BREGMAN ITERATIVE ALGORITHMS FOR ℓ_1-MINIMIZATION WITH APPLICATIONS TO COMPRESSED SENSING

WOTAO YIN*, STANLEY OSHER†, DONALD GOLDFARB‡, AND JEROME DARBON†

Abstract. We propose simple and extremely efficient methods for solving the Basis Pursuit problem

$$\min\{\|u\|_1 : Au = f, u \in \mathbb{R}^n\},$$

which is used in compressed sensing. Our methods are based on Bregman iterative regularization and they give a very accurate solution after solving only a very small number of instances of the unconstrained problem

$$\min_{u\in\mathbb{R}^n} \mu\|u\|_1 + \frac{1}{2}\|Au - f^k\|_2^2,$$

for given matrix A and vector f^k. We show analytically that this iterative approach yields exact solutions in a finite number of steps, and present numerical results that demonstrate that as few as two to six iterations are sufficient in most cases. Our approach is especially useful for many compressed sensing applications where matrix-vector operations involving A and $A^\top$ can be computed by fast transforms. Utilizing a fast fixed-point continuation solver that is solely based on such operations for solving the above unconstrained sub-problem, we were able to solve huge instances of compressed sensing problems quickly on a standard PC.

Key words. ℓ_1-Minimization, Basis Pursuit, Compressed Sensing, Iterative Regularization, Bregman Distances.

1. Introduction. Let $A \in \mathbb{R}^{m\times n}$, $f \in \mathbb{R}^m$, and $u \in \mathbb{R}^n$. The Basis Pursuit problem [23] solves the constrained minimization problem

$$\text{(Basis Pursuit)} \quad \min_u\{\|u\|_1 : Au = f\} \tag{1.1}$$

to determine an ℓ_1-minimal solution u_{opt} of the linear system $Au = f$, typically underdetermined, i.e., $m < n$ (in many cases, $m \ll n$) and $Au = f$ has more than one solution.

The Basis Pursuit problem (1.1) arises in the applications of compressed sensing (CS). A recent burst of research in CS was led by Candés *et al* [12, 14, 16], Donoho *et al* [33, 34, 86], and others [78, 84]. The fundamental principle of CS is that, through optimization, the sparsity of a signal can be exploited for recovering that signal from incomplete measurements of it. Let the vector $\bar{u} \in \mathbb{R}^n$ denote a highly-sparse signal (i.e., $k = \|\bar{u}\|_0 := |\{i : \bar{u}_i \neq 0\}| \ll n$). This principle states that one can encode $\bar{u}$ by a linear transform $f = A\bar{u} \in \mathbb{R}^m$ for some m greater than k, but much smaller than n, and then recover $\bar{u}$ from f by solving (1.1). It is proved that the recovery is perfect, i.e., the solution $u_{\text{opt}} = \bar{u}$, for any $\bar{u}$ whenever k, m, n, and A satisfy certain conditions (e.g., see [13, 30, 37, 42, 78, 95, 96]). While these conditions are computationally intractable to check, it was found in [15, 16] and other work that the types of matrices A allowing a high compression ratio (i.e., $m \ll n$) include random matrices with i.i.d. entries and random ensembles of orthonormal transforms (e.g., matrices formed from random sets of rows of the matrices corresponding to Fourier and cosine transforms).

Recent applications of ℓ_1 minimization can be found in [49, 82, 89, 90] for compressive imaging, [59, 66, 68, 67, 94] for MRI and CT, [3, 4, 48, 52, 74, 91] for multi-sensor networks and distributive sensing,

*Department of Computational and Applied Mathematics, Rice University, wotao.yin@rice.edu. Research supported by an internal faculty research grant from the Dean of Engineering at Rice University.

†Department of Mathematics, UCLA, sjo@ and jerome@math.ucla.edu. Research supported by ONR grant N000140710810.

‡Department of Industrial Engineering and Operations Research, Columbia University, goldfarb@columbia.edu. Research supported in part by NSF Grant DMS 06-06712, ONR Grant N00014-03-0514 and DOE Grant GE-FG01-92ER-25126.

[61, 63, 64, 75, 85] for analog-to-information conversion, and [81] for bio-sensing. ℓ_1 minimization also has applications in image inpainting and missing data recovery; see [40, 80, 96] for example. Also non-convex quasi ℓ_p-norm approaches for $0 \leq p < 1$ have been proposed by Chartrand [20, 21] and he and Yin [22].

Problem (1.1) can be transformed into a linear program and then solved by conventional linear programming solvers. However, such solvers are not tailored for the matrices A that are large-scale and completely dense, or are formed by rows taken from orthonormal matrices corresponding to fast transforms so that Ax and $A^\top x$ can be computed by fast transforms. This, together with the fact that f may contain noise in certain applications, makes solving the unconstrained problem

$$\min_u \mu\|u\|_1 + \frac{1}{2}\|Au - f\|_2^2 \tag{1.2}$$

more preferable than solving the constrained problem (1.1) (e.g., see [24, 25, 27, 35, 39, 47, 56, 60, 87]). Hereafter, we use $\|\cdot\| \equiv \|\cdot\|_2$ to denote the 2-norm. In Section 2.1 below, we give a review of recent numerical methods for solving (1.2). Because (1.2) also allows the constraint $Au = f$ to be relaxed, it is used when the measurement f is contaminated by encoding errors such as noise. However, when there is no encoding error, one must assign a tiny value to μ to heavily weigh the fidelity term $\|Au - f\|^2$ in order for $Au = f$ to be nearly satisfied. Furthermore, one can show that the solution of (1.2) never equals that of (1.1) unless they both have the trivial solution $\mathbf{0}$. In this paper, we introduce a simple method based on Bregman iterative regularization [71], which we review in Section 2.2 below, for finding a solution of problem (1.1) by solving only a small number of instances of the unconstrained problem (1.2). Our numerical algorithm, based on this iterative method, calls the fast fixed-point continuation solver FPC [53, 54] of (1.2), which only involves matrix-vector multiplications (or fast linear transforms) and component-wise shrinkages (defined in (2.4) below). Using a moderate value for the penalty parameter μ, we were able to obtain a very accurate solution to the original Basis Pursuit problem (1.1) for a very small multiple of the cost of solving a single instance of (1.2).

Our results can also be generalized to the constrained problem

$$\min_u \{J(u) : Au = f\}, \tag{1.3}$$

for other types of convex functions J (refer to Section 5). Specifically, a solution of (1.3) can be obtained through a finite number of the Bregman iterations of

$$\min_u \mu J(u) + \frac{1}{2}\|Au - f\|^2. \tag{1.4}$$

In addition, in Section 5.3, we also introduce a two-line algorithm (given in Equations (5.19) and (5.20)) also involving only matrix-vector multiplication and shrinkage operators that generates a sequence $\{u^k\}$ that converges rapidly to an approximate solution of the Basis Pursuit problem (1.1). In fact, the numerical experiments in [32] indicate that this algorithm converges to a true solution if the parameter μ is large enough. Finally, preliminary experiments indicate that our algorithms are robust with respect to a certain amount of noise. This is also implied by our theoretical results stated in Theorems 2.1 and 5.5.

The rest of the paper is organized as follows. In Section 2, we summarize the existing methods for solving the unconstrained problem (1.2) and provide some background on our Bregman iterative regularization scheme. The main Bregman iterative algorithm is described in Section 3.1; its relationship to some previous work [93] is presented in Section 3.2; and its convergence is analyzed in Section 3.3. Numerical results are presented in Section 4. Finally, we extend our results to more general classes of problems in Section 5, including a description and analysis of our linearized Bregman iterative scheme, and conclude the paper in Section 6.

2. Background.

2.1. Solving the Unconstrained Problem (1.2).

Several recent algorithms can efficiently solve (1.2) with large-scale data. The authors of GPSR [46], Figueiredo, Nowak, and Wright [47], reformulate (1.2) as a box-constrained quadratic program, to which they apply the gradient projection method with Barzilai-Borwein steps. The algorithm ℓ_1_ℓ_s [62] by Kim, Koh, Lustig, and Boyd [60] was developed for an ℓ_1 regularization problem equivalent to (1.2). The authors apply an interior-point method to a log-barrier formulation of (1.2). The main step in each interior-point iteration, which involves solving a system of linear equations, is accelerated by using a preconditioned conjugate gradient method, for which the authors developed an efficient pre-conditioner. In the code SPGL1 [88], Van den Berg and Friedlander apply an iterative method for solving the LASSO problem [83], which minimizes $\|Au - f\|$ subject to $\|u\|_1 \leq \sigma$, by using an increasing sequence of σ-values in their algorithm to accelerate the computation. In [69], Nesterov proposes an accelerated multistep gradient method with an error convergence rate $O(1/k^2)$. Under some conditions, the greedy approach StOMP [35] by Donoho, Tsaig, Drori, and Starck can also quickly solve (1.2).

A widely used method by many researchers to solve (1.2) or the general ℓ_1-minimization problems of the form:

$$\min_u \mu\|u\|_1 + H(u) \tag{2.1}$$

for convex and differentiable functions $H(\cdot)$ is an iterative procedure based on shrinkage (also called soft thresholding; see Eq. (2.4) below). It was independently proposed and analyzed by Figueiredo and Nowak in [44, 70] under the expectation-minimization framework for wavelet-based deconvolution, De Mol and Defrise [29] for wavelet inversion, Bect, Blance-Feraud, Aubert, and Chambolle in [5] using an auxiliary variable and the idea from Chambolle's projection method [17], Elad in [36] and he with Matalon, Shtok, and Zibulevsky [38] for sparse representation and other related problems, Daubechies, DeFrise, and DeMol in [27] through an optimization transfer technique, Combettes and Pesquet [24] using operator-splitting, Hale, Yin, and Zhang [53] also using operator-splitting combined with a continuation technique in their code FPC [54], Darbon and Osher [25] through an implicit PDE approach, and others. In addition, related applications and algorithms can be found in Adeyemi and Davies [1] for image sparse representation, Bioucas-Dian [6] for wavelet-based image deconvolution using a Gaussian scale mixture model, Bioucas-Dias and Figueiredo for a recent "two-step" shrinkage-based algorithm [7], Blumensath and Davies [8] for solving a cardinality constrained least-squares problem, Chambolle *et al* [19] for image denoising, Daubechies, Fornasier, and Loris [28] for a direct and accelerated projected gradient method, Elad, Matalon, and Zibulevsky in [39] for image denoising, Fadili and Starck [41] for sparse representation-based image deconvolution, Figueiredo and Nowak [45] for image deconvolution based on a bound optimization and they together with Bioucas-Dias [43] for wavelet-based image denoising using majorization-minimization algorithms, and Kingsbury and Reeves [76] for image coding. These authors, using different approaches, developed or used algorithms based on the iterative scheme

$$u^{k+1} \leftarrow \arg\min_u \mu\|u\|_1 + \frac{1}{2\delta^k}\left\|u - (u^k - \delta^k \nabla H(u^k))\right\|^2 \tag{2.2}$$

for $k = 0, 1, \ldots$ starting from a certain point $u^0 = \mathbf{0}$. The parameter δ^k is positive and serves as the step size at iteration k. Since the unknown variable u is component-wise separable in problem (2.2), each of its components u_i can be independently obtained by the shrinkage operation, which is also referred to as soft thresholding:

$$u_i^{k+1} = \text{shrink}((u^k - \delta^k \nabla H(u^k))_i, \mu\delta^k), \; i = 1, \ldots, n, \tag{2.3}$$

where for $y, \alpha \in \mathbb{R}$, we define

$$\text{shrink}(y, \alpha) := \text{sgn}(y) \max\{|y| - \alpha, 0\} = \begin{cases} y - \alpha, & y \in (\alpha, \infty), \\ 0, & y \in [-\alpha, \alpha], \\ y + \alpha, & y \in (-\infty, -\alpha). \end{cases} \tag{2.4}$$

Among the several approaches giving (2.2), one of the easiest ones is the following: first, $H(u)$ is approximated by its first-order Taylor expansion at u^k, which is $H(u^k) + \langle \nabla H(u^k), u - u^k \rangle$; then, since this approximation is accurate only for u near u^k, u must be made close to u^k so an ℓ_2-penalty term $\|u - u^k\|^2/(2\delta^k)$ is added to the objective; the resulting step is

$$u^{k+1} \leftarrow \arg\min_u \mu\|u\|_1 + H(u^k) + \langle \nabla H(u^k), u - u^k \rangle + \frac{1}{2\delta^k}\|u - u^k\|^2, \tag{2.5}$$

which is equivalent to (2.2) because their objectives differ by only a constant. It is easy to see that the larger the δ^k, the larger the allowable distance between u^{k+1} and u^k. It was proved in [53] that $\{u^k\}$ given by (2.2) converges to an optimum of (1.4) at a q-linear* rate under certain conditions on H and δ^k. Their convergence results are based in part on the previous work by Pang [72], and Luo and Tseng [65] on gradient projection methods. Furthermore, a new result from [53] is that the support (i.e., $\{i : u_i^k \neq 0\}$) and signs of u^k converge finitely; that is, there exists a finite number K such that $\text{sgn}(u^k) \equiv \text{sgn}(u_{\text{opt}})$, $\forall k > K$, where u_{opt} denotes the solution of (1.4); however, an estimate or bound for K is not known.

To improve the efficiency of the iterations (2.2), various techniques have been applied to (2.2), which include generalizing (2.3) by using more parameters [39], employing various types of linear search on δ^k and u^{k+1} [47], and using a decreasing sequence of μ-values [53]. The last technique is called path following or continuation. While our algorithm does not depend on using a specific code, we chose to use FPC [54], one of the fastest codes, to solve each subproblem in (2.2).

In [25], [92] and other work, the iterative procedure (2.2) is adapted for solving the total variation regularization problem

$$\min_u \mu TV(u) + H(u), \tag{2.6}$$

where $TV(u)$ denotes the total variation of u (see [97] for a definition of $TV(u)$ and its properties). Specifically, the regularization term $\mu\|u\|_1$ in (2.2) is replaced by $\mu TV(u)$, yielding

$$u^{k+1} \leftarrow \arg\min_u \mu TV(u) + \frac{1}{2\delta^k}\left\|u - (u^k - \delta^k \nabla H(u^k))\right\|^2. \tag{2.7}$$

Each subproblem (2.7) can be efficiently solved by one of the recent graph/network-based algorithms [18,

*q stands for "quotient"; $\{x^k\}$ converges to x^* q-linearly if $\lim_k \|x^{k+1} - x^*\|/\|x^k - x^*\|$ exists and is less than 1.

26, 51]. In [25] Darbon and Osher also study an algorithm obtained by replacing $\mu TV(u)$ in (2.7) by its Bregman distance (see Subsection 2.2 below) and proved that if $H(u) = 0.5\|Au - f\|^2$ then $\{u^k\}$ converges to the solution of $\min_u\{TV(u) : Au = f\}$. Their algorithm and results are described in Section 5.3. In the next subsection, we give an introduction to Bregman iterative regularization.

2.2. Bregman Iterative Regularization. Bregman iterative regularization was introduced by Osher, Burger, Goldfarb, Xu, and Yin [71] in the context of image processing; it was then extended to wavelet-based denoising [93], nonlinear inverse scale space in [10, 11], and compressed sensing in MR imaging [57]. The authors of [71] extend the Rudin-Osher-Fatemi [79] model

$$\min_u \mu \int |\nabla u| + \frac{1}{2}\|u - b\|^2 \tag{2.8}$$

where u is an unknown image, b is typically an input noisy measurement of a clean image $\bar{u}$, and μ is a tuning parameter, into an iterative regularization model by using the Bregman distance (2.10) below based on the total variation functional:

$$J(u) = \mu TV(u) = \mu \int |\nabla u|. \tag{2.9}$$

Specifically, the Bregman distance [9] based on a convex functional $J(\cdot)$ between points u and v is defined as

$$D_J^p(u, v) = J(u) - J(v) - \langle p, u - v\rangle \tag{2.10}$$

where $p \in \partial J(v)$ is some subgradient in the subdifferential of J at the point v. Because $D_J^p(u, v) \neq D_J^p(v, u)$ in general, $D_J^p(u, v)$ is not a distance in the usual sense. However, it measures the closeness between u and v in the sense that $D_J^p(u, v) \geq 0$ and $D_J^p(u, v) \geq D_J^p(w, v)$ for all points w on the line segment connecting u and v.

Instead of solving (2.8) once, the Bregman iterative regularization procedure of Osher et. al. [71] solves a sequence of convex problems

$$u^{k+1} \leftarrow \min_u D_J^{p^k}(u, u^k) + \frac{1}{2}\|u - b\|^2 \tag{2.11}$$

for $k = 0, 1, \ldots$ starting with $u^0 = \mathbf{0}$ and $p^0 = \mathbf{0}$ (hence, for $k = 0$, one solves the original problem (2.8).) Since $\mu TV(u)$ is not differentiable everywhere, the subdifferential of $\mu TV(u)$ may contain more than one element. However, from the optimality of u^{k+1} in (2.11), it follows that $\mathbf{0} \in \partial J(u^{k+1}) - p^k + u^{k+1} - b$; hence, they set

$$p^{k+1} := p^k + b - u^{k+1},$$

The difference between (2.8) and (2.11) is in the use of regularization. While (2.8) regularizes u by directly minimizing its total variation, (2.11) does the same by minimizing the total variation-based Bregman distance of u to a previous solution u^k.

In [71] two key results for the sequence $\{u^k\}$ generated by (2.11) were proved. First, $\|u^k - b\|$ converges to 0 monotonically; second, u^k also monotonically gets closer to $\bar{u}$, the *unknown* noiseless image, in terms of the Bregman distance $D_{TV}^{p^k}(\bar{u}, u^k)$, at least while $\|u^k - b\| \geq \|\bar{u} - b\|$. Numerical results in [71] demonstrate that for μ sufficiently large, this simple iterative procedure remarkably improves denoising quality over the original model (2.8).

Interestingly, not only for the first iteration $k = 0$, but for all k, the new problem (2.11) can be reduced to the original problem (2.8) with the input $b^{k+1} := b + (b^k - u^k)$ starting with $b^0 = u^0 = \mathbf{0}$, i.e., the iterations (2.11) are equivalent to

$$u^{k+1} \leftarrow \min_u J(u) + \frac{1}{2}\|u - b^{k+1}\|^2, \text{ where } b^{k+1} = b + (b^k - u^k), \tag{2.12}$$

and can be carried out using any existing algorithms for (2.8).

The iterative procedure (2.12) has an intriguing interpretation: Let ω represent the noise in b, i.e., $b = \bar{u} + \omega$, and μ be large. At $k = 0$, $b^k - u^k = \mathbf{0}$, so (2.12) decomposes the input noisy image b into $u^1 + v^1$. Since μ is large, the resulting image u^1 is over smoothed (by total variation minimization) so it does not contain any noise. Consequently, u^1 can be considered to be a portion of the original clean image $\bar{u}$. The residual $v^1 = b - u^1 = (\bar{u} - u^1) + \omega$, hence, is the sum of the unrecovered "good" signal $(\bar{u} - u^1)$ and the "bad" noise ω. We wish to recover $(\bar{u} - u^1)$ from v^1. Intuitively, one would next consider letting v^1 be the new input for (2.8) and solving (2.8). However, Bregman iterative regularization turns out to be both better and "nonintuitive": it adds v^1 back to the original input b. The the new input of (2.12) in the 2nd iteration is

$$b + v^1 = (u^1 + v^1) + v^1 = u^1 + 2(\bar{u} - u^1) + 2\omega.$$

which, compared to the original input $b = u^1 + (\bar{u} - u^1) + \omega$, contains twice as much of both the unrecovered "good" signal $\bar{u} - u^1$ and the "bad" noise ω. What is remarkable is that the new decomposition u^2 is a better approximation to $\bar{u}$ than u^1 (for μ large enough); one explanation is that u^2 not only inherits u^1 but also captures a part of $(\bar{u} - u^1)$, the previously un-captured "good" signal. Of course, as the convergence results indicate, u^k will eventually pick up the noise ω since $\{u^k\}$ converges to $b = \bar{u} + \omega$. However, a high quality image can be found among the sequence $\{u^k\}$: the image u^k that has $\|u^k - b\|$ closest to $\|\bar{u} - b\|$ is cleaner and has a higher contrast than the best image that one could possibly obtain from solving (2.8) one single time.

Formally Bregman iterative regularization applied to the problem

$$\min_u J(u) + H(u) \tag{2.13}$$

is given as Algorithm 1 in which the Bregman distance $D_J^{p^k}(\cdot, \cdot)$ is defined by (2.10).

Algorithm 1 Bregman Iterative Regularization

Require: $J(\cdot)$, $H(\cdot)$

1: Initialize: $k = 0$, $u^0 = \mathbf{0}$, $p^0 = \mathbf{0}$.
2: **while** "not converge" **do**
3: $\quad u^{k+1} \leftarrow \arg\min_u D_J^{p^k}(u, u^k) + H(u)$
4: $\quad p^{k+1} \leftarrow p^k - \nabla H(u^{k+1}) \in \partial J(u^{k+1})$
5: $\quad k \leftarrow k + 1$
6: **end while**

Below we conclude this section by citing some useful convergence results from [71] that are used in Section 3.3 below.

ASSUMPTION 1. *$J(\cdot)$ is convex, $H(\cdot)$ is convex and differentiable, and the solutions u^{k+1} in Step 3 of*

Algorithm 1 exist.

THEOREM 2.1. *Under Assumption 1, the iterate sequence* $\{u^k\}$ *satisfies*

1. *Monotonic decrease in* H: $H(u^{k+1}) \leq H(u^{k+1}) + D_J^{p^k}(u^{k+1}, u^k) \leq H(u^k)$;
2. *Convergence to the original in* H *with exact data: if* $\tilde{u}$ *minimizes* $H(\cdot)$ *and* $J(\tilde{u}) < \infty$, *then* $H(u^k) \leq H(\tilde{u}) + J(\tilde{u})/k$;
3. *Convergence to the original in* D *with noisy data: let* $H(\cdot) = H(\cdot; f)$ *and suppose* $H(\tilde{u}; f) \leq \delta^2$ *and* $H(\tilde{u}; g) = 0$ (f, g, $\tilde{u}$, *and* δ *represent noisy data, noiseless data, perfect recovery, and noise level, respectively); then* $D_J^{p^{k+1}}(\tilde{u}, u^{k+1}) < D_J^{p^k}(\tilde{u}, u^k)$ *as long as* $H(u^{k+1}; f) > \delta^2$.

3. Bregman Iterations for Basis Pursuit.

3.1. Formulations. The main purpose of this paper is to show that the Bregman iterative procedure is a simple but very efficient method for solving the Basis Pursuit problem (1.1), as well as a broader class of problems of the form (1.3), in both theory and practice. Below we first give the details of the algorithm, describe our motivation, and then prove that in a finite number of iterations, u^k becomes a minimizer of $\|u\|_1$ among $\{u : Au = f\}$.

We solve the constrained problem (1.1) by applying Algorithm 1 to (1.2) for $J(u) = \mu\|u\|_1$ and $H(u) = \frac{1}{2}\|Au - f\|^2$:

Version 1:

$$u^0 \leftarrow \mathbf{0},\ p^0 \leftarrow \mathbf{0}, \tag{3.1}$$

For $k = 0, 1, \ldots$ do

$$u^{k+1} \leftarrow \arg\min_u D_J^{p^k}(u, u^k) + \frac{1}{2}\|Au - f\|^2, \tag{3.2}$$

$$p^{k+1} \leftarrow p^k - A^\top(Au^{k+1} - f); \tag{3.3}$$

Version 2:

$$f^0 \leftarrow \mathbf{0},\ u^0 \leftarrow \mathbf{0}, \tag{3.4}$$

For $k = 0, 1, \ldots$ do

$$f^{k+1} \leftarrow f + (f^k - Au^k), \tag{3.5}$$

$$u^{k+1} \leftarrow \arg\min_u J(u) + \frac{1}{2}\left\|Au - f^{k+1}\right\|^2. \tag{3.6}$$

Given u^k and p^k in Version 1, u^{k+1} satisfies the first-order optimality condition:

$$\mathbf{0} \in \partial J(u^{k+1}) - p^k + \nabla H(u^{k+1}) = \partial J(u^{k+1}) - p^k + A^\top(Au^{k+1} - f).$$

Therefore,

$$p^{k+1} = p^k - A^\top(Au^{k+1} - f) \in \partial J(u^{k+1}); \tag{3.7}$$

hence, $D_J^{p^{k+1}}(u, u^{k+1})$ is well-defined. Clearly, if $u_i^{k+1} = 0$, then one can pick any $p_i^{k+1} \in [-1, 1]$ and still have a well-defined $D_J^{p^{k+1}}(u, u^{k+1})$. However, the choice of p^{k+1} as in (3.7) is not only simple but also crucial for the sequence $\{u^k\}$ to converge to the minimizer u_{opt} of the constrained problem (1.1).

THEOREM 3.1. *The Bregman iterative procedure Version 1* (3.1)–(3.3) *and Version 2* (3.4)–(3.6) *are equivalent in the sense that* (3.2) *and* (3.6) *have the same objective functions (they may differ by up to a* *constant) for all k.*

Proof. Let u^k and $\bar{u}^k$ denote the solutions to Versions 1 and 2, respectively. The initialization (3.1) gives $D_J^{p^0}(u, u^0) = J(u)$ while (3.4) gives $f^1 = f$. Therefore, at iteration $k = 0$, (3.2) and (3.6) have the same optimization problem

$$\min_u J(u) + \frac{1}{2}\|Au - f\|^2.$$

We note that this problem, as well as those for all other k, may have more than one solution. We do *not* assume that in this case, u^1 (Version 1) is equal to $\bar{u}^1$ (Version 2). Instead, we use the fact from [53] that $A^\top(f - Au)$ is constant for all optimal solutions u, i.e., $A^\top(f - Au^1) = A^\top(f - A\bar{u}^1)$. According to (3.3), $p^0 = \mathbf{0}$ and $f = f^1$, we have

$$p^1 = p^0 - A^\top(Au^1 - f) = A^\top(f - Au^1) = A^\top(f - A\bar{u}^1) = A^\top(f^1 - A\bar{u}^1).$$

Next, we use induction on $p^k = A^\top(f^k - A\bar{u}^k)$. Given $p^k = A^\top(f^k - A\bar{u}^k)$, we will show the followings: (i) the optimization problems in (3.2) and (3.6) at iteration k are equivalent, (ii) $A^\top(Au^{k+1} - f) = A^\top(A\bar{u}^{k+1} - f)$, and (iii) $p^{k+1} = A^\top(f^{k+1} - A\bar{u}^{k+1})$. Clearly, part (i) proves the theorem.

Part (i): From the induction assumption it follows

$$\begin{aligned} D_J^{p^k}(u, u^k) + \frac{1}{2}\|Au - f\|^2 &= J(u) - \langle p^k, u\rangle + \frac{1}{2}\|Au - f\|^2 + C_1 \\ &= J(u) - \langle f^k - A\bar{u}^k, Au\rangle + \frac{1}{2}\|Au - f\|^2 + C_2 \\ &= J(u) + \frac{1}{2}\|Au - (f + (f^k - A\bar{u}^k))\|^2 + C_3 \\ &= J(u) + \frac{1}{2}\|Au - f^{k+1}\|^2 + C_3, \end{aligned}$$

where C_1, C_2, and C_3 stand for the terms constant in u; hence, (3.2) and (3.6) have the same objective function (up to a constant).

Part (ii): $A^\top(Au^{k+1} - f) = A^\top(A\bar{u}^{k+1} - f)$ follows from part (i) and the result in [53].

Part (iii): It follows from the induction assumption, as well as (3.3), (3.5) and part (ii), that

$$\begin{aligned} p^{k+1} = p^k - A^\top(Au^{k+1} - f) &= p^k - A^\top(A\bar{u}^{k+1} - f) \quad (3.8) \\ &= A^\top(f^k - A\bar{u}^k) - A^\top(A\bar{u}^{k+1} - f) \\ &= A^\top\left(f + (f^k - A\bar{u}^k) - A\bar{u}^{k+1}\right) \\ &= A^\top(f^{k+1} - A\bar{u}^{k+1}). \quad (3.9) \end{aligned}$$

□

Remark: When J is not strictly convex, the subproblems in Versions 1 and 2 may both have more than one solution. The above proof shows, however, even if Versions 1 and 2 generate different intermediate solutions at certain iteration, they remain equivalent thereafter.

Each iteration of (3.6) is an instance of (1.2), which can be solved by the code FPC [54]. Although our convergence result below holds for any strictly positive μ, we choose μ so that (1.2) is solved efficiently by FPC and the total time of the Bregman iterations is nearly optimal.

3.2. Motivation. In [93], Xu and Osher applied Bregman iterative regularization to wavelet based denoising. Briefly, they considered

$$\min_u \mu\|u\|_{1,1} + \frac{1}{2}\|f-u\|_{L^2}^2, \tag{3.10}$$

where $\|u\|_{1,1}$ is the Besov norm defined in [31]; if $u = \sum_j \tilde{u}_j\psi_j$ and $f = \sum_j \tilde{f}_j\psi_j$, for a wavelet basis $\{\psi_j\}$, they solve

$$\min_{\{\tilde{u}_j\}} \mu\sum_j |\tilde{u}_j| + \frac{1}{2}\sum_j |\tilde{f}_j - \tilde{u}_j|^2.$$

It was observed in [93] and elsewhere that this minimization procedure is equivalent to shrinkage, i.e., $\tilde{u}_j = \text{shrink}(\tilde{f}_j, \mu)$, $\forall j$, where shrink$(\cdot,\cdot)$ is defined in (2.4).

What is interesting is that Bregman iterations gives:

$$\tilde{u}_j^k = \begin{cases} \tilde{f}_j, & |\tilde{f}_j| > \frac{\mu}{k-1}, \\ k\tilde{f}_j - \mu\ \text{sign}(\tilde{f}_j), & \frac{\mu}{k} \le |\tilde{f}_j| \le \frac{\mu}{k-1}, \\ 0, & |\tilde{f}_j| \le \frac{\mu}{k}. \end{cases} \tag{3.11}$$

So soft shrinkage becomes firm shrinkage [50] with thresholds $\tau^{(k)} \le \frac{\mu}{k}$ and $\tau^{(k-1)} = \frac{\mu}{(k-1)}$.

In [10, 11] the concept of nonlinear inverse scale space was introduced and analyzed, which is basically the limit of Bregman iteration as k and μ increase with $\frac{k}{\mu} \to t$. This iterative Bregman procedure then approaches hard thresholding:

$$\tilde{u}_j(t) = \begin{cases} \tilde{f}_j, & |\tilde{f}_j| > \frac{1}{t}, \\ 0, & |\tilde{f}_j| \le \frac{1}{t}. \end{cases} \tag{3.12}$$

For Bregman iterations it takes

$$k_j = \text{ smallest integer } \ge \frac{\mu}{|\tilde{f}_j|} \tag{3.13}$$

iterations to recover $\tilde{u}_j(k) = \tilde{f}_j$, for all $k \ge k_j$. This means that spikes return in decreasing orders of their magnitudes and sparse data comes back very quickly.

Next, we consider the trivial example of minimizing $\|u\|_1$ subject to $a^\top u = f$, where $\mathbf{0} \ne a \in \mathbb{R}^n$ and $f \in \mathbb{R}$. Obviously, the solution is $u_{\text{opt}} = (f/a_j)\mathbf{e}_j$, where $\mathbf{e}_j$ is the j-th unit vector and a_j is the component of a with the largest magnitude. Without loss of generality, we suppose $a \ge \mathbf{0}$, $f > 0$, and the largest component of a is $a_1 > 0$, which is strictly larger than the rest (to avoid solution non-uniqueness); hence, $u_{\text{opt}} = (f/a_1)\mathbf{e}_1$. Let $f^k > 0$ then the solution of the Bregman iterative subproblem

$$\min_u \mu\|u\|_1 + \frac{1}{2}(a^\top u - f^k)^2$$

is given by

$$u^k = \begin{cases} \mathbf{0}, & \mu \geq f^k a_1, \\ \frac{f^k a_1 - \mu}{a_1^2}\mathbf{e}_1, & 0 < \mu < f^k a_1. \end{cases} \tag{3.14}$$

The Bregman iterations (3.6) start with $f^1 = f$. If $\mu \geq f^1 a_1$, then $u^1 = \mathbf{0}$ so $f^2 = f + (f^1 - a^\top \mathbf{0}) = 2f$; hence, as long as u^i remains $\mathbf{0}$, $f^{i+1} = (i+1)f$. Therefore, we have $u^j = \mathbf{0}$ and $f^{j+1} = (j+1)f$, for $j = 1, \ldots, J$ for

$$J = \max\{k : \mu \geq f^k a_1\} = \left\lfloor \frac{\mu}{f a_1} \right\rfloor.$$

If $\mu < f^1 a_1$, $J = 0$. In both cases, $u_1^{J+1} = ((J+1)f a_1 - \mu)/a_1^2$ so

$$f^{J+2} = f + (f^{J+1} - a^\top u^{J+1}) = (J+2)f - a_1 \frac{(J+1)f a_1 - \mu}{a_1^2} = f - \frac{\mu}{a_1}$$

and

$$u^{J+2} = \frac{f^{J+2} a_1 - \mu}{a_1^2}\mathbf{e}_1 = \frac{f}{a_1}\mathbf{e}_1,$$

i.e., $u^{J+2} = u_{\text{opt}}$. Therefore, the Bregman iterations give an exact solution in

$$\left\lfloor \frac{\mu}{f \max_i\{|a_i|\}} \right\rfloor + 2$$

steps for any problem with a one-dimensional signal f.

We believe that these simple examples help explain why our procedure works so well in compressed sensing applications.

3.3. Convergence Results. In this section, we show that the Bregman iterative regularization (3.1)–(3.3) (or equivalently (3.4)–(3.6)) described in Section 3.1 generates a sequence of solutions $\{u^k\}$ that converges to an optimum u_{opt} of the Basis Pursuit problem (1.1) in a finite number of steps; that is, there exists a K such that every u^k for $k > K$ is a solution of (1.1). The analytical results of this section are generalized to many other types of ℓ_1 and related minimization problems in Section 5.

We divide our analysis into two theorems. The first theorem shows that if u^k satisfies the linear constraints $Au^k = f$, then it minimizes $J(\cdot) = \mu\|\cdot\|_1$; the second theorem proves that such a u^k is obtained for a finite k.

THEOREM 3.2. *Suppose an iterate u^k from* (3.2) *satisfies $Au^k = f$, then u^k is a solution of the Basis Pursuit problem* (1.1).

Proof. For any u, by the nonnegativity of the Bregman distance, we have:

$$J(u^k) \leq J(u) - \langle u - u^k, p^k \rangle \tag{3.15}$$
$$= J(u) - \langle u - u^k, A^\top(f^k - Au^k) \rangle \tag{3.16}$$
$$= J(u) - \langle Au - Au^k, f^k - Au^k \rangle \tag{3.17}$$
$$= J(u) - \langle Au - f, f^k - f \rangle, \tag{3.18}$$

where the first equality follows from (3.9).

Therefore, u^k satisfies $J(u^k) \leq J(u)$, for any u satisfying $Au = f$; hence, u^k is an optimal solution of the Basis Pursuit problem (1.1). □

THEOREM 3.3. *There exists a number $K < \infty$ such that any u^k, $k \geq K$, is a solution of the Basis Pursuit problem* (1.1).

Proof. Let (I_+^j, I_-^j, E^j) be a partition of the index set $\{1, 2, \ldots, n\}$, and define

$$U^j := U(I_+^j, I_-^j, E^j) = \{u : u_i \geq 0, i \in I_+^j;\; u_i \leq 0, i \in I_-^j;\; u_i = 0, i \in E^j\} \tag{3.19}$$

$$H^j := \min_u \left\{ \frac{1}{2}\|Au - f\|^2 : u \in U^j \right\}. \tag{3.20}$$

There are a finite number of distinct partitions (I_+^j, I_-^j, E^j) and the union of all possible U^j's is $\mathcal{H}$, the entire space of u.

At iteration k, let (I_+^k, I_-^k, E^k) be defined in terms of p^k as follows:

$$I_+^k = \{i : p^k = \mu\},\; I_-^k = \{i : p^k = -\mu\},\; E^k = \{i : p^k \in (-\mu, \mu)\}. \tag{3.21}$$

In light of the definition (3.19) and the fact that $p^k \in \partial J(u^k) = \partial(\mu\|u^k\|_1)$, we have $u^k \in U^k$. To apply Theorem 2.1, we let $\tilde{u}$ satisfy $H(\tilde{u}) = \frac{1}{2}\|A\tilde{u} - f\|^2 = 0$. Using this $\tilde{u}$ in Statement 2 of Theorem 2.1, we see that for each j with $H^j > 0$ there is a sufficiently large K_j such that u^k is not in U^j for $k \geq K_j$. Therefore, letting $K := \max_j\{K_j : H^j > 0\}$, we have $H(u^k) = 0$ for $k \geq K$. That is $Au^k = f$ for $k \geq K$.

Therefore, it follows from (3.3) that $p^K = p^{K+1} = \cdots$, and then from (3.5) that $f^{K+1} = f^{K+2} = \cdots$. Because the minimizers of Bregman iterations (3.2) and (3.6) are not necessarily unique, the u^k for $k > K$ are not necessarily the same. Nevertheless, it follows from Theorem 3.2 that all u^k for $k > K$ are optimal solutions of the Basis Pursuit problem (1.1). □

Both Theorems 3.2 and 3.3 can be extended to a Bregman iterative scheme in which μ takes on varying values $\{\mu^k\}$ as long as this sequence is bounded above. Suppose $J(u) = J^k(u) = \mu^k\|u\|_1$ and $p^k \in \partial J^k(u)$ at k-th Bregman iteration and $J(u) = J^{k+1}(u) = \mu^{k+1}\|u\|_1$ at iteration $k+1$; then, the subproblem (3.2) becomes

$$\text{Version 1:} \quad u^{k+1} \leftarrow \min_u J^{k+1}(u) - \frac{\mu^{k+1}}{\mu^k}\langle p^k, u\rangle + \frac{1}{2}\|Au - f\|^2,$$

where we replace the Bregman distance of J in (3.2) by that of J^{k+1} between u and u^k, so

$$p^{k+1} = \frac{\mu^{k+1}}{\mu^k}p^k - A^\top(Au^{k+1} - f) = -\mu^{k+1}\sum_{j=1}^{k+1}\frac{A^\top(Au^j - f)}{\mu^j}.$$

Using the above identity, the subproblem (3.6), equivalent to (3.2), becomes

$$\text{Version 2:} \quad u^{k+1} \leftarrow \min_u J^{k+1}(u) + \frac{1}{2}\left\|Au - f^{k+1}\right\|^2,$$

for

$$f^{k+1} = f + \mu^{k+1}\sum_{j=1}^{k}\frac{f - Au^j}{\mu^j},$$

or

$$f^{k+1} = f + \frac{\mu^{k+1}}{\mu^k}(f^k - Au^k),\ u^0 = f^0 = \mathbf{0}.$$

We plan to explore varying step sizes to improve the efficiency of our code.

3.4. Equivalence to the augmented Lagrangian method. After we initially submitted this paper, we found that the Bregman iterative method Algorithm 1 is equivalent to the well-know augmented Lagrangian method (a.k.a., the method of multipliers), which was introduced by Hestenes [58] and Powell [73] and was later generalized by Rockafellar [77].

To solve the constrained optimization problem

$$\min_u s(u),\ \text{subject to } c_i(u) = 0,\ i = 1, \dots, m, \tag{3.22}$$

the augmented Lagrangian method minimizes the augmented Lagrangian function

$$L(u; \lambda^k, \nu) := s(u) + \sum_{i=1}^m \lambda_i^k c_i(u) + \frac{1}{2}\sum_{i=1}^m \nu_i c_i^2(u) \tag{3.23}$$

with respect to u at each iteration k, and uses the minimizer u^{k+1} to update

$$\lambda_i^{k+1} \leftarrow \lambda_i^k + \nu_i c_i(u^{k+1}). \tag{3.24}$$

The equivalence between this method and Version 1, (3.1)–(3.3), can be seen by letting

$$\begin{aligned} s(u) &= J(u), \\ c = \begin{bmatrix} c_1 \\ \cdots \\ c_m \end{bmatrix} &= Ax - b, \\ p^k &= -A^\top \lambda^k, \\ \nu_i &\equiv 1,\ \forall i. \end{aligned}$$

Then, we have

$$\begin{aligned} L(u; \lambda^k, \nu) &= J(u) + \langle \lambda^k, Au \rangle + \frac{1}{2}\|Au - b\|^2 + C_1 \\ &= J(u) - \langle p^k, u \rangle + \frac{1}{2}\|Au - b\|^2 + C_1 \\ &= \text{the objective function of (3.2)} + C_2, \end{aligned}$$

where C_1 and C_2 are constant in u, and also (3.24) yields (3.3). Therefore, whenever $u^{\mathbf{0}} = 0$ and $\lambda^0 = \mathbf{0}$, the augmented Lagrangian method is equivalent to Version 1 (3.1)–(3.3). This inspires us to study and apply techniques and results of the augmented Lagrangian method to our problem in the future. Finally, we note that the Bregman iterative regularization is general not equivalent to the augmented Lagrangian method when the constraints are not in the form of $Au = f$.

4. Numerical Results. In this section, we demonstrate the effectiveness of Algorithm 1 for solving the Basis Pursuit problem (1.1), where the constraints $Ax = f$ are under-determined linear equations and f was generated from a sparse signal $\bar{u}$ that has $\|\bar{u}\|_0 << n$, where $\|\bar{u}\|_0$ is defined as the number of nonzeros in u.

Our numerical experiments used two types of A matrices: orthogonalized Gaussian matrices whose elements were generated from i.i.d. normal distributions $\mathcal{N}(0,1)$ (**randn(m,n)** in MATLAB) and whose rows were orthogonalized by QR decompositions, and partial discrete cosine transform (DCT) matrices whose m rows were chosen randomly from the $n \times n$ DCT matrix. These matrices are known to be efficient for compressed sensing in the sense of allowing a good compression ratio m/n with a high probability, and have been widely used by researchers in their numerical experiments. We orthogonalized the rows of A because the subproblem solver FPC tends to be more numerically stable with such As.

The tested *original sparse signals* $\bar{u}$ had numbers of nonzeros equal to $0.1m$ and $0.2m$ rounded to the nearest integers in two sets of experiments, which were obtained by **round(0.1*m)** and **round(0.2*m)** in MATLAB, respectively. Given a number of nonzeros $\|\bar{u}\|_0$, an *original sparse signal* $\bar{u} \in \mathbb{R}^n$ was generated by randomly selecting the locations of these nonzeros, and sampling each of these nonzero elements from $\mathcal{N}(0,4)$ (**2*randn** in MATLAB). Then, f was computed as $A\bar{u}$. When $\|\bar{u}\|_0$ is small enough, we expect the Basis Pursuit problem (1.1), which we solved using Algorithm 1, to yield a solution $u_{\text{opt}} = \bar{u}$ from the inputs A and f.

We used the fast MATLAB code FPC, basic version 1.0, to solve the unconstrained sub-problem (1.2) at each Bregman iteration. This basic version does **not** use any line search techniques to speed up convergence. The reader may use more recent versions of FPC or other solvers of (1.2) such that GPSR [46], ℓ_1_ℓ_s [62], and SPGL1 [88] to repeat the experiments.

While the full Gaussian matrices were explicitly stored in memory, the partial DCT matrices were implicitly stored as fast transforms for which matrix-vector multiplications of the form Ax and $A^\top x$ were computed by the MATLAB commands **dct(x)** and **idct(x)**, respectively. Therefore, we were able to test partial DCT matrices of much larger sizes than Gaussian matrices. The dimensions m-by-n of these matrices are given in the first two columns of Table 4.1.

Our code was written in MATLAB and was run on a Linux (version 2.6.9) workstation with a 1.8GHz AMD Opteron CPU and 3GB memory. The MATLAB version is 7.1.

The computational results given in Table 4.1 were obtained using the stopping tolerance

$$\frac{\|Au^k - f\|}{\|f\|} < 10^{-5}, \tag{4.1}$$

which was sufficient to give a small error $\|u^k - \bar{u}\|/\|\bar{u}\| < O(10^{-6})$. The total number of Bregman iterations and running time of Algorithm 1 heavily depends on μ. Throughout our experiments, we used

$$\mu = \frac{0.02}{\sqrt{\|\bar{u}\|_0}}, \tag{4.2}$$

because μ bounds the maximum residual $\|Au - f\|$ up to a constant factor according to the following: for any minimizer $u_{\text{opt}} \approx \bar{u}$ of the unconstrained subproblem (1.2), it holds that

$$\|Au_{\text{opt}} - f\|^2 = \|A^\top(Au_{\text{opt}} - b)\|^2 \lessapprox O(\|\bar{u}\|_0)\,\|A^\top(Au_{\text{opt}} - b)\|_\infty^2 \leq O(\|\bar{u}\|_0)\mu^2. \tag{4.3}$$

TABLE 4.1
Experiment results using 20 random instances for each configuration of $(n, m, \|\bar{u}\|_0)$.

Results of Algorithm 1										
Bregman stopping tol.	$\|Au^k - f\|/\|f\| < 10^{-5}$									
Sub-problem solver	FPC-basic, ver 1. Stopping tol: **xtol**$= 10^{-4}$, **gtol**$= 10^{-3}$									
		Orthogonalized Gaussian matrices								
		stopping itr. k			relative error $\|u^k - \bar{u}\|/\|\bar{u}\|$			time (sec.)		
		mean	std.	max	mean	std.	max	mean	std.	max
n	m	$\|\bar{u}\|_0/m = 0.1$								
512	256	2.0	0.0	2	2.16e-08	1.37e-08	4.90e-08	0.1	0.0	0.3
1024	512	2.0	0.2	3	2.42e-08	4.53e-08	1.56e-07	0.3	0.0	0.4
2048	1024	2.1	0.4	4	2.74e-07	1.21e-06	5.42e-06	1.1	0.2	2.1
4096	2048	2.2	0.9	6	3.45e-07	1.17e-06	5.11e-06	5.0	1.8	12.4
n	m	$\|\bar{u}\|_0/m = 0.2$								
512	256	2.6	2.5	13	6.11e-07	2.48e-06	1.11e-05	0.2	0.2	0.8
1024	512	2.0	0.2	3	7.48e-08	1.08e-07	4.19e-07	0.5	0.1	0.7
2048	1024	2.5	1.8	10	7.51e-07	2.26e-06	8.93e-06	2.7	1.9	10.4
4096	2048	2.2	0.4	3	7.85e-08	3.26e-07	1.46e-06	10.4	2.1	14.4
		Partial DCT matrices								
n	m	$\|\bar{u}\|_0/m = 0.1$								
1024	512	2.3	1.0	6	9.80e-07	2.99e-06	1.09e-05	0.1	0.0	0.2
4096	2048	2.4	0.7	4	3.57e-08	1.40e-07	6.30e-07	0.4	0.1	0.6
32768	16384	2.0	0.0	2	2.06e-06	2.83e-06	9.19e-06	4.7	0.4	6.3
1048576	524288	2.0	0.0	2	2.33e-07	1.05e-07	3.69e-07	205.4	1.5	208.6
n	m	$\|\bar{u}\|_0/m = 0.2$								
1024	512	2.4	1.1	7	6.16e-08	9.19e-08	4.00e-07	0.2	0.1	0.4
4096	2048	2.5	0.6	4	1.06e-06	3.16e-06	1.04e-05	0.8	0.2	1.2
32768	16384	2.0	0.0	2	1.14e-06	2.10e-06	7.98e-06	8.5	0.2	8.9
1048576	524288	2.0	0.0	2	1.91e-07	7.00e-08	2.97e-07	423.6	2.0	428.3

The first equality in (4.3) follows from $AA^\top = I$ for all orthogonalized Gaussian and partial DCT matrices, the second approximate inequality "$\lesssim$" is an improved estimate over the inequality $\|\cdot\| \leq \sqrt{n}\|\cdot\|_\infty$ in $\mathbb{R}^n$, and the last inequality follows from the optimality of u to (1.2).

For dense Gaussian matrices A, our code was able to solve large-scale problem instances with more than 8 million nonzeros in A (e.g., $n \times m = 4096 \times 2048 = 2^{23} > 8 \times 10^6$) in 11 seconds on average over 20 random instances. For implicit partial DCT matrices A, our code was able to handle problems with matrices of dimension $2^{19} \times 2^{20}$ in less than eight minutes.

It is easy to see that the solver FPC was itself very efficient at solving the subproblem (1.2) for the assigned values of μ in (4.2). However, to yield solutions by a single call to FPC with errors as small as those produced by the Bregman iterations, one needs to use a much smaller value of μ. We tried straight FPC on the same set of test problems using values of μ that are 100 times smaller than those used in the Bregman procedure. This produced solutions with relative errors that were more than 10 times larger, while requiring longer running times. However, we *cannot* conclude that the Bregman iterative procedure accelerates FPC, since the best set of parameters for FPC to run with a tiny μ-value can be very different from those for a normal μ-value, but they are not known to us.

What is interesting is that Bregman iterations yield very accurate solutions even if the sub-problems are *not* solved as accurately. In other words, our approach can tolerate errors in p^k and u^k to a certain extent.

To see this, notice that the stopping tolerances **xtol** (relative error between two subsequence inner iterates) and **gtol** (violation of optimality conditions) for the subproblem solver FPC are much larger (see Table 4.1 for their values) than the relative errors of the final Bregman solutions. The reason for this remains a subject of further study.

Finally, to compare the Bregman iterative procedure based on the solver FPC with other recent ℓ_1 algorithms such as StOMP, one can refer to the CPU times of FPC in the comparative study [55] and multiplying these times by the average numbers of Bregman iterations.

5. Extensions. In this section we present extensions of our results in Section 3 to more general convex functionals $J(\cdot)$ and $H(\cdot)$, and describe a linearized Bregman iterative regularization scheme.

5.1. Finite convergence. Let $J(\cdot)$ and $H(\cdot)$ denote two convex functionals defined on $\mathcal{H}$, a Hilbert space. Moreover, we assume that there exists a $\tilde{u} \in \mathcal{H}$ that minimizes $H(\cdot)$ such that $J(\tilde{u}) < \infty$. Consider the minimization problem

$$\min_{u \in \mathcal{H}} J(u) + H(u). \tag{5.1}$$

Below we study the iterates $\{u^k\}$ of the Bregman iterative procedure Algorithm 1 applied to (5.1) assuming that a solution always exists in Step 3.

THEOREM 5.1. *Let $J(u)$ be convex and $H(u) = h(Au-f)$, for some nonnegative convex and differentiable function $h(\cdot)$ that only vanishes at* $\mathbf{0}$*; and assume that $J(\cdot)$ and $H(\cdot)$ satisfy above assumptions. In* a finite number of iterations, *Algorithm 1 returns a solution of*

$$\min_u \{J(u) : Au = f\} \tag{5.2}$$

under the following conditions: There exists a collection $\mathcal{U} = \{U^j\}$ satisfying:

1. *$\mathcal{H} = \cup_{U \in \mathcal{U}} U$;*
   Define $H^j := \min_u \{H(u) : u \in U^j\}$;
2. *If $H^j = 0$, then a minimum of $\{H(u) : u \in U^j\}$ can be attained;*
3. *$\{U^j \in \mathcal{U} : H^j > 0\}$ is a finite sub-collection;*
4. *There exists a rule to associate each u^k with a $U^{j_k} \ni u^k$ from $\mathcal{U}$ so that if $H^{j_k} = 0$, then $D_J^{p^k}(u, u^k) = 0 \Leftrightarrow u \in U^{j_k}$.*

Theorem 5.1 generalizes Theorems 3.2 and 3.3 in Section 3; therefore, its proof is similar to those of Theorems 3.2 and 3.3, and the only differences are:

1. In the generalized case, $\nabla H(u^k) = A^* \nabla h(Au^k - f)$ and $p^k = A^* \sum_{i=1}^k \nabla h(Au^i - f)$.
2. Equations (3.15)–(3.18) need to be replaced by

$$\begin{aligned} J(u^k) &\le J(u) - \langle u - u^k, p^k \rangle \\ &= J(u) - \langle u - u^k, A^* \sum_{i=1}^k \nabla h(Au^i - f) \rangle \\ &= J(u) - \langle Au - Au^k, \sum_{i=1}^k \nabla h(Au^i - f) \rangle \\ &= J(u) - \langle Au - f, \sum_{i=1}^k \nabla h(Au^i - f) \rangle. \end{aligned}$$

3. The last condition in Theorem 5.1 generalizes (3.21) and (3.19).

5.2. Bregman Iterations for Strictly Convex Functions. We now consider a strictly convex $J(u) \in C^2(\Omega)$, for a compact set $\Omega \subset \mathbb{R}^n$, without the homogeneity assumption (see (5.4) below) that we previously used; we also assume that the sequence $\{u^k\}$ lies in Ω. We have a unique element

$$p(u) \in \partial J(u) \tag{5.3}$$

for all u, and, in general,

$$J(u) \not\equiv \langle u, p(u) \rangle. \tag{5.4}$$

Again we wish to solve the constrained minimization problem:

$$\min_u \{J(u) : Au = f\}. \tag{5.5}$$

Our procedure will be, as before, the Bregman iterations (3.2) or (3.6).

THEOREM 5.2. *If $\|A^\top u\| \geq \delta \|u\|$, $\delta > 0$, then $\|Au^k - f\|$ decays exponentially to zero with k, and $w = \lim_{k\to\infty} u^k$ solves (5.5).*

Proof. Following equation (3.7), we have

$$p^{k+1} - p^k + A^\top (Au^{k+1} - f) = 0. \tag{5.6}$$

By the strict convexity of $J \in C^2(\Omega)$ and the compactness of Ω, there exist $\epsilon > 0$, independent of k, and a positive definite matrix unit: $Q_{k+\frac{1}{2}}$ with $\varepsilon I \prec Q_{k+\frac{1}{2}} \prec \frac{1}{\varepsilon} I$, i.e., both $Q_{k+\frac{1}{2}} - \varepsilon I$ and $\frac{1}{\varepsilon} I - Q_{k+\frac{1}{2}}$ are strictly positive definite, for some $\varepsilon > 0$ with

$$p^{k+1} - p^k = Q_{k+\frac{1}{2}} (u^{k+1} - u^k). \tag{5.7}$$

This leads us to

$$u^{k+1} - u^k + Q^{-1}_{k+\frac{1}{2}} A^\top (Au^{k+1} - f) = 0, \tag{5.8}$$

or

$$(I + AQ^{-1}_{k+\frac{1}{2}} A^\top)(Au^{k+1} - f) = (Au^k - f), \tag{5.9}$$

or

$$\begin{aligned} (Au^{k+1} - f) &= (I + AQ^{-1}_{k+\frac{1}{2}} A^\top)^{-1} (Au^k - f) \\ &= -\prod_{j=0}^{k} (I + AQ^{-1}_{j+\frac{1}{2}} A^\top)^{-1} f, \end{aligned} \tag{5.10}$$

and hence

$$\|Au^k - f\| \leq \left(\frac{1}{1 + \epsilon \delta^2} \right)^k \|f\|. \tag{5.11}$$

By the nonnegativity of the Bregman distance, we have, letting $\tilde{u}$ satisfy $A\tilde{u} = f$:

$$\begin{aligned} J(u^k) &\le J(\tilde{u}) - \langle \tilde{u} - u^k, p^k \rangle \\ &= J(\tilde{u}) + \langle \tilde{u} - u^k, \sum_{j=0}^{k} A^\top (Au^j - f) \rangle \\ &= J(\tilde{u}) - \langle Au^k - f, \sum_{j=0}^{k} Au^j - f \rangle \end{aligned} \tag{5.12}$$

so, by (5.11), taking the limit as $k \to \infty$, we obtain

$$J(w) \le J(\tilde{u}) \text{ with } Aw = f. \tag{5.13}$$

□

5.3. Linearized Bregman Iterations. In [25], Darbon and Osher combined the fixed-point iterations (2.2) with Bregman iterations for solving the image deblurring/deconvolution problem:

$$\min_u \{TV(u) : Au = f\}. \tag{5.14}$$

Let $J(u) = \mu TV(u)$. Their Bregman iterations are

$$u^{k+1} \leftarrow D_J^{p^k}(u, u^k) + \frac{1}{2\delta} \left\| u - \left(u^k - \delta A^\top (Au^k - f) \right) \right\|^2, \; k = 1, 2, \ldots, \tag{5.15}$$

which are different from the fixed-point iterations (2.2) by the use of regularization. While (2.2) minimizes J, (5.15) minimizes the Bregman distance $D_J^{p^k}$ based on J. On the other hand, (5.15) differs from (3.2) by replacing the fidelity term $\|Au - f\|^2/2$ by the sum of its first-order approximation at u^k and an ℓ_2-proximity term at u^k, which are the last three terms in (2.5). This sum is identical to a constant plus the last term in (5.15).

The sequence $\{p^k\}$ in (5.15) is chosen iteratively according the optimality conditions for (5.15):

$$\mathbf{0} = p^{k+1} - p^k + \frac{1}{\delta} \left(u^{k+1} - \left(u^k - \delta A^\top (Au^k - f) \right) \right), \tag{5.16}$$

so each p^{k+1} is uniquely determined from p^k, u^k, and u^{k+1} at the end of iteration k. By noticing that $p^0 = \mathbf{0}$ and $u^0 = \mathbf{0}$, we obtain from (5.16) that

$$p^{k+1} = p^k - A^\top (Au^k - f) - \frac{(u^{k+1} - u^k)}{\delta} = \cdots = \sum_{j=0}^{k} A^\top (f - Au^j) - \frac{u^{k+1}}{\delta}. \tag{5.17}$$

Therefore, $\{p^k\}$ can be computed on the fly. In addition, iterating (5.15) is very simple because it is a component-wise separable problem.

Motivated by Basis Pursuit, we consider the case for which $J(u) = \mu \|u\|_1$. Then, letting

$$v^k = \sum_{j=0}^{k} A^\top (f - Au^j), \tag{5.18}$$

each linearized Bregman iteration (5.15) after rearrangement yields:

$$u_i^{k+1} \leftarrow \delta\,\mathrm{shrink}(v_i^k, \mu),\ \ i = 1, \dots, n, \tag{5.19}$$

$$v^{k+1} \leftarrow v^k + A^\top(f - Au^{k+1}). \tag{5.20}$$

This is an extremely fast algorithm, very simple to program, involving only matrix multiplication and scalar shrinkage.

In [32] we will discuss this method in detail, both in terms of its convergence and speed of execution. We have the following key results under the assumption that $J \in C^2$ is strictly convex over a compact set $\Omega \supset \{u^k\}$ (although $\mu TV(\cdot)$ is not strictly convex, it can be approximated by the strictly convex perturbed functional $\mu \int \sqrt{|\nabla u|^2 + \varepsilon}$ for $\varepsilon > 0$).

THEOREM 5.3. *Let J be strictly convex and u_{opt} be an optimal solution of* $\min\{J(u) : Au = f\}$. *Then if $u^k \to w$, we have*

$$J(w) \le J(u_{\mathrm{opt}}) + \frac{1}{\delta}\langle w, u_{\mathrm{opt}} - w\rangle \tag{5.21}$$

and $\|Au^k - f\|$ decays exponentially in k if

$$I - \frac{\delta}{2}AA^\top \tag{5.22}$$

is strictly positive definite.

Proof. We have

$$u^{k+1} = \arg\min J(u) - J(u^k) - \langle u - u^k, p^k\rangle + \frac{1}{2\delta}\left\|u - u^k + \delta A^\top(Au^k - f)\right\|^2 \tag{5.23}$$

which, by (5.17) and (5.18), becomes

$$\begin{aligned} u^{k+1} = \arg\min\ & J(u) - J(u^k) - \langle u - u^k, v^{k-1}\rangle \\ & + \frac{1}{\delta}\langle u - u^k, u^k\rangle + \frac{1}{2\delta}\left\|u - u^k + \delta A^\top(Au^k - f)\right\|^2. \end{aligned}$$

By nonnegativity of the Bregman distance, we have

$$\begin{aligned} J(u^k) &\le J(u_{\mathrm{opt}}) - \langle u_{\mathrm{opt}} - u^k, p^k\rangle \\ &= J(u_{\mathrm{opt}}) - \langle u_{\mathrm{opt}} - u^k, v^{k-1}\rangle + \frac{1}{\delta}\langle u_{\mathrm{opt}} - u^k, u^k\rangle \\ &= J(u_{\mathrm{opt}}) - \langle f - Au^k, \sum_{j=0}^{k-1}(f - Au^j)\rangle + \frac{1}{\delta}\langle u_{\mathrm{opt}} - u^k, u^k\rangle. \end{aligned}$$

We will show that $\|f - Au^j\|$ decays exponentially with j; then the middle term in the last right-hand side above will vanish and the results follow.

We have

$$p^{k+1} - p^k + \frac{1}{\delta}(u^{k+1} - u^k) = -A^\top(Au^k - f). \tag{5.24}$$

By the strict convexity of J, there exists a symmetric positive definite operator $Q_{k+\frac{1}{2}}$ so that

$$\left(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I\right)(u^{k+1} - u^k) = -A^\top(Au^k - f). \tag{5.25}$$

Then

$$u^{k+1} - u^k = -\left(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I\right)^{-1} A^\top(Au^k - f) \tag{5.26}$$

and

$$Au^{k+1} - f = \left(I - A\left(Q_{k+\frac{1}{2}} + \frac{1}{\delta}I\right)^{-1} A^\top\right)(Au^k - f). \tag{5.27}$$

To have exponential decay in k of $\|Au^k - f\|$, we need the maximum eigenvalue of $(I - A\left(Q_{k+\frac{1}{2}} + \frac{1}{\delta}\right)^{-1} A^\top)$ to be strictly less than 1, or equivalently, the minimum and maximum eigenvalues of $A(Q_{k+\frac{1}{2}} + \frac{1}{\delta})^{-1}A^\top$ to be strictly positive and less than 2, respectively. The former requirement follows from the strict positive definiteness of $A(Q_{k+\frac{1}{2}} + \frac{1}{\delta})^{-1}A^\top$. To have the latter, we note that $A(Q_{k+\frac{1}{2}} + \frac{1}{\delta})^{-1}A^\top \prec \delta AA^\top$, which follows from the Sherman-Morrison-Woodbury formula; hence, it suffices to have $\delta AA^\top \prec 2I$, i.e., (5.22). ◻

We comment that the rate of exponential decay in k of $\|Au^k - f\|$ depends on the value of μ even when δ satisfies the condition $I - \frac{\delta}{2}AA^\top \succ 0$. Since $Q_{k+\frac{1}{2}}$ is linear in μ (assuming u^k and u^{k+1} are fixed), we let $Q_{k+\frac{1}{2}} = \mu\overline{Q}_{k+\frac{1}{2}}$. When μ is much larger than δ, $(Q_{k+\frac{1}{2}} + \frac{1}{\delta})^{-1}$ is dominated by $\frac{1}{\mu}\overline{Q}_{k+\frac{1}{2}}^{-1}$ so the minimum eigenvalue of $A(Q_{k+\frac{1}{2}} + \frac{1}{\delta})^{-1}A^\top$ diminishes linearly to 0 in μ.

If we let μ be very large, then w approaches a minimizer of $\|u\|_1$ subject to $Au = f$. We also have a simple estimate from [25].

Theorem 5.4. *If $\delta A^\top A < I$, then*

$$\|Au^{k+1} - f\|^2 + \left(\frac{1}{\delta} - \|A^\top A\|\right)\|u^{k+1} - u^k\|^2 \le \|Au^k - f\|^2.$$

Proof. Since the Bregman distance used in (5.23) is nonnegative we have

$$\|u^{k+1} - u^k + \delta A^\top(Au^k - f)\|^2 \le \|\delta A^\top(Au^k - f)\|^2 \tag{5.28}$$

or

$$\|u^{k+1} - u^k\|^2 + 2\delta\langle u^{k+1} - u^k, A^\top(Au^k - f)\rangle \le 0 \tag{5.29}$$

or

$$\|u^{k+1} - u^k\|^2 + \delta\|Au^{k+1} - f\|^2 - \delta\langle u^{k+1} - u^k, A^\top A(u^{k+1} - u^k)\rangle \le \delta\|Au^k - f\|^2 \tag{5.30}$$

or

$$\|Au^{k+1} - f\|^2 + \left(\frac{1}{\delta} - \|A^\top A\|\right)\|u^{k+1} - u^k\|^2 \le \|Au^k - f\|^2. \tag{5.31}$$

◻

Finally we have a result which typifies the effectiveness of Bregman iteration in the presence of noisy data. Our argument below follows that of [2].

THEOREM 5.5. *Let $J(\tilde{u})$ and $\|\tilde{u}\|$ be finite and $I - 2\delta AA^\top$ be strictly positive definite. Then the generalized Bregman distance*

$$\tilde{D}_J^{p^k}(\tilde{u}, u^k) = J(\tilde{u}) - J(u^k) - \langle \tilde{u} - u^k, p^k \rangle + \frac{1}{2\delta}\|\tilde{u} - u^k\|^2$$

diminishes with increasing k as long as $\|A\tilde{u} - f\| < (1 - 2\delta\|AA^\top\|)\|Au^k - f\|$.

Proof. Using (5.24) and the fact that $\langle u^{k+1} - u^k, p^{k+1} - p^k \rangle \geq 0$, we have

$$\|p^{k+1} - p^k\|^2 \leq -\langle p^{k+1} - p^k, A^\top(Au^k - f)\rangle$$

so

$$\|p^{k+1} - p^k\| \leq \|A^\top(A^k u - f)\|. \tag{5.32}$$

Also, following [25], we have for any $\tilde{u}$ for which $J(\tilde{u})$, $\|\tilde{u}\|$ are both finite

$$\begin{aligned} D_J^{p^{k+1}}(\tilde{u}, u^{k+1}) - D_J^{p^k}(\tilde{u}, u^k) + \frac{1}{2\delta}\|\tilde{u} - u^{k+1}\|^2 - \frac{1}{2\delta}\|\tilde{u} - u^k\|^2 \\ \leq \langle p^{k+1} - p^k + \frac{1}{\delta}(u^{k+1} - u^k),\ u^{k+1} - u^k\rangle \\ + \langle p^{k+1} - p^k + \frac{1}{\delta}(u^{k+1} - u^k), u^k - \tilde{u}\rangle. \end{aligned} \tag{5.33}$$

Using (5.24), the first term on the right side of (5.33) equals

$$\delta\|A^\top(Au^k - f)\|^2 - \delta\langle A^\top(Au^k - f), p^{k+1} - p^k\rangle \leq 2\delta\|A^\top(Au^k - f)\|^2 \tag{5.34}$$

The second term on the right side of (5.33) equals

$$\langle -A^\top(Au^k - f), u^k - \tilde{u}\rangle = -\|Au^k - f\|^2 + \langle Au^k - f, A\tilde{u} - f\rangle. \tag{5.35}$$

Adding (5.34) and (5.35) gives us

$$\begin{aligned} D_J^{p^{k+1}}(\tilde{u}, u^{k+1}) - D_J^{p^k}(\tilde{u}, u^k) + \frac{1}{2\delta}\|\tilde{u} - u^{k+1}\|^2 - \frac{1}{\delta}\|\tilde{u} - u^k\|^2 \\ \leq 2\delta\|A^\top(Au^k - f)\|^2 - \|Au^k - f\|^2 + \langle Au^k - f, A\tilde{u} - f\rangle. \end{aligned} \tag{5.36}$$

This means that this generalized Bregman distance $\tilde{D}_J^{p^k}(u, u^k)$ between $\tilde{u}$ and u^k diminishes in k as long as

$$\|A\tilde{u} - f\| < (1 - 2\delta\|AA^\top\|)\|Au^k - f\| \tag{5.37}$$

i.e. as long as $\|Au^k - f\|$ is not too small compared with $\|A\tilde{u} - f\|$ for $\tilde{u}$, the "denoised" solution. Of course if $\tilde{u}$ is a solution of the Basis Pursuit problem, then this generalized Bregman distance monotonically decreases in k. ☐

6. Conclusion. For solving the Basis Pursuit problem (1.1), which is traditionally formulated and solved as a constrained linear program, we show that a simple Bregman iterative scheme applied to its unconstrained Lagrangian relaxation (1.2) yields an exact solution in a finite number of iterations. Using a moderate value of the relaxation penalty parameter μ, very few iterations are required for most problem instances.

Solving (1.1) via (1.2) enables one to use recently developed fast codes designed to solve (1.2) that require only matrix-vector products and thus take advantage of available fast transforms. As a result, we are able to solve huge compressed sensing problems on a standard PC to high accuracies.

Our discovery that certain types of constrained problems can be exactly solved by iteratively solving a sequence of unconstrained subproblems generated by a Bregman iterative regularization scheme is new. We extend this result in several ways. One yields even simpler iterations (5.19) and (5.20). We hope that our discovery and its extensions will lead to efficient algorithms for even broader classes of problems.

Acknowledgement. We want to thank Bin Dong (UCLA), Elaine Hale (Rice), Yu Mao (UCLA), and Yin Zhang (Rice) for their helpful discussions and comments. We would like to acknowledge the anonymous referees for reviewing this article and making helpful suggestions.

REFERENCES

[1] T. Adeyemi and M. Davies, *Sparse representations of images using overcomplete complex wavelets*, 2005 IEEE/SP 13th Workshop on Statistical Signal Processing, (2005).

[2] M. Bachmayr, *Iterative total variation methods for nonlinear inverse problems*, Master thesis, Johannes Kepler Universität, Linz, Austria, (2007).

[3] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, *Compressive wireless sensing*, Proc. Int. Conf. on Information Processing in Sensor Networks (IPSN), Nashville, Tennessee, (2006).

[4] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk, *Distributed compressed sensing*, Preprint, (2005).

[5] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle, *A ℓ_1-unified variational framework for image restoration*, European Conference on Computer Vision, Prague, Lecture Notes in Computer Sciences 3024, (2004), pp. 1–13.

[6] J. Bioucas-Dias, *Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors*, IEEE Transactions on Image Processing, 15 (2006), pp. 937–951.

[7] J. Bioucas-Dias and M. Figueiredo, *Two-step algorithms for linear inverse problems with non-quadratic regularization*, IEEE International Conference on Image Processing ICIP' 2007, San Antonio, TX, (2007).

[8] T. Blumensath and M. Davies, *Iterative thresholding for sparse approximations*, Preprint, (2007).

[9] L. Bregman, *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming*, USSR Computational Mathematics and Mathematical Physics, 7 (1967), pp. 200–217.

[10] M. Burger, G. Gilboa, S. Osher, and J. Xu, *Nonlinear inverse scale space methods*, Communications in Mathematical Sciences, 4 (2006), pp. 175–208.

[11] M. Burger, S. Osher, J. Xu, and G. Gilboa, *Nonlinear inverse scale space methods for image restoration*, Lecture Notes in Computer Science, 3752 (2005), pp. 25–36.

[12] E. Candès and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Foundations of Computational Mathematics, 6 (2006), pp. 227–254.

[13] E. Candès and J. Romberg, *Sparsity and incoherence in compressive sampling*, Inverse Problems, 23 (2007), pp. 969–985.

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

[15] E. Candès and T. Tao, *Decoding by linear programming*, IEEE Transactions on Information Theory, 51 (2005), pp. 4203–4215.

[16] ———, *Near optimal signal recovery from random projections: universal encoding strategies*, IEEE Transactions on Information Theory, 52 (2006), pp. 5406–5425.

[17] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97.

[18] ———, *Total variation minimization and a class of binary MRF models*, Tech. Rep. UMR CNRS 7641, Ecole Polytechnique, 2005.

[19] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Transactions on Image Processing, 7 (1998), pp. 319–335.

[20] R. Chartrand, *Exact reconstructions of sparse signals via nonconvex minimization*, IEEE Signal Processing Letters, to appear, (2007).

[21] ———, *Nonconvex compressed sensing and error correction*, Proceedings of the ICASSP, (2007).

[22] R. Chartrand and W. Yin, *Iteratively reweighted algorithms for compressive sensing*, Submitted to ICASSP'08, (2007).

[23] S. Chen, D. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Journal on Scientific Computing, 20 (1998), pp. 33–61.

[24] P. L. Combettes and J.-C. Pesquet, *Proximal thresholding algorithm for minimization over orthonormal bases*, To appear in SIAM Journal on Optimization, (2007).

[25] J. Darbon and S. Osher, *Fast discrete optimization for sparse approximations and deconvolutions*, Preprint, (2007).

[26] J. Darbon and M. Sigelle, *Image restoration with discrete constrained total variation, Part I: fast and exact optimization*, Journal of Mathematical Imaging and Vision, 26 (2006), pp. 261–276.

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

[28] I. Daubechies, M. Fornasier, and I. Loris, *Accelerated projected gradient method for linear inverse problems with sparsity constraints*, arXiv:0706:4297, (2007).

[29] C. De Mol and M. Defrise, *A note on wavelet-based inversion algorithms*, Contemporary Mathematics, 313 (2002), pp. 85–96.

[30] R. A. DeVore, *Deterministic constructions of compressed sensing matrices*, Submitted, (2007).

[31] R. A. DeVore and B. J. Lucier, *Wavelets*, Acta Numerica, 1 (1992), pp. 54–81.

[32] B. Dong, Y. Mao, S. Osher, and W. Yin, *Fast linearized Bregman iteration for compressed sensing and related problems*, working paper, (2007).

[33] D. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, 52 (2006), pp. 1289–1306.

[34] D. Donoho and J. Tanner, *Neighborliness of randomly-projected simplices in high dimensions*, Proc. National Academy of Sciences, 102 (2005), pp. 9452–9457.

[35] D. Donoho, Y. Tsaig, I. Drori, and J.-C. Starck, *Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit*, Submitted to IEEE Transactions on Information Theory, (2006).

[36] M. Elad, *Why simple shrinkage is still relevant for redundant representations?*, IEEE Transactions on Information Theory, 52 (2006), pp. 5559–5569.

[37] M. Elad and A. Bruckstein, *A generalized uncertainty principle and sparse representations in pairs of bases*, IEEE Transactions on Information Theory, 48 (2002), pp. 2558–2567.

[38] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, *A wide-angle view at iterated shrinkage algorithms*, SPIE (Wavelet XII), San-Diego CA, August 26-29, 2007., (2007).

[39] M. Elad, B. Matalon, and M. Zibulevsky, *Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization*, Journal on Applied and Computational Harmonic Analysis, (2006).

[40] M. Elad, J.-C. Starck, P. Querre, and D. Donoho, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Journal on Applied and Computational Harmonic Analysis, 19 (2005).

[41] M. J. Fadili and J. L. Starck, *Sparse representation-based image deconvolution by iterative thresholding*, Astronomical Data Analysis ADA'06, Marseille, France, (2006).

[42] A. Feuer and A. Nemirovski, *On sparse representation in pairs of bases*, IEEE Transactions on Information Theory, 49 (2003), pp. 1579–1581.

[43] M. Figueiredo, J. Bioucas-Dias, and R. Nowak, *Majorization-minimization algorithms for wavelet-based image restoration*, IEEE Transactions on Image Processing, 16 (2007), pp. 2980–2991.

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

[45] ———, *A bound optimization approach to wavelet-based image deconvolution*, Proceedings of the IEEE International Conference on Image Processing (ICIP), (2005).

[46] M. Figueiredo, R. Nowak, and S. J. Wright, *GPSR: Gradient projection for sparse reconstruction, http://www.lx.it.pt/~mtf/gpsr/*, (2007).

[47] ———, *Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems*, To appear in the IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing, (2007).

[48] M. Fornasier and H. Rauhut, *Recovery algorithms for vector valued data with joint sparsity constraints*, Preprint, (2006).

[49] L. Gan, *Block compressed sensing of natural images*, Proc. Int. Conf. on Digital Signal Processing (DSP), Cardiff, UK, (2007).

[50] H.-Y. Gao and A. G. Bruce, *WaveShrink with firm shrinkage*, Statistica Sinica, 7 (1997), pp. 855–874.

[51] D. Goldfarb and W. Yin, *Parametric maximum flow algorithms for fast total variation minimization*, Submitted, (2007).

[52] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, *Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms.* `http://hal.inria.fr/inria-00146660`, (2007).

[53] E. Hale, W. Yin, and Y. Zhang, *A fixed-point continuation method for ℓ_1-regularization with application to compressed sensing*, Rice University CAAM Technical Report TR07-07, (2007).

[54] ———, *FPC: A fixed-point continuation method for ℓ_1-regularization,* `http://www.caam.rice.edu/~optimization/l1`, (2007).

[55] ———, *A numerical study of fixed-point continuation applied to compressed sensing*, To be submitted to IEEE Transactions on Information Theory, (2007).

[56] E. T. Hale, W. Yin, and Y. Zhang, *Fixed-point continuation for ℓ_1-minimization: methodology and convergence*, Submitted to SIAM Journal on Optimization, (2007).

[57] L. He, T.-C. Chang, S. Osher, T. Fang, and P. Speier, *MR image reconstruction by using the iterative refinement method and nonlinear inverse scale space methods*, UCLA CAM Report 06-35, (2006).

[58] M. R. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications, 4 (1969), pp. 303–320.

[59] H. Jung, J. Ye, and E. Kim, *Improved k-t blask and k-t sense using focuss*, Phys. Med. Biol., 52 (2007), pp. 2101–3226.

[60] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, *A method for large-scale ℓ_1-regularized least squares problems with applications in signal processing and statistics,* `http://www.stanford.edu/~boyd/l1_ls.html`, (2007).

[61] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk, *Analog-to-information conversion via random demodulation*, in Proceedings of the IEEE Dallas Circuits and Systems Workshop (DCAS), Dallas, Texas, 2006.

[62] K. Koh, S.-J. Kim, and S. Boyd, *ℓ_1_ℓ_s: A simple MATLAB solver for ℓ_1-regularized least squares problems,* `http://www.stanford.edu/~boyd/l1_ls/`, (2007).

[63] J. Laska, S. Kirolos, M. Duarte, T. Ragheb, R. Baraniuk, and Y. Massoud, *Theory and implementation of an analog-to information converter using random demodulation*, in Proceedings of the IEEE International Symposium on Circuites and Systems (ISCAS), New Orleans, Louisiana, 2007.

[64] J. Laska, S. Kirolos, Y. Massoud, R. Baraniuk, A. Gilbert, M. Iwen, and M. Strauss, *Random sampling for analog-to-information conversion of wideband signals*, in Proceedings of the IEEE Dallas Circuits and Systems Workshop, Dallas, Texas, 2006.

[65] Z.-Q. Luo and P. Tseng, *On the linear convergence of descent methods for convex essentially smooth minimization*, SIAM Journal on Control and Optimization, 30 (1990), pp. 408–425.

[66] M. Lustig, D. Donoho, and J. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Preprint, (2007).

[67] M. Lustig, J. Santos, D. Donoho, and J. Pauly, *k-t sparse: High frame rate dynamic MRI exploiting spatio-temporal sparsity*, Proc. 14th. Annual Meeting of ISMRM, Seattle, Washington, (2006).

[68] M. Lustig, J. Santos, J.-H. Lee, D. Donoho, and J. Pauly, *Application of compressed sensing for rapid MR imaging*, Preprint, (2007).

[69] Y. Nesterov, *Gradient methods for minimizing composite objective function*, `www.optimization-online.org`, CORE Discussion Paper 2007/76, (2007).

[70] R. Nowak and M. Figueiredo, *Fast wavelet-based image deconvolution using the EM algorithm*, Proceedings of the 35th Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, (2001).

[71] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterated regularization method for total variation based image restoration*, SIAM Journal on Multiscale Modeling and Simulation, 4 (2005), pp. 460–489.

[72] J.-S. Pang, *A posteriori error bounds for the linearly-constrained variational inequality problem*, Mathematical Methods of Operations Research, 12 (1987), pp. 474–484.

[73] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1972, pp. 283–298.

[74] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, *Decentralized compression and predistribution via randomized gossiping*,

Proc. Int. Conf. on Information Processing in Sensor Networks (IPSN), Nashville, Tennessee, (2006).

[75] T. Ragheb, S. Kirolos, J. Laska, A. Gilbert, M. Strauss, R. Baraniuk, and Y. Massoud, *Implementation models for analog-to-information conversion via random sampling*, Proc. Midwest Symposium on Circuits and Systems (MWSCAS), (2007).

[76] T. H. Reeves and N. G. Kingsbury, *Overcomplete image coding using iterative projection-based noise shaping*, 2002 International Conference on Image Processing, 3 (2002), pp. 24–28.

[77] R. T. Rockafellar, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Mathematical Programming, 5 (1973), pp. 354–373.

[78] M. Rudelson and R. Vershynin, *Geometric approach to error correcting codes and reconstruction of signals*, International Mathematical Research Notices, 64 (2005), pp. 4019–4041.

[79] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[80] I. Selesnick, R. Van Slyke, and O. Guleryuz, *Pixel recovery via ℓ_1 minimization in the wavelet domain*, 2004 International Conference on Image Processing, 3 (2004), pp. 1819–1822.

[81] M. Sheikh, O. Milenkovic, and R. Baraniuk, *Compressed sensing dna microarrays*, Rice ECE Department Technical Report TREE 0706, (2007).

[82] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk, *A new compressive imaging camera architecture using optical-domain compression*, in Proceedings of Computational Imaging IV at SPIE Electronic Image, San Jose, California, 2006.

[83] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal Royal Statistical Society B, 58 (1996), pp. 267–288.

[84] J. Tropp, *Just relax: Convex programming methods for identifying sparse signals*, IEEE Transactions on Information Theory, 51 (2006), pp. 1030–1051.

[85] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, *Random filters for compressive sampling and reconstruction*, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, 2006.

[86] Y. Tsaig and D. Donoho, *Extensions of compressed sensing*, Signal Processing, 86 (2005), pp. 533–548.

[87] E. Van den Berg and M. P. Friedlander, *In pursuit of a root*, UBC Computer Science Technical Report TR-2007-16, (2007).

[88] ———, *SPGL1: A MATLAB solver for large-scale sparse reconstruction,* `http://www.cs.ubc.ca/labs/scl/index.php/main/spgl1`, (2007).

[89] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, *An architecture for compressive image*, in Proceedings of the International Conference on Image Processing (ICIP), Atlanta, Georgia, 2006.

[90] ———, *Compressive imaging for video representation and coding*, in Proceedings of Picture Coding Symposium (PCS), Beijing, China, 2006.

[91] W. Wang, M. Garofalakis, and K. Ramchandran, *Distributed sparse random projections for refinable approximation*, Proc. Int. Conf. on Information Processing in Sensor Networks (IPSN), Nashville, Tennessee, (2007).

[92] Y. Wang, W. Yin, and Y. Zhang, *A fast fixed-point algorithm for convex total variation regularization*, Working paper, (2007).

[93] J. Xu and S. Osher, *Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising*, IEEE Transactions on Image Processing, 16 (2006), pp. 534–544.

[94] J. Ye, *Compressed sensing shape estimation of star-shaped objects in Fourier imaging*, Preprint, (2007).

[95] Y. Zhang, *A simple proof for recoverability of ℓ_1-minimization: go over or under?*, Rice University CAAM Technical Report TR05-09, (2005).

[96] ———, *When is missing data recoverable?*, Rice University CAAM Technical Report TR06-15, (2006).

[97] W. P. Ziemer, *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation*, Graduate Texts in Mathematics, Springer, 1989.