < \Delta s^{k+1}, \Delta u^{k+1} > . \end{aligned}$$

By Lemma 1, we have:

$$\|\Delta u^{k+1}\|_L^2 + \|\Delta f^{k+1}\|^2 + \|u^{k+1} - u^k\|_L^2 + \|f^{k+1} - f^k\|^2 \le \|\Delta u^k\|_L^2 + \|\Delta f^k\|^2.$$

By induction, we get:

$$\|\Delta u^{k+1}\|_L^2 + \|\Delta f^{k+1}\|^2 \le \|\Delta u^0\|_L^2 + \|\Delta f^0\|^2,$$

Summation of the inequality implies:

$$\sum_{k=0}^{\infty} \|u^{k+1} - u^k\|_L^2 + \|f^{k+1} - f^k\|^2 \le \|\Delta u^0\|_L^2 + \|\Delta f^0\|^2.$$

Together with the uniform boundness of s^k, we conclude that there exists a convergent subsequence of (u^k, s^k, f^k). Furthermore, if $(u^{k_l}, s^{k_l}, f^{k_l})$ is such a convergent subsequence, then

$$\left\{ \begin{array}{rcl} \mu s^{k+1} + A^T(f - f^{k+1}) & = & L(u^k - u^{k+1}) \to 0 \\ f - Au^{k+1} & = & f^{k+1} - f^k \to 0. \end{array} \right.$$

Thus the limit is a solution of (4). □

Theorem 3. *If* $0 < \delta < \frac{1}{\|A^T A\|}$*, let the sequence* (u^k, s^k, p^k) *be generated by the algorithm II given in (25). Then the sequence converges along subsequences to solutions of (4).*

Proof: The proof is analogous to that of the above theorem by considering an equivalent problem:

$$\min_{\tilde{u}} J(P\tilde{u}), \qquad \text{s.t.} \quad B\tilde{u} = f, \tag{30}$$

where $P = \delta^{-\frac{1}{2}}(A^T A + \frac{1}{\delta})^{-\frac{1}{2}}$ and $B = AP$. Since P is invertible and symmetric, then when $\overline{\tilde{u}}$ is a solution of (30), then $\overline{u} = P\overline{\tilde{u}}$ is a solution of (4). The Lagrangian form of the (30) is:

$$L(\tilde{u}, \tilde{p}) = \mu J(P\tilde{u}) - < B\tilde{u} - f, \tilde{p} - f > \qquad \text{and} \qquad B\tilde{u} = f.$$

Then the optimality condition is:

$$\begin{array}{rcl} \mu P^T s(P\overline{\tilde{u}}) + B^T(f - \overline{\tilde{p}}) & = & 0 \\ B\overline{\tilde{u}} - f & = & 0. \end{array}$$

where $\overline{\tilde{u}}$ and $\overline{\tilde{p}}$ are the optimal solutions of (30) and $s(P\overline{\tilde{u}})$ is a subgradient of J at $P\overline{\tilde{u}}$. Let recall that the sequence (u^k, s^k, f^k) generated by the Algorithm II satisfies:

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

where $s^{k+1} \in \partial J(u^{k+1})$. Let $\tilde{u}^k = P^{-1}u^k$, $\tilde{s}^k = Ps^k$, then $\tilde{s}^k = P^T s^k \in \partial J(P\tilde{u}^k)$. Therefore the first equality of the above optimality condition is equivalent to

$$\begin{array}{rl} & \mu\delta P^{-T}\tilde{s}^{k+1} + P\tilde{u}^{k+1} = P\tilde{u}^k - \delta P^2 A^T(AP\tilde{u}^k - f^k) \\ \Longrightarrow & \mu\delta P^{-T}\tilde{s}^{k+1} + P\tilde{u}^{k+1} = P(\tilde{u}^k - \delta B^T(B\tilde{u}^k - f^k)) \\ \Longrightarrow & \mu\delta\tilde{s}^{k+1} + P^T P\tilde{u}^{k+1} = P^T P(\tilde{u}^k - \delta B^T(B\tilde{u}^k - f^k)) \end{array}$$

Then we have the new sequences $(\tilde{u}^{k+1}, \tilde{s}^{k+1}, f^{k+1})$ such that

$$\left\{ \begin{array}{rcl} \mu\tilde{s}^{k+1} + \frac{1}{\delta}P^T P\tilde{u}^{k+1} & = & \frac{1}{\delta}P^T P(1 - \delta B^T B)\tilde{u}^k + P^T P B^T f^k \\ f^{k+1} & = & f^k + f - B\tilde{u}^{k+1} \end{array} \right.$$

Let

$$L = \frac{1}{\delta}P^T P(1 - \delta B^T B) = \frac{1}{\delta}P^T(1 - \delta B^T B)P = \frac{1}{\delta^2}(A^T A + \frac{1}{\delta})^{-1}(1 - \delta B^T B)$$

thus L is symmetric.

By applying the same procedure as in Theorem 2, we can prove that a subsequence of $(\tilde{u}^k, \tilde{s}^k, f^k)$ converges to a solution of (30), therefore a subsequence of $u^k = P\tilde{u}^k$ converges to a solution of (4). □

5.1 Inexact Uzawa Methods Interpretation

The above two algorithms can be interpreted as an inexact Uzawa method [29] applied to the augmented Lagrangian of the original problem as follows:

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

where $J(u)$ is general convex functional. If we apply inexact Uzawa method on this formulation, we get the following algorithm:

$$\left\{ \begin{array}{lrcl} \text{step 1:} & u^{k+1} & = & \min_u \left(\mu J(u) + \frac{1}{2}||Au - f||^2 - < Au - f, f^k > + \frac{1}{2}||u - u^k||_D^2 \right) \\ \text{step 1:} & f^{k+1} & = & f^k - (Au^{k+1} - f) \end{array} \right. \tag{31}$$

where L is a positive-definite preconditioner. The sequence (u^k, s^k, f^k) generated by (31) provides us:

$$\left\{ \begin{array}{lcl} \mu s^{k+1} + (D + A^T A)u^{k+1} & = & Du^k + A^T A f - A^T f^k \\ f^{k+1} & = & f^k - (Au^k - f) \end{array} \right.$$

When $D = \frac{1}{\delta} - A^T A$, we get back the Algorithm I (BOS) defined in (23).

6 Algorithms

We describe the algorithms introduced in this paper.

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
```

7 Experimental Results

We present two applications : compressive sensing with Fourier measurement and image deconvolution. We compare the nonlocal TV regularization with standard TV regularization and wavelet based l^1 regularization with the GPSR algorithm [15] available at $http : //www.lx.it.pt/\ mtf/GPSR/GPSR\_5.0.zip$.

We compute the weight function as follows. A full computation of the nonlocal weight function is computationally expensive. 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 21×21 centered at x and the 4 nearest neighbors on comparing 5×5 patches with the formula (13).

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

```
Initialization: : $u^0 = v^0 = 0$, $f^0 = f, h, \mu, \delta, K, M, N$.
  for $k = 0$ to $K$ 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
      $v^{k+1} = u^k - (A^T A + \frac{1}{\delta})^{-1} A^T (Au^k - f^k)$
    end if
    if ( $\mod (K, M) == 0$) then
      Every $m$ steps update the nonlocal weight $w^{(k)} = w(v^{k+1})$ // using the formula (13)
    end if
    Inner denoising step: Performing $N$ steps of the nonlocal TV denoising iteration with input $v^{k+1}, \delta\mu$.
    if $||Au^{k+1} - f|| \geq btol$ and $\|u^{k+1} - u^k\| < xtol$ then
      Break.
    end if
    Update $f^{k+1} = f^k + f - Au^{k+1}$.
  end for
```

For the nonlocal TV denoising step, we implemented both projection method and split Bregman method (Algorithm 1) with 20 steps. Although the split Bregman approach produced slightly better results, the performance of both algorithms are quite similar. In the following, we present the result with split Bregman denoising method.

7.1 Fast Nonlocal TV Deconvolution

We run the BOS and PBOS methods on the Cameraman image for the image deconvolution problem. In [20], a fixed weight is computed based on a Tikhonov-based deblurred image by using an optimal μ and estimated noise level. The gradient descent algorithm is usually applied to solve the unconstrained formulation:

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

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

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

by using Bregmanized operator splitting (BOS) and preconditioned Bregman operator splitting (PBOS) until the residual noise level is around σ, that is , we set the stopping criterion $btol = \tau\sigma$, with $\tau = 0.99$. Figure 1 shows the results and compare different algorithms. The images reconstructed by NLTV+BOS and NLTV+PBOS present better contrasts and edges. Compared to the result in [20], the algorithm BOS and PBOS takes less than 20 steps to meet the stopping criterion and the results are similar.

We also tested the weight updating scheme, it appears that there is no significant improvement compared to a fixed weight function. This means that the weight function computed from a deblurred image is good enough to express structured information in the nonlocal TV regularization.

7.2 Compressive Sensing

In this section, we focus on exploring the sparsity of natural images with non-local regularization operators. The general 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 $m \times (N^2)$. As we expect, 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. Figures 2 and 3 present the results for the Barbara picture and a five-texture picture. Images are obtained from random sampled Fourier coefficients with $m = \text{round}(0.3 * N * N)$. 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 optimal thresholding parameter setting in a wide range. Since there is no noise considered in these two examples, we solve the equality constrained problem by activating the continuation option in the GPSR code. The nonlocal regularization schemes with Bregman iteration (BOS/PBOS) achieve the best reconstruction result. 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. For both examples, we use $K = 500$ outer iterations and $N = 20$ inner iterations for nonlocal TV denoising. The weights are updated every $M = 20$ steps. The scale parameters are fixed to be $\mu = 10$ and $\delta = 1$.

8 Discussion

In this paper, we propose two general algorithms for convex minimization problems with equality constraints. 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 [17], 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, it may be also possible to apply directly the split Bregman method for inverse problems instead of using operator splitting. We will investigate more carefully this question in the future. Furthermore, as mentioned in [24], it is also important to better understand the weight updating strategy in a theoretical framework.

Acknowledgements

Xiaoqun Zhang was supported by ARO MURI subcontract from the University of South Carolina and NSF DMS 0312222. The work of Martin Burger has been 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”. Xavier Bresson was supported by ONR N00014-03-1-0071, ONR MURI subcontract from Stanford University. Stanley Osher was supported by ONR N000140710810, ONR N00014-08-1-1119 and NSF DMS-07-14087. Martin Burger and Stanley Osher thank Fondazione CIME for a summer school in stimulating atmosphere, initiating a part of this project.

References

[1] T. Brox and D. Cremers. Iterated nonlocal means for texture restoration. In SpringerLink, editor, *Scale Space and Variational Methods in Computer Vision*, volume 4485/2007, pages 13–24. Springer Berlin / Heidelberg, 2007. 7

[2] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.*, 4(2), 2005. 6

[3] A. Buades, B. Coll, and J. M. Morel. Image enhancement by non-local reverse heat equation. Technical Report 22, CMLA, 2006. 7

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

[5] 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. 5, 6

[6] J.-F. Cai, S. Osher, and Z. Shen. Linearized Bregman iterations for Frame-based Image Deblurring. CAM Report 08-57, UCLA, May 2008. 12

[7] 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

[8] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004. 5, 9

[9] F. R. K. Chung. *Spectral graph theory*, pages 1–212. American Mathematical Society, 1997. 7

[10] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting, 2005. 4, 5, 14

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

[12] J. Darbon and S. Osher. Fast discrete optimization for sparse approximations and deconvolutions. preprint, UCLA, 2007. 3, 5

[13] 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. 6

[14] A. Elmoataz, O. Lezoray, and S. Bougleux. Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing, 2008. 7

[15] 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. 18

[16] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. CAM Report 07-23, UCLA, July 2007. 7, 9

[17] T. Goldstein and S. Osher. The split bregman method for $l1$ regularized problems. CAM Report 08-29, UCLA, 2008. 3, 5, 9, 10, 20

[18] 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. 5

[19] 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. 7

[20] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi. Image recovery via nonlocal operators. CAM Report 08-35, UCLA, May 2008. 12, 19

[21] 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. 4

[22] 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. 3

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

[24] G. Peyré, S. Bougleux, and L. Cohen. Non-local regularization of inverse problems. In *Proceedings of Proc. ECCV 2008*, 2008. 14, 20

[25] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. 1, 3, 9

[26] W. Yin. On dual penality methods for basis pursuit. Private communication, Rice University, 2008. 6

[27] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algoruithms for $l1$ minimization with applications to compressed sensing. *SIAM J. Imaging Sci.*, 1:143–168, 2008. 3, 5

[28] D. Zhou and B. Schölkopf. Regularization on discrete spaces. *Pattern Recognition*, 361:361–368, 2005. 7

[29] W. Zulehner. Analysis of iterative methods for saddle point problems: A unified approach. *Mathematics of Computation*, 71(238):479–505, 2001. 2, 18

Original Image

Blurry and noisy

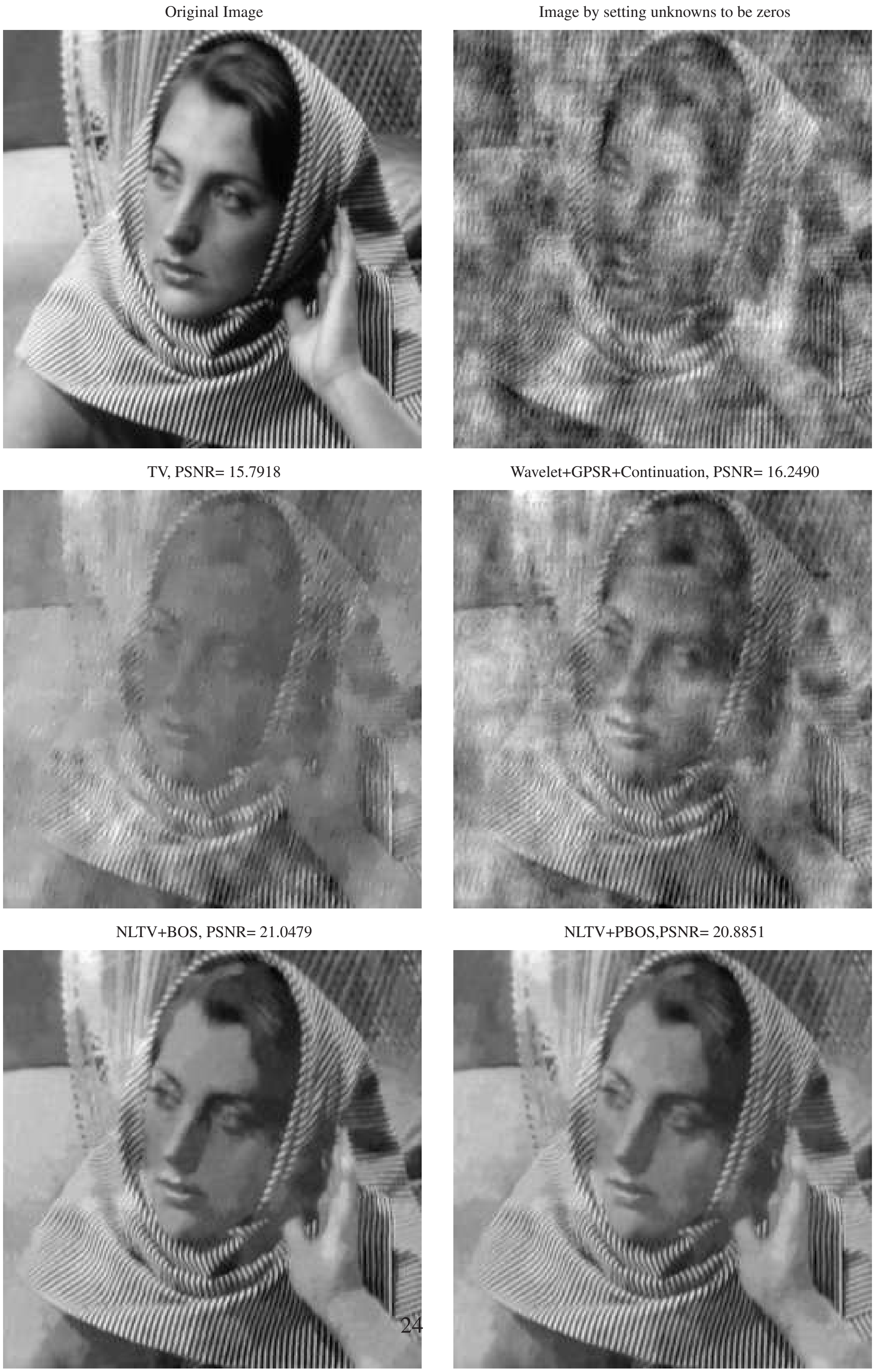
TV, PSNR= 25.0088

Wavelet+GPSR, PSNR= 23.2196

NLTV+BOS, PSNR= 25.0566

NLTV+PBOS, PSNR= 25.0635

Figure 1: Deconvolution example: Cameraman (256×256), box average kernel of 9×9, Gaussian noise $\sigma = 3$. Weight fixed.



Original Image

Image by setting unknowns to be zeros

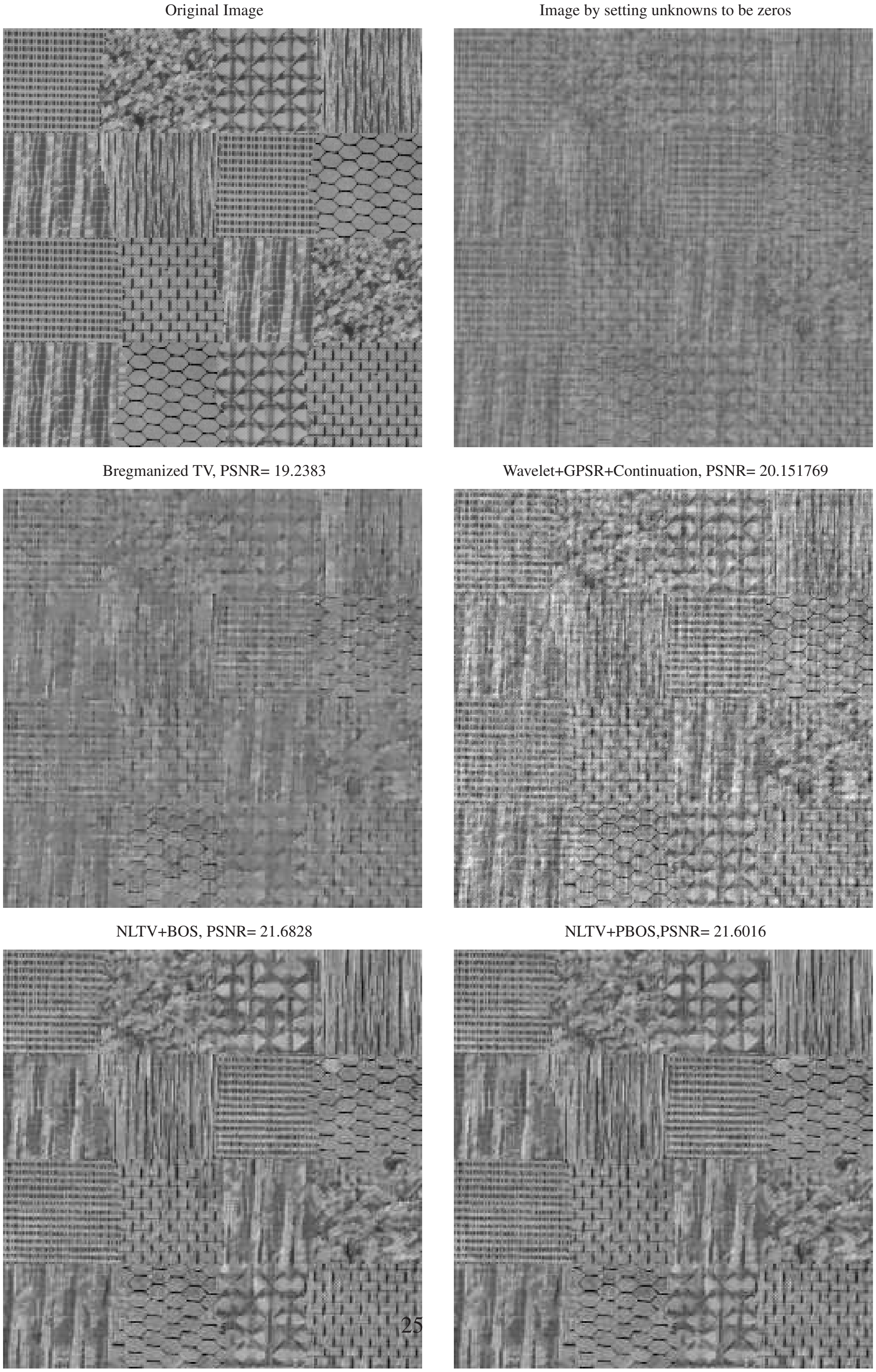
TV, PSNR= 15.7918

Wavelet+GPSR+Continuation, PSNR= 16.2490

NLTV+BOS, PSNR= 21.0479

NLTV+PBOS,PSNR= 20.8851

Figure 2: Compressive sensing example: Barbara (256×256), 30% randomly chosen Fourier coefficients, noiseless. Weight updating.



Original Image

Image by setting unknowns to be zeros

Bregmanized TV, PSNR= 19.2383

Wavelet+GPSR+Continuation, PSNR= 20.151769

NLTV+BOS, PSNR= 21.6828

NLTV+PBOS,PSNR= 21.6016

Figure 3: Compressive sensing example: Textures (256 × 256), 30% randomly chosen Fourier coefficients, noiseless. Weight updating.