LINEARIZED BREGMAN ITERATIONS FOR COMPRESSED SENSING

JIAN-FENG CAI[*], STANLEY OSHER[†], AND ZUOWEI SHEN[‡]

Abstract. Finding a solution of a linear equation $Au = f$ with various minimization properties arises from many applications. One such application is compressed sensing, where an efficient and robust-to-noise algorithm to find a minimal ℓ_1 norm solution is needed. This means that the algorithm should be tailored for large scale and completely dense matrices A, while Au and $A^T u$ can be computed by fast transforms and the solution we seek is sparse. Recently, a simple and fast algorithm based on linearized Bregman iteration was proposed in [28, 32] for this purpose. This paper is to analyze the convergence of linearized Bregman iterations and the minimization properties of their limit. Based on our analysis here, we derive also a new algorithm that is proven to be convergent with a rate. Furthermore, the new algorithm is simple and fast in approximating a minimal ℓ_1 norm solution of $Au = f$ as shown by numerical simulations. Hence, it can be used as another choice of an efficient tool in compressed sensing.

AMS subject classifications. Primary 65K05, 65F22; Secondary 65T99.

1. Introduction. Let $A \in \mathbb{R}^{m \times n}$ with $n > m$ and $f \in \mathbb{R}^m$ be given. The aim of a basis pursuit problem is to find $u \in \mathbb{R}^n$ by solving the following constrained minimization problem

$$\min_{u \in \mathbb{R}^n} \{J(u) : Au = f\}, \tag{1.1}$$

where $J(u)$ is a continuous convex function. In other words, it is to find a solution of the linear system $Au = f$, which has a minimal energy $J(u)$. Throughout this paper, we assume that A is a surjective map, i.e., AA^T is invertible. Therefore, $Au = f$ is under-determined, and has at least one solution, i.e., $u = A^T(AA^T)^{-1}f$ which minimizes the ℓ_2 norm among all the solutions of the equation. The set of all solutions of $Au = f$ is convex. We assume also that $J(u)$ is coercive, i.e., whenever $\|u\| \to \infty$, $J(u) \to \infty$. This implies that the set of all solutions of (1.1) is a nonempty convex set. Furthermore, when $J(u)$ is strictly or strongly convex, the solution of (1.1) is unique.

This basis pursuit problem arises from many applications. For example, in a recent burst of research in compressed sensing, it amounts to solving (1.1) with J being the ℓ_1 norm to obtain a sparse solution of the equation. The interested reader should consult, for example, [2, 8, 9, 10, 11, 12, 19, 21, 22, 23, 24, 29, 30, 31, 33, 34] for details. The problem (1.1) can be transformed into a linear programming one, and then solved by a conventional linear programming solver in many cases. However, such solvers are not tailored for matrices A that are large scale and completely dense while the solution to find is sparse. It does not use, for example, the fact that matrices A are normally formed by rows of some orthonormal matrices corresponding to fast transforms either where Au and $A^T u$ can be computed by fast transforms. This fact

is, indeed, the case in some applications of compressed sensing. Hence, the algorithm should utilize the fact that the solution to seek is sparse and Au and $A^T u$ can be computed by fast transforms. More importantly, it should also be robust to noise.

1.1. Linearized Bregman Iteration. To solve (1.1) with emphasis on $J(u) = \|u\|_1$, a linearized Bregman iteration was proposed in [28, 32], which was inspired by the work in [16]. The idea of the linearized Bregman iteration is to combine a fixed point iteration and the Bregman iteration in [27, 32]. Given $p^0 = u^0 = 0$, the linearized Bregman iteration is generated by

$$\begin{cases} u^{k+1} = \arg\min_{u \in \mathbb{R}^n} \left\{ \mu(J(u) - J(u^k) - \langle u - u^k, p^k \rangle) + \frac{1}{2\delta} \|u - (u^k - \delta A^T(Au^k - f))\|^2 \right\}, \\ p^{k+1} = p^k - \frac{1}{\mu\delta}(u^{k+1} - u^k) - \frac{1}{\mu} A^T(Au^k - f), \end{cases}$$

(1.2)

where δ is a fixed step size.

This paper is to provide a complete analysis for the linearized Bregman iteration (1.2). In particular, we prove the convergence of (1.2) (both of the sequences $\{u^k\}_{k \in \mathbb{N}}$ and $\{p^k\}_{k \in \mathbb{N}}$) under the assumption that the convex function $J(u)$ is continuously differentiable and there exists a positive constant β such that

$$\|\partial J(u) - \partial J(v)\|^2 \le \beta \langle \partial J(u) - \partial J(v), u - v \rangle, \quad \forall u, v \in \mathbb{R}^n, \tag{1.3}$$

where $\partial J(u)$ is the gradient of $J(u)$. This implies that $\partial J(u)$ is Lipshitz continuous. Note that we also use $\partial J(u)$ to denote the subdifferential of the convex function $J(u)$ throughout the paper. Moreover, the limit of $\{u^k\}_{k \in \mathbb{N}}$ is the unique solution of

$$\min_{u \in \mathbb{R}^n} \left\{ \mu J(u) + \frac{1}{2\delta} \|u\|^2 : Au = f \right\}. \tag{1.4}$$

Though (1.4) is not the same as (1.1), we will show that (1.4) approximates (1.1) with a large μ.

1.2. Linearized Bregman Iteration for $J(u) = \|u\|_1$. When $J(u) = \|u\|_1$, after reformulation, algorithm (1.2) can be rewritten as

$$\begin{cases} v^{k+1} = v^k - A^T(Au^k - f), \\ u^{k+1} = T_{\mu\delta}(\delta v^{k+1}), \end{cases} \tag{1.5}$$

where $u^0 = v^0 = 0$, and

$$T_\lambda(w) := [t_\lambda(w(1)), t_\lambda(w(2)), \cdots, t_\lambda(w(n))]^T \tag{1.6}$$

is the soft thresholding operator [20] with

$$t_\lambda(\xi) = \begin{cases} 0, & \text{if } |\xi| \le \lambda, \\ \text{sgn}(\xi)(|\xi| - \lambda), & \text{if } |\xi| > \lambda. \end{cases} \tag{1.7}$$

This algorithm is shown to be accurate and efficient in many cases by numerical simulations in [28]. As shown in [28], algorithm (1.5) is an extremely fast algorithm that is very simple to program, involving only matrix-vector multiplications and scalar shrinkages. We remark here that it is not the first time in signal and image processing to combine an iterative algorithm together with a soft thresholding to obtain a solution that minimizes the ℓ_1 norm in some sense. In fact, one can find many of them in the

literature; see, e.g., [3, 4, 5, 6, 7, 13, 14, 15, 17, 18] for details. With all of these, we have good reason to believe that algorithm (1.5) works, as the numerical simulation of [28] has already illustrated.

Here, we would like to emphasize that algorithm (1.5) is robust to noise. In the first step of (1.5), it multiplies the error $f - Au^k$ by A^T and adds it into v^k to obtain v^{k+1}. This step does not blow up the noise. In the second step of (1.5), it thresholds δv^{k+1} by $T_{\mu\delta}$. This step produces a sparse vector u^{k+1} and removes the noise. In fact, if we choose a large μ (as we will see later, this is the case in both theory and practice), only large components in v^{k+1} are nonzeroes in u^{k+1}. This implies that u^{k+1} is a sparse vector, and the noise contained can be efficiently removed. Furthermore, since we start with $u^0 = 0$, algorithm (1.5) adds very important nonzero entries into u step by step. Altogether, it indicates that algorithm (1.5) is robust to noise. This means that, by properly choosing stopping criteria (e.g., stopping whenever $\|f - Au^k\|^2$ is less than the standard deviation of the noise), algorithm (1.5) can handle the case when f is very noisy. Indeed, even for the case that the noise dominates the signal in f, this algorithm gives an impressive result; see [28].

1.3. Application to Compressed Sensing. The original goal in [28, 32] of the linearized Bregman iteration is to solve the basis pursuit problem in compressed sensing, i.e.,

$$\min_{u \in \mathbb{R}^n} \{\|u\|_1 : Au = f\}. \tag{1.8}$$

Therefore, another theme of this paper is to apply the theory and analysis presented here to approximate a solution of (1.8). Hence, algorithms given here can be used as efficient tools for compressed sensing. Our strategy is as follows.

First, we use the solution of

$$\min_{u \in \mathbb{R}^n} \left\{ \mu\|u\|_1 + \frac{1}{2\delta}\|u\|^2 : Au = f \right\} \tag{1.9}$$

to approximate a solution of (1.8). We prove that, as $\mu \to \infty$, the solution of (1.9) tends to a solution of (1.8). The algorithm to find the solution of (1.9) is the linearized Bregman iteration (1.5). Since $\|u\|_1$ is not differentiable, we cannot prove the convergence of (1.5) by applying the theory here. However, we show that, when $\{u^k\}_{k \in \mathbb{N}}$ in (1.5) converge, the limit $u_{0,\mu}^*$ of $\{u^k\}_{k \in \mathbb{N}}$ is the solution of (1.9).

Second, in order to have a proven convergent algorithm, we further approximate the solution of (1.8) by the solution of

$$\min_{u \in \mathbb{R}^n} \left\{ \mu J_\epsilon(u) + \frac{1}{2\delta}\|u\|^2 : Au = f \right\} \tag{1.10}$$

by properly chosen J_ϵ that satisfies (1.3). In particular, we choose

$$J_\epsilon(u) = \sum_{i=1}^n F_\epsilon(u(i)), \quad \text{with} \quad F_\epsilon(\xi) = \begin{cases} \frac{\xi^2}{2\epsilon}, & \text{if } |\xi| \leq \epsilon, \\ |\xi| - \frac{\epsilon}{2}, & \text{if } |\xi| > \epsilon. \end{cases} \tag{1.11}$$

The function $J_\epsilon(u)$ is Moreau-Yosida C^1 regularization for $\|u\|_1$; see [25]. It is also known as Huber's norm in statistics [26]. The function J_ϵ is continuously differentiable and the gradient $\partial J_\epsilon(u)$ satisfies (1.3); see Figure 1.1. Since the function $\|u\|_1$, which is denoted by $J_0(u) = \|u\|_1$, can be approximated by J_ϵ, we prove that the solution
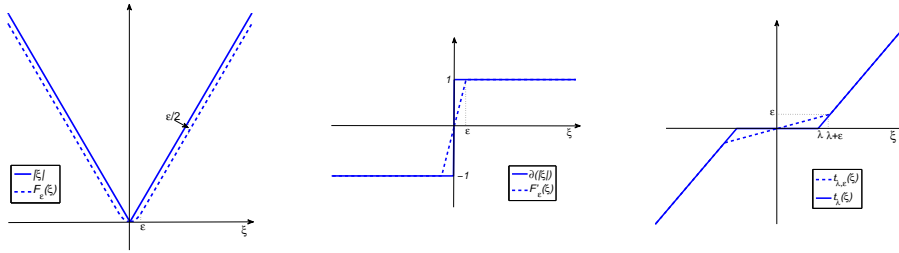
FIG. 1.1. *The left figure shows the approximation of F_ϵ to the absolute value function, the middle one shows the approximation of F'_ϵ to the subdifferential of the absolute value function, and the right one shows the approximation of $t_{\lambda,\epsilon}$ to the soft-thresholding function t_λ.*

of (1.10) can be as close to that of (1.9) as possible as long as $\mu\epsilon$ is chosen sufficiently small. This, in turn, reveals that, as $\mu \to \infty$ and $\mu\epsilon \to 0$, the solution of (1.10) converges to a solution of (1.8).

The algorithm for solving (1.10) is the linearized Bregman iteration (1.2) with $J = J_\epsilon$, which is guaranteed to be convergent by applying the convergence result of this paper. Furthermore, iteration (1.2) with $J = J_\epsilon$, which has a compact form as (1.5), can be reformulated as

$$\begin{cases} v^{k+1} = v^k - A^T(Au^k - f), \\ u^{k+1} = T_{\mu\delta,\epsilon}(\delta v^{k+1}), \end{cases} \tag{1.12}$$

where $u^0 = v^0 = 0$, and

$$T_{\lambda,\epsilon}(u) := [t_{\lambda,\epsilon}(u(1)), t_{\lambda,\epsilon}(u(2), \ldots, t_{\lambda,\epsilon}(u(n)]^T \tag{1.13}$$

with

$$t_{\lambda,\epsilon}(\xi) := \begin{cases} \frac{\epsilon}{\lambda+\epsilon}\xi & \text{if } |\xi| \le \lambda + \epsilon, \\ \text{sgn}(\xi)(|\xi| - \lambda) & \text{if } |\xi| > \lambda + \epsilon. \end{cases} \tag{1.14}$$

By comparing (1.14) with (1.7), we see that $T_{\lambda,\epsilon}$ in (1.12) is very simple to implement and very close to T_λ in (1.5), where $T_{\lambda,0} = T_\lambda$ as it is also illustrated by Figure 1.1. In fact, as we will see, $T_{\lambda,\epsilon}$ is essentially T_λ numerically, when ϵ is sufficiently small. Therefore, iteration (1.5) can be identified with iteration (1.12) numerically for sufficiently small ϵ. Furthermore, we will show that iteration (1.12) converges to the limit that can be a good approximation of a solution of (1.8) by choosing sufficiently large μ and small $\mu\epsilon$. Hence, it can be used in various applications in compressed sensing.

1.4. Organization. The paper is organized as follows. In Section 2, we reformulate the linearized Bregman iteration defined by (1.2) which helps us to understand the algorithm better. In Section 3, a complete analysis of (1.2) is presented, including the convergence of the iteration and minimization properties of its limit. The theory developed here is then used to derive algorithms approximating a solution of (1.8) in Section 4. It is shown that the derived algorithms are as effective as (1.5) to approximate a solution of (1.8).

2. Linearized Bregman Iteration. Before giving a full analysis of the convergence of the linearized Bregman iteration (1.2), we reformulate it into two different forms to better understand the algorithm.

We first interpret (1.2) as a one-step approximation of the Bregman iteration in [32]. Then, we show that the linearized Bregman iteration for $J(u)$ is, in fact, also a linearized Bregman iteration for another cost functional $J(u) + \frac{\tau}{2}\|u\|^2$ with different parameters $0 < \tau < \frac{1}{\mu\delta}$. This immediately leads to understanding why (1.2) converges to the minimizer of (1.4).

2.1. Reformulation. Here, we reformulate the linearized Bregman iteration (1.2) as a one-step approximation for the Bregman iteration in [32].

The Bregman iteration is defined in terms of the Bregman distance. Recall that the Bregman distance (see, e.g., [1] for details) based on the convex function J between points u and v is defined as

$$D_J^p(u, v) = J(u) - J(v) - \langle u - v, p \rangle,$$

where $p \in \partial J(v)$ is a 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. Moreover, whenever J is convex, $D_J^p(u, v) \geq 0$; whenever J is strictly convex, $D_J^p(u, v) > 0$ for $u \neq v$; and whenever J is strongly convex, there exists a constant $\nu > 0$ such that $D_J^p(u, v) \geq \nu\|u - v\|^2$.

To solve (1.1), the Bregman iteration is proposed in [32]. Given $u^0 = p^0 = 0$, define

$$\begin{cases} u^{k+1} = \arg\min_{u \in \mathbb{R}^n} \left\{ \mu(J(u) - J(u^k) - \langle u - u^k, p^k \rangle) + \frac{1}{2}\|Au - f\|^2 \right\}, \\ p^{k+1} = p^k - \frac{1}{\mu}A^T(Au^{k+1} - f). \end{cases} \quad (2.1)$$

Since, by induction, $p^k \in \partial J(u^k)$, the first equation can be written in terms of the Bregman distance as

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

It was proven in [32] that, if $J(u) \in C^2(\Omega)$ and $J(u)$ is strictly convex in a compact set Ω, then $\|Au^k - f\|$ decays exponentially in k whenever $u^k \in \Omega$ for all k. Furthermore, when the limit of u^k exists, it is a solution of (1.1). It was also proven in [32] that, when $J(u) = \|u\|_1$, or J is a convex function satisfying some additional conditions (see [32] for details), iteration (2.1) leads to a solution of (1.1) in finitely many steps.

As shown in [32], the Bregman iteration (2.1) can be written as

$$\begin{cases} f^{k+1} = f^k + (f - Au^k), \\ u^{k+1} = \arg\min_{u \in \mathbb{R}^n} \left\{ \mu J(u) + \frac{1}{2}\|Au - f^{k+1}\|^2 \right\}, \end{cases} \quad (2.2)$$

where $f^0 = 0$ and $u^0 = 0$. Indeed, by the second equation of (2.1), we have

$$-\mu p^k = A^T \sum_{i=1}^k (Au^i - f). \quad (2.3)$$

Let $E(u) = \mu(J(u) - J(u^k) - \langle u - u^k, p^k \rangle) + \frac{1}{2}\|Au - f\|^2$. Then, by (2.3),

$$\frac{\partial(E(u) - \mu J(u))}{\partial u} = -\mu p^k + A^T(Au - f) = A^T(Au - (f - \sum_{i=1}^k (Au^i - f)))$$

and

$$\frac{\partial^2(E(u) - J(u))}{\partial u^2} = A^T A.$$

Therefore, $E(u) - \mu J(u)$ is a quadratic function, and

$$E(u) = \mu J(u) + \frac{1}{2}\|Au - (f - \sum_{i=1}^{k}(Au^i - f))\|^2 + C,$$

where C is independent of u. By the definition of u^{k+1} in (2.1), we have that

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^n} E(u) = \arg\min_{u \in \mathbb{R}^n}\{\mu J(u) + \frac{1}{2}\|Au - (f - \sum_{i=1}^{k}(Au^i - f))\|^2\}. \quad (2.4)$$

Define $f^0 = 0$, $f^1 = f$ and $f^{k+1} = f - \sum_{i=1}^{k}(Au^i - f)$. Then

$$f^{k+1} = f^k + (f - Au^k), \quad f^0 = 0. \quad (2.5)$$

With this, (2.4) becomes

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^n}\{\mu J(u) + \frac{1}{2}\|Au - f^{k+1}\|^2\}, \quad (2.6)$$

which is (2.2). By (2.2), the Bregman iteration (2.1) is essentially using the solution of the Lagrange multiplier relaxed problem

$$\min_{u \in \mathbb{R}^n}\left\{\mu J(u) + \frac{1}{2}\|Au - f\|^2\right\} \quad (2.7)$$

as a solver to approximate the solution of (1.1). The Bregman iteration (2.2) applies this process iteratively. The first equation of (2.2) is to update the residues from the result of the solver, and the second equation of (2.2) is to derive a new approximation by solving (1.1) from the updated data.

Since, generally, there is no explicit expression for the solver (i.e., the second equation of (2.2) or the first equation in (2.1)), we have to turn to iterative methods. Therefore, we need an inner-outer iteration for (2.1), where the inner iteration is for solving the first equation of (2.1) at each step, i.e., (2.8), and the outer one is to solve (2.1). In the following, we show that the linearized Bregman iteration (1.2) is just an approximation of this inner-outer iteration, where only one step of the inner iteration is performed for each outer iteration.

The inner iteration used for solving

$$u^{k+1} = \arg\min_{u}\left\{\mu(J(u) - J(u^k) - \langle u - u^k, p^k\rangle) + \frac{1}{2}\|Au - f\|^2\right\} \quad (2.8)$$

is the proximal forward-backward splitting (PFBS) iteration of [15]. The goal of the PFBS iteration is to solve the minimization problem

$$\min_{x \in \mathbb{R}^n} F(x) \quad \text{with} \quad F(x) = F_1(x) + F_2(x), \quad (2.9)$$

by the (simplified) iteration

$$x^{k+1} = \text{prox}_{\delta F_1}(x^k - \delta \nabla F_2(x^k)), \quad (2.10)$$

where δ is a constant, and $\text{prox}_{\delta F_1}$ is the proximity for δF_1 defined by

$$\text{prox}_{\delta F_1}(x) = \arg \min_{y \in \mathbb{R}^n} \{\delta F_1(y) + \frac{1}{2}\|x - y\|^2\}. \tag{2.11}$$

The following convergence theorem is an immediate consequence of the main results in [15].

THEOREM 2.1. *Assume that F_1 and F_2 satisfy*
  1. *F_1 is a proper, convex and continuous function whose range is in $\mathbb{R} \cup +\infty$, and*
  2. *F_2 is a proper, convex, differentiable function, whose gradient satisfies*

$$\|\partial F_2(x) - \partial F_2(y)\| \leq \beta \|x - y\|, \tag{2.12}$$

and whose range is in $\mathbb{R}$. Suppose that there exists at least one solution for (2.9). Then, for any initial guess x^0, iteration (2.10) converges to a solution of (2.9), when $0 < \delta < \frac{2}{\beta}$.

The PFBS iteration has been used to analyze the convergence of tight wavelet frame based algorithms for deblurring and inpainting in [3, 4, 5, 6, 7]. Next we apply the PFBS iteration (2.10) to the minimization problem (2.8) by splitting the energy function in (2.8) as $F_1(u) + F_2(u)$, where

$$F_1(u) = \mu(J(u) - J(u^k) - \langle u - u^k, p^k \rangle) \qquad \text{and} \qquad F_2(u) = \frac{1}{2}\|Au - f\|^2. \tag{2.13}$$

By the definition of the proximity operator, we have

$$\text{prox}_{\delta F_1}(v) = \arg \min_{u \in \mathbb{R}^n} \{\mu\delta(J(u) - J(u^k) - \langle u - u^k, p^k \rangle) + \frac{1}{2}\|u - v\|^2\}.$$

Then we obtain the PFBS iteration for (2.8) as

$$u^{k,i+1} = \arg \min_{u \in \mathbb{R}^n} \left\{\mu(J(u) - J(u^k) - \langle u - u^k, p^k \rangle) + \frac{1}{2\delta}\|u - (u^{k,i} - \delta A^T(Au^{k,i} - f))\|^2\right\}. \tag{2.14}$$

Based on Theorem 2.1, we have the following convergence result for (2.14).

PROPOSITION 2.2. *Assume that $0 < \delta < \frac{2}{\|AA^T\|}$. Then iteration (2.14) converges to a solution of (2.8) whenever (2.8) has a solution.*

Proof. It is easy to see that the assumptions on F_1 and F_2 in Theorem 2.1 are satisfied. $\square$

Based on this, iteration (2.1) can be written as

$$\begin{cases} u^{k,0} = u^k, \\ u^{k,i+1} = \arg \min_{u \in \mathbb{R}^n} \left\{\mu(J(u) - J(u^k) - \langle u - u^k, p^k \rangle) + \frac{1}{2\delta}\|u - (u^{k,i} - \delta A^T(Au^{k,i} - f))\|^2\right\}, \\ u^{k+1} = u^{k,M_k}, \\ p^{k+1} = p_k - \frac{1}{\mu\delta}(u^{k+1} - u^{k,M_k-1}) - \frac{1}{\mu}A^T(Au^{k,M_k-1} - f), \end{cases} \tag{2.15}$$

where $u^0 = p^0 = 0$.

The last equation in (2.15) for updating p^{k+1} is to ensure that $p^{k+1} \in \partial J(u^{k+1})$. Theoretically, to make this iteration to be identical with (2.1), the numbers M_k in each iteration should be infinity. Practically, the number M_k in each iteration is

chosen to be finite. If we choose M_k to be its minimal, i.e., $M_k = 1$ for all k, we obtain the algorithm

$$
\begin{cases}
u^{k+1} = \arg\min_{u \in \mathbb{R}^n} \left\{ \mu(J(u) - J(u^k) - \langle u - u^k, p^k \rangle) + \frac{1}{2\delta} \| u - (u^k - \delta A^T (Au^k - f)) \|^2 \right\}, \\
p^{k+1} = p^k - \frac{1}{\mu\delta}(u^{k+1} - u^k) - \frac{1}{\mu} A^T (Au^k - f),
\end{cases}
$$

(2.16)

which is exactly the linearized Bregman iteration (1.2).

2.2. Functional $J(u) + \frac{\tau}{2}\|u\|^2$. The reason why iteration (1.2) converges to the minimizer of (1.4) is because the linearized Bregman iteration for the functional $J(u)$ is the same as the linearized Bregman iteration for the cost functional $J(u) + \frac{\tau}{2}\|u\|^2$, $0 < \tau < \frac{1}{\mu\delta}$, with a different step size. This is shown in the next proposition.

PROPOSITION 2.3. Let $H(\widetilde{u}) = J(\widetilde{u}) + \frac{\tau}{2}\|\widetilde{u}\|^2$ with $0 < \tau < \frac{1}{\mu\delta}$, and $\frac{1}{\mu\widetilde{\delta}} = \frac{1}{\mu\delta} - \tau$. Define the sequence $\{\widetilde{u}^k\}_{k \in \mathbb{N}}$ and $\{\widetilde{p}^k\}_{k \in \mathbb{N}}$ by

$$
\begin{cases}
\widetilde{u}^{k+1} = \arg\min_{\widetilde{u} \in \mathbb{R}^n} \left\{ \mu(H(\widetilde{u}) - H(\widetilde{u}^k) - \langle \widetilde{u} - \widetilde{u}^k, \widetilde{p}^k \rangle) + \frac{1}{2\widetilde{\delta}} \| \widetilde{u} - (\widetilde{u}^k - \widetilde{\delta} A^T (A\widetilde{u}^k - f)) \|^2 \right\}, \\
\widetilde{p}^{k+1} = \widetilde{p}^k - \frac{1}{\mu\widetilde{\delta}}(\widetilde{u}^{k+1} - \widetilde{u}^k) - \frac{1}{\mu} A^T (A\widetilde{u}^k - f),
\end{cases}
$$

(2.17)

where $\widetilde{u}^0 = \widetilde{p}^0 = 0$. Let $\{u^k\}_{k \in \mathbb{N}}$ and $\{p^k\}_{k \in \mathbb{N}}$ be generated by (1.2). Then

$$
\widetilde{u}^k = u^k, \qquad \widetilde{p}^k = p^k + \tau u^k, \quad \forall k \in \mathbb{N}.
$$

(2.18)

Proof. We prove (2.18) by induction. It is obvious that it holds for $k = 0$. Assume that (2.18) holds until $k = j$. Then, by its definition, we have

$$
\widetilde{u}^{j+1} = \arg\min_{\widetilde{u} \in \mathbb{R}^n} \{ \mu(J(\widetilde{u}) + \frac{\tau}{2}\|\widetilde{u}\|^2 - J(u^j) - \frac{\tau}{2}\|u^j\|^2 - \langle \widetilde{u} - u^j, p^j \rangle - \tau\langle \widetilde{u} - u^j, u^j \rangle)
$$
$$
+ \frac{1}{2\widetilde{\delta}} \| \widetilde{u} - (u^j - \widetilde{\delta} A^T (Au^j - f)) \|^2 \}.
$$

(2.19)

Denote the energy function in the above minimization problem by $E(\widetilde{u})$. Then we have

$$
\frac{\partial (E(\widetilde{u}) - \mu J(\widetilde{u}) + \mu J(u^j))}{\partial \widetilde{u}} = \mu(\tau\widetilde{u} - (p^j + \tau u^j)) + \frac{1}{\widetilde{\delta}}[\widetilde{u} - (u^j - \widetilde{\delta} A^T (Au^j - f))]
$$
$$
= \mu(\tau + \frac{1}{\mu\widetilde{\delta}})(\widetilde{u} - u^j) + A^T (Au^j - f)
$$
$$
= \frac{1}{\delta}[\widetilde{u} - (u^j - \delta A^T (Au^j - f))]
$$

and

$$
\frac{\partial^2 (E(\widetilde{u}) - \mu J(\widetilde{u}) + \mu J(u^j))}{\partial(\widetilde{u})^2} = \frac{1}{\delta}.
$$

Hence, the function $E(\widetilde{u}) - \mu J(\widetilde{u}) + \mu J(u^j)$ is quadratic. Then it follows that

$$
E(\widetilde{u}) = \mu(J(\widetilde{u}) - J(u^j)) + \frac{1}{2\delta} \| \widetilde{u} - (u^j - \delta A^T (Au^j - f)) \|^2 + C(u^j).
$$

Since $C(u^j)$ is independent of $\widetilde{u}$, by (2.19), we have

$$
\widetilde{u}^{j+1} = \arg\min_{\widetilde{u} \in \mathbb{R}^n} \{ \mu(J(\widetilde{u}) - J(u^j)) + \frac{1}{2\delta} \| \widetilde{u} - (u^j - \delta A^T (Au^j - f)) \|^2 \}.
$$

Compared with the definition of u^{j+1} in (1.2), the above equality implies that

$$\widetilde{u}^{j+1} = u^{j+1}. \tag{2.20}$$

Substituting (2.20) and (2.18) with $k = j$ into the second equation in (2.17), we get

$$\begin{aligned}
\widetilde{p}^{j+1} &= p^j + \tau u^j - \frac{1}{\mu\widetilde{\delta}}(u^{j+1} - u^j) - \frac{1}{\mu}A^T(Au^j - f) \\
&= p^j + \tau u^{j+1} - (\tau + \frac{1}{\mu\widetilde{\delta}})(u^{j+1} - u^j) - \frac{1}{\mu}A^T(Au^j - f) \\
&= p^j - \frac{1}{\mu\delta}(u^{j+1} - u^j) - \frac{1}{\mu}A^T(Au^j - f) + \tau u^{j+1} \\
&= p^{j+1} + \tau u^{j+1}.
\end{aligned}$$

Thus, (2.18) holds for $k = j + 1$. $\square$

3. Analysis of Linearized Bregman Iteration. In this section, we give a complete analysis for the linearized Bregman iteration (1.2). We prove the convergence of (1.2) when $J(u)$ is continuous differentiable and the gradient satisfies (1.3). Moreover, we show that, whenever the sequence $\{u^k\}_{k\in\mathbb{N}}$ converges and $\{p^k\}_{k\in\mathbb{N}}$ is bounded, the limit of $\{u^k\}_{k\in\mathbb{N}}$ is the minimizer of

$$\min_{u\in\mathbb{R}^n} \left\{ \mu J(u) + \frac{1}{2\delta}\|u\|^2 : Au = f \right\}. \tag{3.1}$$

3.1. Convergence. To prove the convergence of the linearized Bregman iteration (1.2), we first estimate the residual $\|Au^k - f\|$. The following lemma follows from ideas in the proofs of Theorems 5.2 and 5.3 in [32]. To make the proof work, the key step is to construct a symmetric, positive semi-definite matrix Q_k^θ at each step k satisfying (3.4) and $C_2 I \leq Q_k^\theta + (\frac{1}{\delta} - \theta)I \leq C_1 I$, where $0 < C_2 \leq C_1 < \infty$ are independent of k. It is compared with the proof of Theorem 5.2 in [32] for the Bregman iteration (2.1), where one needs to find Q_k satisfying $C_2 I \leq Q_k \leq C_1 I$ and $Q_k(u^{k+1} - u^k) = -A^T(Au^{k+1} - f)$. Thanks to the additional terms in the linearized Bregman iteration, the matrix Q_k^θ demands weaker conditions than those imposed to Q_k in the proof of the corresponding result (Theorem 5.2 in [32]) for the Bregman iteration. The stronger conditions on Q_k in [32] force that the cost functional J there must meet stronger conditions.

LEMMA 3.1. *Suppose that $J(u)$ is convex and continuously differentiable satisfying (1.3). Let $\{u^k\}_{k\in\mathbb{N}}$ be generated by the linearized Bregman iteration (1.2) with $0 < \delta < \frac{2}{\|AA^T\|}$. Assume that $u^{k+1} \neq u^k$. Then*

$$\|Au^{k+1} - f\| \leq \eta\|Au^k - f\|, \tag{3.2}$$

where $\eta = \max\{\|I - \delta AA^T\|, \|I - \frac{\delta}{1+\delta\beta\mu}AA^T\|\} < 1$.
    Proof. Since

$$p^k \in \partial J(u^k), \qquad p^{k+1} \in \partial J(u^{k+1}),$$

and $J(u)$ is convex, the nonnegativity of the Bregman distance implies that

$$J(u^{k+1}) - J(u^k) - \langle u^{k+1} - u^k, p^k \rangle \geq 0$$

and

$$J(u^k) - J(u^{k+1}) - \langle u^k - u^{k+1}, p^{k+1} \rangle \geq 0.$$

By summing the above two equations together, we obtain

$$\langle u^{k+1} - u^k, p^{k+1} - p^k \rangle \geq 0. \tag{3.3}$$

Let $\theta \in (0, \frac{1}{\mu\delta})$ be a constant. Define $\gamma_k^\theta := \langle u^{k+1} - u^k, \mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k) \rangle$. Then $\gamma_k^\theta \geq \theta \|u^{k+1} - u^k\|^2$. When $u^{k+1} \neq u^k$, $\gamma_k^\theta > 0$. Therefore, we can define a matrix

$$Q_k^\theta = \frac{1}{\gamma_k^\theta} [\mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k)][\mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k)]^T,$$

where $\mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k)$ is viewed as a column vector. It is obvious that Q_k^θ is symmetric positive semi-definite and satisfies

$$\mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k) = Q_k^\theta(u^{k+1} - u^k). \tag{3.4}$$

Furthermore,

$$\begin{aligned}
\|Q_k^\theta\| &= \frac{1}{\gamma_k^\theta} \langle \mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k), \mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k) \rangle \\
&= \frac{1}{\gamma_k^\theta} (\langle \mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k), \mu(p^{k+1} - p^k) \rangle + \theta \gamma_k^\theta) \\
&\leq \frac{1}{\gamma_k^\theta} (\beta\mu + \theta) \langle u^{k+1} - u^k, \mu(p^{k+1} - p^k) \rangle + \theta \\
&\leq \frac{1}{\gamma_k^\theta} (\beta\mu + \theta) \langle u^{k+1} - u^k, \mu(p^{k+1} - p^k) + \theta(u^{k+1} - u^k) \rangle + \theta \\
&= \beta\mu + 2\theta,
\end{aligned}$$

where the first inequality is obtained by applying (1.3). Note that the second equation in (1.2) gives

$$\mu(p^{k+1} - p^k) + \frac{1}{\delta}(u^{k+1} - u^k) = -A^T(Au^k - f). \tag{3.5}$$

By substituting (3.4) into (3.5), we have

$$(Q_k^\theta + (\frac{1}{\delta} - \theta)I)(u^{k+1} - u^k) = -A^T(Au^k - f),$$

or equivalently,

$$u^{k+1} - u^k = -\delta(\delta Q_k^\theta + (1 - \delta\theta)I)^{-1} A^T(Au^k - f).$$

Therefore,

$$Au^{k+1} - f = (I - \delta A(\delta Q_k^\theta + (1 - \delta\theta)I)^{-1} A^T) (Au^k - f).$$

In the following, for given matrices B and B', the inequality $B \leq B'$ (or $B \geq B'$) means $w^T B w \leq w^T B' w$ (or $w^T B w \geq w^T B' w$) holds for any vector w of the right size. Since $0 \leq Q_k^\theta \leq (\beta\mu + 2\theta)I$, we have

$$(1 - \delta\theta)I \leq \delta Q_k^\theta + (1 - \delta\theta)I \leq (1 + \delta\beta\mu + \delta\theta)I,$$

which implies

$$\frac{1}{1-\delta\theta}I \geq (\delta Q_k^\theta + (1-\delta\theta)I)^{-1} \geq \frac{1}{1+\delta\beta\mu+\delta\theta}I.$$

This leads to

$$\frac{1}{1-\delta\theta}AA^T \geq A(\delta Q_k^\theta + (1-\delta\theta)I)^{-1}A^T \geq \frac{1}{1+\delta\beta\mu+\delta\theta}AA^T,$$

which further leads to

$$I - \frac{\delta}{1-\delta\theta}AA^T \leq I - \delta A(\delta Q_k^\theta + (1-\delta\theta)I)^{-1}A^T \leq I - \frac{\delta}{1+\delta\beta\mu+\delta\theta}AA^T.$$

Since all of matrices in the above equation are symmetric, we obtain

$$\|I - \delta A(\delta Q_k^\theta + (1-\delta\theta)I)^{-1}A^T\| \leq \max\{\|I - \frac{\delta}{1-\delta\theta}AA^T\|, \|I - \frac{\delta}{1+\delta\beta\mu+\delta\theta}AA^T\|\}.$$

Consequently,

$$\|Au^{k+1} - f\| \leq \max\{\|I - \frac{\delta}{1-\delta\theta}AA^T\|, \|I - \frac{\delta}{1+\delta\beta\mu+\delta\theta}AA^T\|\}\|Au^k - f\|.$$

Since θ can be any positive number in $(0, \frac{1}{\delta})$, and the norm is a continuous function of θ, by letting θ tend to 0, we obtain (3.2) with

$$\eta = \max\{\|I - \delta AA^T\|, \|I - \frac{\delta}{1+\delta\beta\mu}AA^T\|\}.$$

Furthermore, since $0 < \delta < \frac{2}{\|AA^T\|}$, we have

$$-I < I - \delta AA^T \leq I - \frac{\delta}{1+\delta\beta\mu}AA^T < I.$$

The above inequality also implies $\eta < 1$. $\square$

Next, we prove the convergence theorem for (1.2).

THEOREM 3.2. *Suppose that $J(u)$ is convex and continuously differentiable satisfying* (1.3). *Then both the sequences $\{u^k\}_{k\in\mathbb{N}}$ and $\{p^k\}_{k\in\mathbb{N}}$ generated by the linearized Bregman iteration* (1.2) *with $0 < \delta < \frac{2}{\|AA^T\|}$ converge with a rate*

$$\eta = \max\{\|I - \delta AA^T\|, \|I - \frac{\delta}{1+\delta\beta\mu}AA^T\|\} < 1.$$

Proof. Suppose that there exists an integer K such that $u^{K+1} = u^K$. Then, by the definition, we have $p^{K+1} \in \partial J(u^{K+1})$ and $p^K \in \partial J(u^K)$. Since, by the assumption, J is continuously differentiable, $p^{K+1} = p^K$. Therefore, from (1.2), we see that $u^{k+1} = u^k$ and $p^{k+1} = p^k$ for all $k \geq K$. Thus, both the sequences $\{u^k\}_{k\in\mathbb{N}}$ and $\{p^k\}_{k\in\mathbb{N}}$ converge.

Otherwise, suppose that $u^{k+1} \neq u^k$ for all k. We take the inner product of both sides of (3.5) with respect to $u^{k+1} - u^k$, then we have

$$\langle \mu(p^{k+1} - p^k), u^{k+1} - u^k \rangle + \frac{1}{\delta}\|u^{k+1} - u^k\|^2 = \langle -A^T(Au^k - f), u^{k+1} - u^k \rangle.$$

This, together with $\langle p^{k+1} - p^k, u^{k+1} - u^k \rangle \geq 0$, implies that

$$\frac{1}{\delta}\|u^{k+1} - u^k\|^2 \leq \langle -A^T(Au^k - f), u^{k+1} - u^k \rangle \leq \|A^T\|\|Au^k - f\|\|u^{k+1} - u^k\|.$$

Therefore,

$$\|u^{k+1} - u^k\| \leq \delta\|A^T\|\|Au^k - f\|.$$

On the other hand, since $0 < \delta < \frac{2}{\|AA^T\|}$, by Lemma 3.1,

$$\|Au^{k+1} - f\| \leq \eta\|Au^k - f\|.$$

Therefore,

$$\|Au^k - f\| \leq \eta^k\|Au^0 - f\|.$$

Thus, for any j and k such that $j > k$,

$$\|u^k - u^j\| \leq \sum_{i=k}^{j-1}\|u^i - u^{i+1}\| \leq \delta\|A^T\|\sum_{i=k}^{j-1}\|Au^i - f\| \leq \delta\|A^T\|\sum_{i=k}^{j-1}\eta^i\|Au^0 - f\|$$

$$\leq \delta\|A^T\|\|Au^0 - f\|\frac{\eta^k}{1 - \eta}. \tag{3.6}$$

We see from the above equation that $\{u^k\}_{k\in\mathbb{N}}$ is a Cauchy sequence, hence converges. Let u^* be its limit. By letting $j \to \infty$ in (3.6), we obtain

$$\|u^k - u^*\| \leq \frac{\delta\|A^T\|\|Au^0 - f\|}{1 - \eta}\eta^k.$$

It means that the convergence rate of $\{u^k\}_{k\in\mathbb{N}}$ is η. The convergence of $\{p^k\}_{k\in\mathbb{N}}$ can be shown analogously. ☐

3.2. Minimization. In this subsection, we study minimization properties of the limit for the linearized Bregman iteration (1.2). We have the following theorem, which holds for any continuous convex function $J(u)$ including $\|u\|_1$. Some ideas of this proof are motivated by those in the proof of Theorem 5.3 in [32] and the proof of Theorem 3.1 in [28].

THEOREM 3.3. *Suppose that J is a continuous convex function. Let $\{u^k\}_{k\in\mathbb{N}}$ and $\{p^k\}_{k\in\mathbb{N}}$ be generated by (1.2). Assume that $\{u^k\}_{k\in\mathbb{N}}$ converges and $\{p^k\}_{k\in\mathbb{N}}$ is bounded. Then the limit u^* of $\{u^k\}_{k\in\mathbb{N}}$ is the unique solution of (3.1).*

Proof. We first prove that $Au^* = f$. The proof of this fact given here is from [28]. It is proven by contradiction. Assume $Au^* \neq f$. Then $A^T(Au^* - f) \neq 0$ since A^T has full rank. This means that, for some i, $(A^T(Au^k - f))_i$ converges to a nonzero value, which by (1.2) implies that $p_i^{k+1} - p_i^k$ does as well. On the other hand, by the assumption, we have $\{p^k\}_{k\in\mathbb{N}}$, hence $\{p_i^k\}_{k\in\mathbb{N}}$ is bounded. However, if $p_i^{k+1} - p_i^k$ converges to a nonzero limit, $\{p_i^k\}_{k\in\mathbb{N}}$ is not bounded. This contradicts the assumption that $\{p^k\}_{k\in\mathbb{N}}$ is bounded. Therefore, $Au^* = f$.

Since $p^0 = u^0 = 0$, by the second equation in (1.2), we have

$$\mu p^k + \frac{1}{\delta}u^k = A^T\sum_{j=0}^{k-1}(f - Au^j).$$

Define $w^k = \sum_{j=0}^{k-1}(f - Au^j)$. Then, since A is surjective and both $\{p^k\}_{\mathbb{N}}$ and $\{u^k\}_{k\in\mathbb{N}}$ are bounded, we have that $\{w^k\}_{k\in\mathbb{N}}$ is bounded, i.e., $\|w^k\| \le C$ for all k.

Let $H(u) = \mu J(u) + \frac{1}{2\delta}\|u\|^2$. By the definition, $p^k \in \partial J(u^k)$. Therefore,

$$\mu p^k + \frac{1}{\delta}u^k \in \partial H(u^k).$$

Since $H(u)$ is a strongly convex function, there exists a unique solution of (3.1), which is denoted by u^{opt}. By the nonnegativity of the Bregman distance for $H(u)$, we obtain

$$H(u^k) \le H(u^{\text{opt}}) - \langle u^{\text{opt}} - u^k, \mu p^k + \frac{1}{\delta}u^k \rangle = H(u^{\text{opt}}) - \langle u^{\text{opt}} - u^k, A^T w^k \rangle$$
$$= H(u^{\text{opt}}) - \langle A(u^{\text{opt}} - u^k), w^k \rangle. \tag{3.7}$$

On the other hand, by Cauchy-Schwarz inequality, we have

$$|\langle A(u^{\text{opt}} - u^k), w^k \rangle| \le \|A(u^{\text{opt}} - u^k)\|\|w^k\| \le C\|A(u^{\text{opt}} - u^k)\|.$$

Letting $k \to \infty$, and noticing that $Au^{\text{opt}} = Au^* = f$, we obtain that

$$\lim_{k\to\infty} |\langle A(u^{\text{opt}} - u^k), w^k \rangle| = 0.$$

Since $J(u)$ hence $H(u)$ is continuous, by letting $k \to \infty$ in (3.7), we have

$$H(u^*) \le H(u^{\text{opt}}).$$

This and $Au^* = f$ means that u^* is the unique solution of (3.1). $\square$

4. Application to Compressed Sensing. As for applications of compressed sensing, one needs to solve (1.8). In this section, we develop algorithms for (1.2) by choosing $J = J_\epsilon$, where J_ϵ is defined in (1.11), to approximate a special solution of (1.8).

We first show that the linearized Bregman iteration (1.2) with $J = J_\epsilon$ becomes iteration (1.12). Therefore, it converges to the solution of (1.10). Furthermore, we will see that iteration (1.12), which is as simple as iteration (1.5), is iteration (1.5) numerically for a sufficiently small ϵ. Then we prove that, as $\mu\epsilon$ tends to 0, the solution of (1.10) tends to the solution of (1.9). Moreover, as μ goes to infinity, the solution of (1.9) goes to a particular solution of (1.8). Therefore, by properly choosing μ and $\mu\epsilon$, iteration (1.12) converges to a good approximation of a solution of (1.8). We further discuss the sparsity of the solution of (1.9). Numerical simulations are also given to illustrate the effectiveness of iteration (1.12) in compressed sensing.

4.1. Algorithms. In this subsection, we derive the linearized Bregman iteration (1.2) for $J = J_\epsilon$ in (1.11). The first equation in (1.2) becomes

$$u^{k+1} = \arg\min_{u\in\mathbb{R}^n} \left\{ \mu(J_\epsilon(u) - J_\epsilon(u^k) - \langle u - u^k, p^k \rangle) + \frac{1}{2\delta}\|u - (u^k - \delta A^T(Au^k - f))\|^2 \right\}. \tag{4.1}$$

It can be easily verified that the minimization problem (4.1) can be decoupled into the following n one variable minimization problems. Let $w^k = u^k - \delta A^T(Au^k - f)$. Then

$$u^{k+1}(i) = \arg\min_{\zeta\in\mathbb{R}}\{\mu F_\epsilon(\zeta) - \mu F_\epsilon(u^k(i)) - \mu(\zeta - u^k(i))p^k(i) + \frac{1}{2\delta}(\zeta - w^k(i))^2\}, \quad i = 1, \ldots, n, \tag{4.2}$$

where F_ϵ is defined in (1.11), that is also shown in Figure 1.1. Since the above minimization problem is strongly convex, it has a unique solution. Hence, the solution $u^{k+1}(i)$ of (4.2) is the unique solution of the following equation:

$$\mu\delta F_\epsilon'(\zeta) + \zeta = w^k(i) + \delta\mu p^k(i). \tag{4.3}$$

Furthermore, $t_{\lambda,\epsilon}(\xi)$ defined in (1.14) is the unique solution of

$$\lambda F_\epsilon'(\zeta) + \zeta = \xi. \tag{4.4}$$

Indeed, the derivative of F_ϵ is given by

$$F_\epsilon'(\zeta) = \begin{cases} \frac{\zeta}{\epsilon} & \text{if } |\zeta| \leq \epsilon, \\ \text{sgn}(\zeta) & \text{if } |\zeta| > \epsilon. \end{cases} \tag{4.5}$$

The conclusion follows by substituting $\zeta = t_{\lambda,\epsilon}(\xi)$ into (4.4).

Therefore, the solution $u^{k+1}(i)$ of (4.2) is

$$u^{k+1}(i) = t_{\mu\delta,\epsilon}(w^k(i) + \delta\mu p^k(i)).$$

Due to the definition of $T_{\lambda,\epsilon}$ in (1.13), we have

$$u^{k+1} = T_{\mu\delta,\epsilon}(w^k + \delta\mu p^k) = T_{\mu\delta,\epsilon}(u^k + \delta\mu p^k - \delta A^T(Au^k - f)). \tag{4.6}$$

Now the linearized Bregman iteration (1.2) with $J = J_\epsilon$ becomes

$$\begin{cases} u^{k+1} = T_{\mu\delta,\epsilon}(u^k + \delta\mu p^k - \delta A^T(Au^k - f)), \\ p^{k+1} = p^k - \frac{1}{\mu\delta}(u^{k+1} - u^k) - \frac{1}{\delta}A^T(Au^k - f), \end{cases} \tag{4.7}$$

where $u^0 = p^0 = 0$. By introducing a variable $v^k = \mu p^k + \frac{1}{\delta}u^k$, we obtain the final linearized Bregman iteration (1.12)

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

where $u^0 = v^0 = 0$. We see that iteration (1.5), which has been illustrated by [28], an extremely fast algorithm, is just (1.12) with $\epsilon = 0$. Note that $T_{\lambda,0}$ equals T_λ, the soft thresholding operator in (1.6), and that $T_{\lambda,\epsilon}$ is essentially T_λ numerically, when ϵ is sufficiently small. Hence, iteration (1.12) is essentially the same as iteration (1.5) numerically, when ϵ is sufficiently small.

By applying Theorems 3.2 and 3.3, we have the following convergence result for (1.12).

COROLLARY 4.1. *Let $\epsilon > 0$, and $0 < \delta < \frac{2}{\|AA^T\|}$. Then the sequence $\{u^k\}_{k\in\mathbb{N}}$ generated by (1.12) converges to the unique solution of (1.10), i.e.,*

$$\min_{u\in\mathbb{R}^n}\left\{\mu J_\epsilon(u) + \frac{1}{2\delta}\|u\|^2 : Au = f\right\}.$$

Proof. From (4.5), we see that F_ϵ' is a Lipshitz function with constant $\frac{1}{\epsilon}$. Therefore, $\partial J_\epsilon(u)$ satisfies (1.3) with $\beta = \frac{1}{\epsilon}$. The rest follows immediately from Theorems 3.2 and 3.3. □

Furthermore, by Theorem 3.3, we have the following corollary for $\epsilon = 0$.

COROLLARY 4.2. *Let $\epsilon = 0$. Assume that the sequence $\{u^k\}_{k \in \mathbb{N}}$ generated by (1.12), or equivalently (1.5), converges. Then the limit is the unique solution of (1.9), i.e.,*

$$\min_{u \in \mathbb{R}^n} \left\{ \mu \|u\|_1 + \frac{1}{2\delta} \|u\|^2 : Au = f \right\}.$$

Proof. Since $\partial J_0(u) \in [-1, 1]^n$ and $p^k \in \partial J_0(u^k)$, $\{p^k\}_{k \in \mathbb{N}}$ is bounded. Therefore, the corollary follows from Theorem 3.3. $\square$

4.2. Approximation. Let $u^*_{\epsilon,\mu} = \lim_{k \to \infty} u^k$, where u^k is generated by (1.12). Then $u^*_{\epsilon,\mu}$ for $\epsilon > 0$ is the unique solution of (1.10). This subsection is to show that $u^*_{\epsilon,\mu}$ can be a good approximation to a solution of (1.8).

First, we show that for fixing μ, $u^*_{\epsilon,\mu}$ tends to the unique solution of (1.9) as $\mu\epsilon \to 0$.

THEOREM 4.3. *Let $u^*_{\epsilon,\mu}$, where $\epsilon > 0$, be the limit of (1.12), and $u^*_{0,\mu}$ the unique solution of (1.9). Then*

$$\|u^*_{0,\mu} - u^*_{\epsilon,\mu}\| \leq \sqrt{\delta\mu n\epsilon}. \tag{4.8}$$

Proof. Let $H_\epsilon(u) = \mu J_\epsilon(u) + \frac{1}{2\delta} \|u\|^2$ and $H_0(u) = \mu \|u\|_1 + \frac{1}{2\delta} \|u\|^2$. Since, for any $u(i)$, $|F_\epsilon(u(i)) - |u(i)|| \leq \frac{\epsilon}{2}$, we obtain that, for any u,

$$|H_\epsilon(u) - H_0(u)| \leq \frac{1}{2}\mu n\epsilon. \tag{4.9}$$

This, together with $H_0(u^*_{0,\mu}) \leq H_0(u^*_{\epsilon,\mu})$ and $H_\epsilon(u^*_{0,\mu}) \geq H_\epsilon(u^*_{\epsilon,\mu})$, implies that

$$\begin{aligned}
0 &\leq H_0(u^*_{\epsilon,\mu}) - H_0(u^*_{0,\mu}) \\
&= (H_0(u^*_{\epsilon,\mu}) - H_\epsilon(u^*_{\epsilon,\mu})) + (H_\epsilon(u^*_{\epsilon,\mu}) - H_\epsilon(u^*_{0,\mu})) + (H_\epsilon(u^*_{0,\mu}) - H_0(u^*_{0,\mu})) \\
&\leq (H_0(u^*_{\epsilon,\mu}) - H_\epsilon(u^*_{\epsilon,\mu})) + (H_\epsilon(u^*_{0,\mu}) - H_0(u^*_{0,\mu})) \\
&\leq |H_0(u^*_{\epsilon,\mu}) - H_\epsilon(u^*_{\epsilon,\mu})| + |H_\epsilon(u^*_{0,\mu}) - H_0(u^*_{0,\mu})| \\
&\leq \mu n\epsilon.
\end{aligned}$$

Define the indicator function

$$I(u) = \begin{cases} 0, & \text{if } Au = f, \\ +\infty, & \text{otherwise.} \end{cases}$$

Let $K(u) = H_0(u) + I(u)$. Since the set $\{u : Au = f\}$ is a nonempty convex set, the function $I(u)$ is a proper, lower semi-continuous and convex function. This, together with $H_0(u)$ is a strongly convex function, implies that $K(u)$ is a strongly convex function. In particular, for any element $p_K(u) \in \partial K(u)$, we have

$$K(v) - K(u) - \langle v - u, p_K(u) \rangle \geq \frac{1}{\delta} \|u - v\|^2. \tag{4.10}$$

On the other hand, it is obvious that the minimization problem $\min_u \{K(u)\}$ is equivalent to (1.9). By the first-order optimality condition, $0 \in \partial K(u^*_{0,\mu})$. Thus, by (4.10),

$$K(u^*_{\epsilon,\mu}) - K(u^*_{0,\mu}) \geq \frac{1}{\delta} \|u^*_{\epsilon,\mu} - u^*_{0,\mu}\|^2.$$

However, $K(u_{\epsilon,\mu}^*) = H_0(u_{\epsilon,\mu}^*)$, due to the fact that $Au_{\epsilon,\mu}^* = f$. Similarly, $K(u_{0,\mu}^*) = H_0(u_{0,\mu}^*)$. Therefore,

$$\|u_{\epsilon,\mu}^* - u_{0,\mu}^*\|^2 \le \delta |H_0(u_{\epsilon,\mu}^*) - H_0(u_{0,\mu}^*)| \le \delta\mu n\epsilon.$$

□

The above result says that when $\mu\epsilon$ is sufficiently small, the limit of (1.12) is sufficiently close to the unique solution of (1.9). Next, we show that when μ is sufficiently large, the unique solution of (1.9) is sufficiently close to a solution of (1.8).

Let $\mathcal{S}$ be the set of all solutions of (1.8) and define

$$u_1 := \arg\min_{u \in \mathcal{S}}\{\|u\|^2\}, \tag{4.11}$$

where, as always, $\|\cdot\|$ denotes the ℓ_2 norm. Since the set $\mathcal{S}$ is convex, and $\|u\|^2$ is coercive and strongly convex, u_1 is uniquely determined. We have the following result.

THEOREM 4.4. *Let* $u_{0,\mu}^*$ *be the unique solution of* (1.9) *and* u_1 *be given in* (4.11). *Then*

$$\lim_{\mu\to\infty} \|u_{0,\mu}^* - u_1\| = 0. \tag{4.12}$$

Proof. Since $u_{0,\mu}^*$ and u_1 are solutions of (1.9) and (1.8) respectively, we have that

$$\|u_{0,\mu}^*\|_1 + \frac{1}{2\mu\delta}\|u_{0,\mu}^*\|^2 \le \|u_1\|_1 + \frac{1}{2\mu\delta}\|u_1\|^2 \tag{4.13}$$

and

$$\|u_1\|_1 \le \|u_{0,\mu}^*\|_1. \tag{4.14}$$

Since $\frac{1}{2\mu\delta} > 0$, summing the above two equations gives

$$\|u_{0,\mu}^*\| \le \|u_1\|. \tag{4.15}$$

This means that $\|u_{0,\mu}^*\|$ is uniformly bounded in μ, which leads to the fact that $\|u_{0,\mu}^*\|_1$ is also uniformly bounded by the norm equivalence in finite dimensional space. By letting $\mu \to \infty$ in (4.13), we see that

$$\limsup_{\mu\to\infty} \|u_{0,\mu}^*\|_1 \le \|u_1\|_1.$$

By letting $\mu \to \infty$ in (4.14), we see that

$$\|u_1\|_1 \le \liminf_{\mu\to\infty} \|u_{0,\mu}^*\|_1.$$

Therefore, $\lim_{\mu\to\infty} \|u_{0,\mu}^*\|_1$ exists, and

$$\lim_{\mu\to\infty} \|u_{0,\mu}^*\|_1 = \|u_1\|_1. \tag{4.16}$$

Next, we show that, if a sequence $\{u_{0,\mu_k}^*\}_{k\in\mathbb{N}}$, where $\lim_{k\to\infty}\mu_k = \infty$, converges, then it must converge to u_1. Let u_c be the limit of $\{u_{0,\mu_k}^*\}_{k\in\mathbb{N}}$, i.e.,

$$\lim_{k\to\infty}\|u_{0,\mu_k}^* - u_c\| = 0.$$

Since $Au_{0,\mu_k}^* = f$ for all k, and A is a continuous linear operator, we must have that $Au_c = f$. This, together with (4.16), implies that u_c is a solution of (1.8). Therefore,

$$\|u_c\| \geq \|u_1\|, \tag{4.17}$$

by the definition of u_1. On the other hand, $\|u_{0,\mu_k}^*\| \leq \|u_1\|$ by (4.15). Letting $k\to\infty$ leads to $\|u_c\| \leq \|u_1\|$. This, together with (4.17), implies that $\|u_c\| = \|u_1\|$. Since u_c is a solution of (1.8) with $\|u_c\| = \|u_1\|$, we conclude that $u_c = u_1$ due to the fact that u_1 is uniquely defined.

Finally, we prove (4.12) by contradiction. Suppose that $u_{0,\mu}^*$ does not converge to u_1. Then there exists a $\sigma > 0$ and a sequence $\{\mu_k\}_{k\in\mathbb{N}}$ with $\lim_{k\to\infty}\mu_k = \infty$ such that

$$\|u_{0,\mu_k}^* - u_1\| > \sigma, \qquad \forall k. \tag{4.18}$$

Since the sequence $\{u_{0,\mu_k}^*\}_{k\in\mathbb{N}}$ is bounded, there is a convergent subsequence that must converge to u_1 by the discussions above. This leads to a contradiction. $\square$

Combining Theorems 4.3 and 4.4 together, we conclude that, by properly choosing μ and $\mu\epsilon$, one can use the limit of the linearized Bregman iteration (1.12) to approximate to a special solution of the basis pursuit problem (1.8).

THEOREM 4.5. *Let u_1 be given in* (4.11). *For given $\sigma > 0$, one can choose μ, and $\epsilon \leq \frac{\sigma^2}{4n\delta\mu}$, such that*

$$\|u_{\epsilon,\mu}^* - u_1\| \leq \sigma,$$

where $u_{\epsilon,\mu}^*$ is the limit of (1.12) *and the unique solution of* (1.10).

Proof. By Theorem 4.4, there exists a μ such that

$$\|u_{0,\mu}^* - u_1\| \leq \frac{\sigma}{2}.$$

Fixing this μ, since we choose $\epsilon \leq \frac{\sigma^2}{4n\delta\mu}$, we have $\sqrt{\delta\mu n\epsilon} \leq \frac{\sigma}{2}$. By Theorem 4.3, we have

$$\|u_{0,\mu}^* - u_{\epsilon,\mu}^*\| \leq \frac{\sigma}{2}.$$

This leads to

$$\|u_{\epsilon,\mu}^* - u_1\| \leq \|u_{\epsilon,\mu}^* - u_{0,\mu}^*\| + \|u_{0,\mu}^* - u_1\| \leq \sigma.$$

$\square$

4.3. Sparsity. The ultimate goal of all these efforts is to find a sparse solution of the equation $Au = f$. The sparsest solution amounts to solving the problem

$$\min_u\{\|u\|_0 : Au = f\}, \tag{4.19}$$

where $\|u\|_0$ is the number of nonzero entries in u. As pointed out by [2], it is a conceptually challenging and computationally hard problem. Due to the combinatory

complexity of (4.19), it is usually relaxed to its "nearest" convex counterpart (1.8). In fact, the equation (1.8) can lead to a much sparser solution than the least square solution given by $u = A^T(AA^T)^{-1}f$. This paper proves that the simple and fast iteration given in (1.12) converges and its limit can be a good approximation of a solution of (1.8). Recently, there have been a vast number of papers (e.g., see [8, 19, 23, 24, 29, 33, 34]) concentrating on proving the equivalence between (1.8) and (4.19) under suitable assumptions. A recent survey article [2] gives a detailed account of this, and interested readers should consult [2] and the literature cited there for details. One of the results in this direction (e.g., Theorem 4 in [2]) roughly says that, if the columns of the matrix A are incoherent, then (1.8) is equivalent to (4.19). The following result indicates the same flavor of this. It says that, to have a sparse solution of (1.9), the columns of matrix A have to be incoherent.

THEOREM 4.6. *Assume that the columns A_i of A are normalized to 1, i.e., $\|A_i\| = 1$ for all i. Let $u_{0,\mu}^*$ be the unique solution of (1.9), and $\rho_{ij} = \langle A_i, A_j \rangle$ be the correlation between A_i and A_j. Then, if $u_{0,\mu}^*(i)u_{0,\mu}^*(j) > 0$, there exists a constant κ such that*

$$|u_{0,\mu}^*(i) - u_{0,\mu}^*(j)| \leq \kappa\sqrt{2(1 - \rho_{ij})}.$$

Proof. In the Lagrange multiplier formulation, solving (1.9) is equivalent to solving

$$\begin{cases} 0 \in \partial(\mu\|u_{0,\mu}^*\|_1) + \frac{1}{\delta}u_{0,\mu}^* + A^T w, \\ Au_{0,\mu}^* = f, \end{cases} \tag{4.20}$$

where w is the Lagrange multiplier for the constraint $Au = f$. By the first equation in (4.20), we have for each i, j,

$$0 \in \partial\mu|u_{0,\mu}^*(i)| + \frac{1}{\delta}u_{0,\mu}^*(i) + \langle A_i, w \rangle \tag{4.21}$$

and

$$0 \in \partial\mu|u_{0,\mu}^*(j)| + \frac{1}{\delta}u_{0,\mu}^*(j) + \langle A_j, w \rangle. \tag{4.22}$$

By the assumption that $u_{0,\mu}^*(i)u_{0,\mu}^*(j) > 0$, both $\partial\mu|u_{0,\mu}^*(i)|$ and $\partial\mu|u_{0,\mu}^*(j)|$ contain only one element. Hence, the "$\in$" in (4.21) and (4.22) can be replaced by "$=$", and $\partial\mu|u_{0,\mu}^*(i)| = \partial\mu|u_{0,\mu}^*(j)|$. Subtracting (4.21) from (4.22) gives

$$|u_{0,\mu}^*(i) - u_{0,\mu}^*(j)| = \delta|\langle A_i - A_j, w \rangle| \leq \delta\|A_i - A_j\|\|w\|. \tag{4.23}$$

Since $\|A_i\| = \|A_j\| = 1$, we have $\|A_i - A_j\|^2 = 2(1 - \langle A_i, A_j \rangle) = 2(1 - \rho_{ij})$.

Next we estimate $\|w\|$. The first equation in (4.20) gives

$$A^T w \in -\partial(\mu\|u_{0,\mu}^*\|_1) - \frac{1}{\delta}u_{0,\mu}^*.$$

To have a solution, there must exist a vector $p \in \partial(\|u_{0,\mu}^*\|_1)$ such that

$$A^T w = -\mu p - \frac{1}{\delta}u_{0,\mu}^*.$$

Hence, the right hand side belongs to the range of A^T. Then, by multiplying both sides by A, we have that

$$w = -\mu(AA^T)^{-1}Ap - \frac{1}{\delta}(AA^T)^{-1}Au^*_{0,\mu}$$
$$= -\mu(AA^T)^{-1}Ap - \frac{1}{\delta}(AA^T)^{-1}f.$$

Therefore,

$$\|w\| \leq \mu\|(AA^T)^{-1}A\|\|p\| + \frac{1}{\delta}\|(AA^T)^{-1}\|\|f\|. \tag{4.24}$$

Because each entry of $p \in \partial(\|u^*_{0,\mu}\|_1)$ is in $[-1,1]$, we have

$$\|p\| \leq \sqrt{n}. \tag{4.25}$$

Note that we also have

$$\|(AA^T)^{-1}A\| = \sqrt{\|((AA^T)^{-1}A)^T(AA^T)^{-1}A\|} = \sqrt{\|A^T(AA^T)^{-2}A\|} = \sqrt{\|(AA^T)^{-1}\|}. \tag{4.26}$$

By combining (4.23), (4.24), (4.25) and (4.26) together, we finally obtain that

$$|u^*_{0,\mu}(i) - u^*_{0,\mu}(j)| \leq \left(\delta\mu\sqrt{n\|(AA^T)^{-1}\|} + \|(AA^T)^{-1}\|\|f\|\right)\sqrt{2(1-\rho_{ij})}.$$

□

This result indicates that the entries of the solution of (1.9) corresponding to strongly correlated columns tend to be close. This may affect the sparsity of the solution (1.9), when the columns of A have a strong coherence. The similar phenomenon is known for the solution of (1.8) (see, e.g., [2]).

Finally, we remark that the minimization of the cost functional $\mu\|u\|_1 + \frac{1}{2\delta}\|u\|^2$ in (1.9) is known as the elastic net [35] in variable selections in statistics. The proof of Theorem 4.6 is motivated by the corresponding results of [35].

4.4. Simulations. Although numerical simulations are not the major focus of this paper, as a comprehensive numerical study on the linearized Bregman iteration based on (1.5) for compressed sensing given in [28], we give some numerical results to show that iteration (1.12) for (1.10) is numerically the same as iteration (1.5) for (1.9) when ϵ is sufficiently small, as the analysis of this paper already says. Hence, we can foresee that iteration (1.12) will be widely used in compressed sensing as (1.5), since it is also shown to be a convergent iteration with a rate.

In the numerical experiments, we choose A to be partial discrete cosine transform (DCT) matrices, whose rows are randomly chosen from the $n \times n$ DCT matrices. These matrices are known to be efficient for compressed sensing. We store A in the computer implicitly. Only the indices for the chosen rows are stored, so that the matrix-vector multiplications Au and A^Tu can be computed very efficiently by the fast DCT or the fast inverse DCT.

The numbers of nonzeroes of the tested original sparse signals $\bar{u}$ are $0.05n$ or $0.02n$ rounded to the nearest integers. The positions of the nonzeroes of $\bar{u}$ are randomly selected, and the values of the nonzeroes are randomly drawn from the uniform distribution in the interval $(-0.2, 0.2) \pm 1$, where the signs $+$ and $-$ are randomly selected with equal probabilities. Then the observed data f are computed by $f = A\bar{u}$. Since $\bar{u}$

n	m	$\|\bar{u}\|_0$	# of iters for $\epsilon =$			$\|u^*_{\epsilon,\mu} - \bar{u}\|/\|\bar{u}\|$ for $\epsilon =$			time(s) for $\epsilon =$		
			0	10^{-8}	10^{-4}	0	10^{-8}	10^{-4}	0	10^{-8}	10^{-4}
4000	2000		51.4	51.4	51.4	9.1e-6	9.1e-6	1.7e-4	0.21	0.21	0.22
20000	10000	$0.05n$	53	53	53	8.3e-6	8.3e-6	1.7e-4	1.22	1.19	1.18
50000	25000		78.4	78.4	78.4	1.1e-5	1.1e-5	2.4e-4	4.48	4.42	4.51
4000	1327		68.1	68.1	68.1	9.3e-6	9.3e-6	2.8e-4	0.31	0.29	0.28
20000	7923	$0.02n$	57.4	57.4	57.4	7.6e-6	7.6e-6	2.2e-4	1.23	1.25	1.26
50000	21640		50.2	50.2	50.2	8.0e-6	8.0e-6	1.9e-4	2.81	2.77	2.72

TABLE 4.1
Experiment results for partial DCT matrices.

is sparse enough, we expect that the solution of (1.8), which is approximated by our algorithms (1.5) and (1.12), can yield $\bar{u}$.

We choose $\mu = 10$ and $\delta = 1.9$. The stopping criterion is

$$\frac{\|Au^k - f\|}{\|f\|} < 10^{-5}.$$

The results are summarized in Table 4.1, where all the quantities are the averages of the outputs obtained from 10 random instances. From Table 4.1, we see that both iteration (1.5) and iteration (1.12) with small $\epsilon = 10^{-8}$ are very effective in solving (1.1) arising from compressed sensing. Furthermore, as predicted, when $\epsilon = 10^{-8}$, that is, very close to 0, the results obtained by (1.12) are almost identical to that obtained by (1.5) (the case $\epsilon = 0$). However, if we choose a larger $\epsilon = 10^{-4}$, then the difference between the results obtained by (1.5) (the case $\epsilon = 0$) and those by (1.12) becomes larger; see the errors $\frac{\|u^*_{\epsilon,\mu} - \bar{u}\|}{\|\bar{u}\|}$ for different ϵ's in the table.

As illustrated in [28], algorithm (1.12) is robust to noise. Next, we show that, similar to algorithm (1.5), algorithm (1.12) is also robust to noise. The settings are the same as those in the previous experiment, but we add a Gaussian noise of standard deviation $\sigma = 0.03$ into $A\bar{u}$ to obtain f. We choose $\mu = 10$ and $\delta = 1.9$. We stop the iteration when the square error is less than the variance of the noise, i.e., when

$$\|Au^k - f\|^2 \leq m\sigma^2.$$

We show the results in Table 4.2. From this table, we see that both iteration (1.5) and iteration (1.12) with small ϵ are robust to noise. Furthermore, again, when $\epsilon = 10^{-8}$, that is, very close to 0, the results obtained by (1.12) are almost identical to that obtained by (1.5) (the case $\epsilon = 0$).

In this simulation, $\bar{u}$ has a relatively small dynamic range. When $\bar{u}$ has a large dynamic range, a new and simple numerical device called "kicking" (which resembles line search) is introduced and used in [28] to speed up convergence of (1.5) numerically. A similar "kicking" device can also be applied to iteration (1.12) to speed up the convergence numerically for a large dynamic range of $\bar{u}$. We forgo the detailed discussions of the "kicking" device here, since it is not the focus of this paper. The interested readers should consult [28] for details.

n	m	$\|\bar{u}\|_0$	# of iters for $\epsilon =$			$\|u^*_{\epsilon,\mu} - \bar{u}\|/\|\bar{u}\|$ for $\epsilon =$			time(s) for $\epsilon =$		
			0	10^{-8}	10^{-4}	0	10^{-8}	10^{-4}	0	10^{-8}	10^{-4}
4000	2000		44.4	44.4	44.4	4.9e-2	4.9e-2	4.9e-2	0.19	0.19	0.19
20000	10000	$0.05n$	42.4	42.4	42.4	5.4e-2	5.4e-2	5.4e-2	0.99	0.98	1.03
50000	25000		42.2	42.2	42.2	5.4e-2	5.4e-2	5.4e-2	2.38	2.31	2.38
4000	1327		78.9	78.9	78.2	6.4e-2	6.4e-2	6.4e-2	0.36	0.35	0.33
20000	7923	$0.02n$	49.9	49.9	49.8	5.2e-2	5.2e-2	5.2e-2	1.13	1.08	1.08
50000	21640		43.3	43.3	43.1	4.9e-2	4.9e-2	5.0e-2	2.43	2.37	2.39

TABLE 4.2

Experiment results for noisy f with noise $\sigma = 0.03$ for partial DCT matrices.

REFERENCES

[1] L. M. BRÈGMAN, *A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming*, Ž. Vyčisl. Mat. i Mat. Fiz., 7 (1967), pp. 620–631.

[2] A. M. BRUCKSTEIN, D. L. DONOHO, AND M. ELAD, *From sparse solutions of systems of equations to sparse modeling of signals and images*, 2008. to appear in SIAM Review.

[3] J.-F. CAI, R. H. CHAN, L. SHEN, AND Z. SHEN, *Restoration of chopped and nodded images by framelets*, SIAM J. Sci. Comput., 30 (2008), pp. 1205–1227.

[4] J.-F. CAI, R. H. CHAN, L. SHEN, AND Z. SHEN, *Simultaneously Inpainting in Image and Transformed Domains*, 2008. preprint.

[5] J.-F. CAI, R. H. CHAN, AND Z. SHEN, *A framelet-based image inpainting algorithm*, Appl. Comput. Harmon. Anal., 24 (2008), pp. 131–149.

[6] J.-F. CAI, R. H. CHAN, AND Z. SHEN, *Simultaneous Cartoon and Texture Inpainting*, 2008. preprint.

[7] J.-F. CAI AND Z. SHEN, *Deconvolution: A wavelet frame approach, II*, 2008. preprint.

[8] E. CANDÈS AND J. ROMBERG, *Sparsity and incoherence in compressive sampling*, Inverse Problems, 23 (2007), pp. 969–985.

[9] E. J. CANDÈS, *Compressive sampling*, in International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 1433–1452.

[10] E. J. CANDÈS AND J. ROMBERG, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. Comput. Math., 6 (2006), pp. 227–254.

[11] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[12] E. J. CANDES AND T. TAO, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.

[13] A. CHAI AND Z. SHEN, *Deconvolution: A wavelet frame approach*, Numer. Math., 106 (2007), pp. 529–587.

[14] P. L. COMBETTES AND J.-C. PESQUET, *Proximal thresholding algorithm for minimization over orthonormal bases*, SIAM J. Opt., 18 (2007), pp. 1351–1376 (electronic).

[15] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200 (electronic).

[16] J. DARBON AND S. OSHER, *Fast discrete optimization for sparse approximations and deconvolutions*, 2007. preprint.

[17] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.

[18] I. DAUBECHIES, G. TESCHKE, AND L. VESE, *Iteratively solving linear inverse problems under general convex constraints*, Inverse Probl. Imaging, 1 (2007), pp. 29–46.

[19] R. A. DEVORE, *Deterministic constructions of compressed sensing matrices*, J. Complexity. 23 (2007), pp. 918–925. preprint.

[20] D. L. DONOHO, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.

[21] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[22] D. L. DONOHO AND J. TANNER, *Neighborliness of randomly projected simplices in high dimensions*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 9452–9457 (electronic).

[23] M. ELAD AND A. M. BRUCKSTEIN, *A generalized uncertainty principle and sparse representation in pairs of bases*, IEEE Trans. Inform. Theory, 48 (2002), pp. 2558–2567.

[24] A. FEUER AND A. NEMIROVSKI, *On sparse representation in pairs of bases*, IEEE Trans. Inform. Theory, 49 (2003), pp. 1579–1581.

[25] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex analysis and minimization algorithms. I*, vol. 305 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993. Fundamentals.

[26] P. J. HUBER, *Robust regression: Asymptotics, conjectures and Monte Carlo*, Ann. Statist., 1 (1973), pp. 799–821.

[27] S. OSHER, M. BURGER, D. GOLDFARB, J. XU, AND W. YIN, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4 (2005), pp. 460–489 (electronic).

[28] S. OSHER, Y. MAO, B. DONG, AND W. YIN, *Fast Linearized Bregman Iteration for Compressed Sensing and Sparse Denoising*, UCLA CAM Report (08-37), 2008.

[29] M. RUDELSON AND R. VERSHYNIN, *Geometric approach to error-correcting codes and reconstruction of signals*, Int. Math. Res. Not., (2005), pp. 4019–4041.

[30] J. A. TROPP, *Just relax: Convex programming methods for identifying sparse signals in noise*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1030–1051.

[31] Y. TSAIG AND D. L. DONONO, *Extensions of compressed sensing*, Signal Processing, 86 (2005), pp. 533–548.

[32] W. YIN, S. OSHER, D. GOLDFARB, AND J. DARBON, *Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing*, SIAM J. Imaging Sci., 1 (2008), pp. 143–168.

[33] Y. ZHANG, *A simple proof for recoverability of ℓ_1-minimization: Go over or under?*, 2005. Rice University CAAM Technical Report TR05-09.

[34] Y. ZHANG, *When is missing data recoverable?*, 2006. Rice University CAAM Technical Report TR06-15.

[35] H. ZOU AND T. HASTIE, *Regularization and variable selection via the elastic net*, J. R. Stat. Soc. Ser. B Stat. Methodol., 67 (2005), pp. 301–320.