SOLVING REGULARIZED LINEAR LEAST-SQUARES PROBLEMS BY THE ALTERNATING DIRECTION METHOD WITH APPLICATIONS TO IMAGE RESTORATION[*]

JIANJUN ZHANG[†] AND BENEDETTA MORINI[‡]

Abstract. We present and analyze ways to apply the Alternating Direction Method (ADM) to bound-constrained quadratic problems including ℓ_1 and ℓ_2 regularized linear least-squares problems. The resulting ADM schemes require the solution of two subproblems at each iteration: the first one is a linear system, the second one is a bound-constrained optimization problem with closed-form solution. Numerical results on image deblurring problems are provided and comparisons are made with a Newton-based method and a first-order method for bound-constrained optimization.

Key words. Linear least-squares problems, ℓ_1 and ℓ_2 regularization, bound-constraints, alternating direction method, image deblurring.

AMS subject classifications. 65F22, 65K10, 65T50, 68U10, 90C25

1. Introduction. In this paper we address the solution of bound-constrained quadratic problems by the Alternating Direction Method (ADM) proposed originally in [10]. The problems considered are

$$(1.1) \qquad \min_{l \leq x \leq u} \frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{2}\alpha^2\|Bx\|_2^2,$$

and

$$(1.2) \qquad \min_{l \leq x \leq u} \frac{1}{2}\|Ax - b\|_2^2 + \alpha^2 e^T x,$$

where A, $B \in \mathbb{R}^{m \times n}$, $m \geq n$, $b \in \mathbb{R}^m$, $e = (1, \dots, 1)^T \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$ and the dimension n is large. The vectors l, $u \in \mathbb{R}^n$ have finite components and the inequalities $l \leq x \leq u$ are meant element-wise. Since l is finite, without loss of generality we assume it is the zero vector. Thus (1.2) is equivalent to the ℓ_1 regularized least-squares problem

$$\min_{0 \leq x \leq u} \frac{1}{2}\|Ax - b\|_2^2 + \alpha^2\|x\|_1.$$

The emphasis of our work is on problems where A and B have a specific structure and on optimization techniques capable of exploiting the structure. A motivation and relevant application are recovering images from noisy and blurry observations in image processing. In fact, (1.1) and (1.2) model image deblurring problems and A and B have structures depending on the boundary conditions used [23]. Next we assume that the quadratic functions in (1.1) and (1.2) are strictly convex, thus both problems admit an unique solution x^*.

Typical examples for our problems are imaging systems which capture an image x and return degraded data b. A common model of the degradation process is

$$b = Ax + \eta,$$

[†]Department of Mathematics, Shanghai University, Shanghai 200444, China.
(jjzhang@staff.shu.edu.cn).

[‡]Dipartimento di Ingegneria Industriale, Università di Firenze, viale G.B. Morgagni 40, 50134 Firenze, Italia.
(benedetta.morini@unifi.it)

where $b \in \mathbb{R}^m$, $\eta \in \mathbb{R}^m$ is an additive noise and A is a linear operator; e.g., a convolution by a blurring kernel followed by a subsampling. Recovering x from b is usually an ill-posed inverse problem which should be regularized by using some prior information; this leads to the minimization problem

$$\min_x \Phi_{fid}(x, b) + \alpha^2 \Phi_{reg}(x),$$

where $\Phi_{fid}(x, b)$ measures the violation of the relation between x and its observation b, $\Phi_{reg}(x)$ regularizes the solution by enforcing certain prior constraints on x, and α is a scalar weight.

In the literature, for the Gaussian noise the common fidelity term is $\Phi_{fid}(x, b) = \|Ax - b\|_2^2/2$. The use of Tikhonov regularization yields a minimization problem of the form (1.1), while an ℓ_1 regularization, favoring the recovery of a sparse solution, gives rise to the so-called *lasso* problem [28]. Interestingly, Problem (1.1) may also arise in fast total variation minimization algorithms for image restoration, see; e.g., [16]. The constraints l and u on x represent the dynamic range of the image; for example, in 8-bit gray-scale images the light intensities lie in a box with l being the zero vector and $u = 255\,e$. We also mention the ℓ_2-ℓ_1 regularization introduced in [34] and note that strict convexity of the objective function in (1.1) and (1.2) can be enforced by using ℓ_2 regularization.

A variety of iterative algorithms have been proposed for the above problems in both a general setting and image processing. In the latter context, a practical approach is to first solve the unconstrained minimization problems, and then to project or scale the unconstrained minimizer into the box $\{x \in \mathbb{R}^n : l \leq x \leq u\}$. However, in general the resulting images are of lower quality than those produced by methods which determine the solution of the constrained problems [7, 20]. This fact motivates the recent growing interest in methods for constrained optimization which provide fast solvers for image deblurring problems.

Newton-based methods for (1.1) and (1.2) such as primal dual interior point methods, affine scaling methods, conjugate gradient methods, require the solution of a linear system at each iteration, see e.g, [2, 17, 21, 19, 20, 26]. An efficient implementation of this linear algebra phase is an open problem in image restoration. In fact, due to the presence of bounds, the coefficient matrices of the systems do not preserve the properties of A and B and this fact slows down the performance of the solvers [7]. This difficulty can be overcome by gradient-based methods which avoid the solution of linear systems and are very effective when modest accuracy is required, see; e.g., [11, 12, 27, 31]. Another option is to use the alternating direction method which may be competitive with gradient-based methods although it calls for the solution of linear systems. Recently ADM has been successfully used in image processing restoration [1, 7, 24, 30, 32, 33].

The goal of this paper is to solve (1.1) and (1.2) by the alternating direction method. We propose formulations of our problems that can be tackled by ADM and analyze the resulting schemes, requiring the solution of two subproblems at each iteration. The first subproblem is a linear system which can be efficiently solved taking advantage of the structure of A and B. The second subproblem is a bound-constrained optimization problem with closed-form solution. As a result, each iteration of the procedures is cheap to perform and modest accuracy can be achieved effectively. Convergence analysis of the procedures is provided along with experimental results on image deblurring problems that illustrate the performance of ADM and compare it with a Newton-based method and a gradient-based algorithm.

The rest of the paper is organized as follows. In Section 2, we briefly review the classical alternating direction method and present its application to reformulations for ℓ_1 and ℓ_2 regularized linear least-squares. The convergence analysis of the proposed ADM is given in

Section 3. Experimental results and conclusions are presented in Sections 4 and 5, respectively.

Throughout this paper the n-dimensional box $\{x \in \mathbb{R}^n : l \leq x \leq u\}$ is also denoted as $[l, u]$ and $P(\cdot)$ is the projection map onto the box; i.e., $P(x) = \min(u, \max(x, l))$ for $x \in \mathbb{R}^n$. Given $x, y \in \mathbb{R}^n$, $\langle x, y \rangle$ denotes the inner product $x^T y$. The superscript k represents the iteration number in an algorithm. $\| \cdot \|$ denotes $\| \cdot \|_2$, the Euclidean norm.

2. Solving the regularized least-squares problems via ADM. In this section, first we briefly review the alternating direction method for linearly constrained convex programming problems with separable structure. Second, we propose reformulations of problems (1.1), (1.2) that can be tackled via ADM and analyze the resulting procedures.

ADM solves problems of the form

$$(2.1) \qquad \min_{x \in X, y \in Y} \ f_1(x) + f_2(y), \quad \text{subject to} \quad Cx + Ey = d,$$

where $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ are given convex sets, $f_1 : X \rightarrow \mathbb{R}$ and $f_2 : Y \rightarrow \mathbb{R}$ are closed proper convex functions, $C \in \mathbb{R}^{h \times n}$, $E \in \mathbb{R}^{h \times m}$ and $d \in \mathbb{R}^h$ are given, see [10]. The augmented Lagrangian function for (2.1) is

$$L(x, y, \lambda) = f_1(x) + f_2(y) + \lambda^T (Cx + Ey - d) + \frac{\beta}{2} \|Cx + Ey - d\|^2,$$

where $\lambda \in \mathbb{R}^h$ is the Lagrangian multiplier to the linear constraints and $\beta > 0$ is the penalty parameter for the violation of the linear constraints.

Given an initial λ^0, the kth iteration of the method of the Lagrangian multipliers for solving (2.1) has the form

$$\begin{cases} (x^k, y^k) = \underset{x \in X, y \in Y}{\operatorname{argmin}} \ L(x, y, \lambda^k), \\ \lambda^{k+1} = \lambda^k + \beta(Cx^k + Ey^k - d), \end{cases}$$

where $k \geq 0$ and the superscripts are the iteration counters. The update of x^k and y^k requires a joint minimization of the Lagrangian function with respect to x and y, while the update of the Lagrangian multiplier is performed by using the dual ascent method [14, 25]. In the rest of the section we will keep β fixed, but it is known that varying the penalty parameter through the iterations improves the convergence in practice.

In order to exploit the nice separable structure emerging in both the objective function and the constraint of (2.1), ADM minimizes the objective function with respect to x by fixing y and vice versa. More specifically, given $y^{-1} \in Y$ and λ^0, the kth iteration of ADM consists of the following steps

$$\begin{cases} x^k = \underset{x \in X}{\operatorname{argmin}} \ L(x, y^{k-1}, \lambda^k), \\ y^k = \underset{y \in Y}{\operatorname{argmin}} \ L(x^k, y, \lambda^k), \\ \lambda^{k+1} = \lambda^k + \beta(Cx^k + Ey^k - d). \end{cases}$$

with $k \geq 0$.

In many applications, the optimization problems providing x^k and y^k in ADM are either easily solvable or have closed-form solution. Therefore ADM iterations can be performed at a low computational cost. Taking into account that ADM can be very slow to achieve high accuracy but often provides approximate solution of modest accuracy in a limited number of iterations, ADM is attractive for many applications such as statistical and machine learning,

signal and image processing [5]. A very recent result establishes that ADM is convergent with the $O(1/N)$ rate, where N denotes the iteration number [13].

In order to apply ADM to our problems, we introduce an auxiliary variable $z \in \mathbb{R}^n$ and cast Problems (1.1), (1.2) into the form

$$(2.2) \quad \min \frac{1}{2}\|Ax - b\|^2 + \frac{1}{2}\alpha^2\|Bx\|^2, \quad \text{subject to} \quad z - x = 0, \; x \in \mathbb{R}^n, \; l \le z \le u,$$

$$(2.3) \quad \min \frac{1}{2}\|Ax - b\|^2 + \alpha^2 e^T z, \quad\quad \text{subject to} \quad z - x = 0, \; x \in \mathbb{R}^n, \; l \le z \le u,$$

respectively. The augmented Lagrangian functions for the above problems are

$$(2.4) \quad L(x, z, \lambda) = \frac{1}{2}\|Ax - b\|^2 + \frac{1}{2}\alpha^2\|Bx\|^2 + \lambda^T(z - x) + \frac{\beta}{2}\|z - x\|^2,$$

$$L(x, z, \lambda) = \frac{1}{2}\|Ax - b\|^2 + \alpha^2 e^T z + \lambda^T(z - x) + \frac{\beta}{2}\|z - x\|^2,$$

and, given $z^{-1} \in [l, u]$ and λ^0, the kth ADM iteration becomes

$$(2.5) \quad\quad\quad x^k = \operatorname*{argmin}_{x \in \mathbb{R}^n} L(x, z^{k-1}, \lambda^k),$$

$$(2.6) \quad\quad\quad z^k = \operatorname*{argmin}_{l \le z \le u} L(x^k, z, \lambda^k),$$

$$(2.7) \quad\quad\quad \lambda^{k+1} = \lambda^k + \beta(z^k - x^k).$$

By the strict convexity of (1.1) and (1.2), the two problems (2.5) and (2.6) are strictly convex and admit unique solution. We now provide details on their solution.

Algorithm 2.1: ADM FOR PROBLEM (1.1)

Given the scalar $\beta > 0$, $z^{-1} \in [l, u]$ and $\lambda^0 \in \mathbb{R}^n$.
For $k = 0, 1, 2, \ldots$
  Compute the solution x^k to Problem (2.5),

$$(2.8) \quad\quad\quad (A^T A + \alpha^2 B^T B + \beta I)x^k = A^T b + \beta z^{k-1} + \lambda^k.$$

  Compute the solution z^k to Problem (2.6),

$$(2.9) \quad\quad\quad z^k = P\left(x^k - \frac{\lambda^k}{\beta}\right).$$

  Set $\lambda^{k+1} = \lambda^k + \beta(z^k - x^k)$.

The algorithm for ADM applied to the ℓ_2 regularized problem is sketched in Algorithm 2.1. The solution x^k to Problem (2.5) solves the shifted linear systems (2.8), where $A^T A + \alpha^2 B^T B$ is the Hessian of the quadratic function in (1.1) and I denotes the identity matrix of dimension n. The solution z^k to (2.6) has the closed-form (2.9), since

$$z^k = \operatorname*{argmin}_{l \le z \le u} L(x^k, z, \lambda^k) = \operatorname*{argmin}_{l \le z \le u} \|z - x^k + \frac{\lambda^k}{\beta}\|^2 = P\left(x^k - \frac{\lambda^k}{\beta}\right),$$

with P being the projection map onto the box $[l, u]$. Proceeding analogously for (2.3), Algorithm 2.2 is obtained.

Algorithm 2.2: ADM FOR PROBLEM (1.2)

Given the scalar $\beta > 0$, $z^{-1} \in [l, u]$ and $\lambda^0 \in \mathbb{R}^n$.
For $k = 0, 1, 2, \ldots$
    Compute the solution x^k to Problem (2.5),

$$(2.10) \qquad (A^T A + \beta I)x^k = A^T b + \beta z^{k-1} + \lambda^k.$$

Compute the solution z^k to Problem (2.6),

$$z^k = P\left(x^k - \frac{\alpha^2 e}{\beta} - \frac{\lambda^k}{\beta}\right).$$

Set $\lambda^{k+1} = \lambda^k + \beta(z^k - x^k)$.

The major computational effort of the sketched algorithms is the solution of the linear systems (2.8), (2.10) at each iteration. This task can be efficiently performed both in a general setting and for a specific problem at hand. In a general setting, if a direct method is used, then a factorization of the coefficient matrix can be evaluated and cached as long as β does not change. If an iterative solver is used, say the Preconditioned Conjugate Gradient method [15], and a factorized preconditioner for $A^T A + \alpha^2 B^T B$ (or $A^T A$) is available, then a preconditioner for each shifted system can be cheaply computed by updating techniques [3, 4].

In the context of image restoration, depending on the boundary conditions, matrices $A^T A + \alpha^2 B^T B$ and $A^T A$ have specific structures and can be diagonalized by fast transforms [23]. Then, such diagonalizations can be reused through different iterations. For example, if A models out-of-focus blur, B is the gradient matrix and Neumann boundary conditions are used, then $A^T A + \alpha^2 B^T B$ is a block-Toeplitz-plus-Hankel matrix with Toeplitz-plus-Hankel blocks and it can be diagonalized via Fast Cosine Transform in $O(n \log n)$ operations [23]. For doubly symmetric point spread function, $A^T A$ can be diagonalized by 2D Discrete Cosine Transform. On the other hand, if the point spread function is not doubly symmetric, then $A^T A$ can still be diagonalized by 2D Discrete Fourier Transform using the periodic boundary conditions. The matrix $B^T B$ is diagonalizable, too.

We conclude this section with two remarks. The first one concerns formulations of (1.1) alternative to (2.2). Very recently Chan et al. [7] proposed to apply ADM to the reformulated problem

$$(2.11) \quad \min \frac{1}{2}\|Ax - b\|^2 + \frac{1}{2}\alpha^2\|Bz\|^2, \quad \text{subject to} \quad z - x = 0, \ x \in \mathbb{R}^n, \ l \le z \le u.$$

In this case, computing the iterate z^k amounts to solving a bound-constrained quadratic problem which admits closed-form solution only if $B = I$. For this reason, we consider Problem (2.2) computationally more convenient than (2.11).

Finally we remark that, for ℓ_1 and ℓ_2 regularization, our procedures can be viewed as a generalization of the Split Augmented Lagrangian Shrinkage Algorithm given in [1] to the case where the unknown is subject to bound constraints.

3. Convergence analysis. In this section, we show the convergence properties of the procedures proposed and use the saddle-point problem for the augmented Lagrangian func-

tion L:

Find $(x^*, z^*, \lambda^*) \in \mathbb{R}^n \times [l, u] \times \mathbb{R}^n$, such that

$$(3.1) \qquad L(x^*, z^*, \lambda) \leq L(x^*, z^*, \lambda^*) \leq L(x, z, \lambda^*), \ \forall (x, z, \lambda) \in \mathbb{R}^n \times [l, u] \times \mathbb{R}^n.$$

Here we report the results for the ℓ_2 regularization. The same results hold for the ℓ_1 case, and can be derived proceeding along the lines of the proofs given below. The results obtained parallel those obtained in [32].

In the rest of this section, we let q denote the quadratic function

$$q(x) = \frac{1}{2}\|Ax - b\|^2 + \frac{1}{2}\alpha^2\|Bx\|^2,$$

and $\langle x, y \rangle$ denotes the inner product in $\mathbb{R}^n$. The relation between Problems (2.2) and (3.1) is stated in the following theorem.

THEOREM 3.1. *Let L be the function in (2.4). The vector $(x^*, z^*) \in \mathbb{R}^n \times [l, u]$ solves (2.2) if and only if there exists $\lambda^* \in \mathbb{R}^n$ such that (x^*, z^*, λ^*) solves (3.1).*

Proof. Suppose that (x^*, z^*, λ^*) is a solution of (3.1). From the first inequality in (3.1), we have $z^* = x^*$. This relation, together with the second inequality in (3.1), gives

$$(3.2) \qquad q(x^*) \leq q(x) + \langle \lambda^*, z - x \rangle + \frac{\beta}{2}\|z - x\|^2, \ \forall \, (x, z, \lambda) \in \mathbb{R}^n \times [l, u] \times \mathbb{R}^n.$$

Taking $x = z \in [l, u]$ in (3.2), it follows that x^* is the solution of (1.1), and accordingly, (x^*, z^*) is the solution of (2.2).

Let us now suppose that (x^*, z^*) solves (2.2). The first inequality in (3.1) trivially follows. To complete the proof, we first note that the first-order optimality condition for Problem (1.1) is

$$(3.3) \quad \langle (A^T A + \alpha^2 B^T B)x^* - A^T b, z - x^* \rangle = \langle \nabla q(x^*), z - x^* \rangle \geq 0, \quad \forall z \in [l, u].$$

Then Problem (2.2) admits a unique solution (x^*, z^*) such that

$$(3.4) \qquad\qquad (A^T A + \alpha^2 B^T B)x^* - A^T b - \lambda^* = 0,$$

$$(3.5) \qquad\qquad \langle \lambda^*, z - z^* \rangle \geq 0, \quad \forall z \in [l, u],$$

$$z^* - x^* = 0,$$

with $\lambda^* \in \mathbb{R}^n$. To show this fact, note that by $x^* = z^*$, inequality (3.5) becomes

$$\langle \lambda^*, z - x^* \rangle \geq 0, \quad \forall z \in [l, u],$$

and using (3.4) we obtain (3.3).

Let now $\lambda^* \in \mathbb{R}^n$ be the vector in (3.4). The function $L(x, z, \lambda^*)$, with $(x, z) \in \mathbb{R}^n \times [l, u]$, is strictly convex and any stationarity point $(\tilde{x}, \tilde{z}) \in \mathbb{R}^n \times [l, u]$ satisfies

$$(A^T A + \alpha^2 B^T B + \beta I)\tilde{x} = A^T b + \beta \tilde{z} + \lambda^*,$$

$$\langle \lambda^* + \beta(\tilde{z} - \tilde{x}), z - \tilde{z} \rangle \geq 0, \quad \forall z \in [l, u].$$

Since (x^*, z^*) satisfies these conditions, the proof is completed.   □

The next lemma establishes conditions for finite termination of ADM.

LEMMA 3.2. *If*

$$(3.6) \qquad\qquad z^k = z^{k-1} \quad and \quad x^k = z^k,$$

then $x^k = x^*$ where x^* is the unique solution of (1.1).

    Proof. By (2.5) and the convexity of $L(x, z^{k-1}, \lambda^k)$, we have

$$(3.7) \qquad q(x) - q(x^k) + \langle -\lambda^k + \beta(x^k - z^{k-1}), x - x^k \rangle \geq 0, \ \ \forall\, x \in \mathbb{R}^n,$$

see also [9, p. 170], while by (2.6) it follows

$$(3.8) \qquad \langle \lambda^k + \beta(z^k - x^k), z - z^k \rangle \geq 0, \ \ \forall\, z \in [l, u].$$

Then, letting x^* be the solution to (1.1), (3.6) gives

$$q(x^*) - q(x^k) + \langle -\lambda^k, x^* - x^k \rangle \geq 0,$$
$$\langle \lambda^k, x^* - x^k \rangle \geq 0;$$

i.e., $q(x^*) - q(x^k) \geq \langle \lambda^k, x^* - x^k \rangle \geq 0$. Hence, $x^k \in [l, u]$ and $q(x^k) \leq q(x^*)$ shows that $x^k = x^*$. $\qquad\square$

    The main convergence results is given below.

    THEOREM 3.3. *Let L be the function in (2.4) and (x^*, z^*, λ^*) be a saddle-point of L. Then the sequence (x^k, z^k) generated by Algorithm 2.1 satisfies*

$$\lim_{k \to \infty} (x^k, z^k) = (x^*, z^*).$$

    Proof. Let us define $\bar{x}^k$, $\bar{z}^k$ and $\bar{\lambda}^k$ as

$$\bar{x}^k = x^k - x^*, \ \ \bar{z}^k = z^k - z^*, \ \ \bar{\lambda}^k = \lambda^k - \lambda^*.$$

Then (2.7) gives $\bar{\lambda}^{k+1} = \bar{\lambda}^k + \beta(z^k - x^k)$ and consequently

$$(3.9) \qquad \|\bar{\lambda}^k\|^2 - \|\bar{\lambda}^{k+1}\|^2 = -2\beta\langle \bar{\lambda}^k, z^k - x^k \rangle - \beta^2\|z^k - x^k\|^2.$$

    Since (x^*, z^*, λ^*) is a saddle-point of $L(x, z, \lambda)$, by Theorem 3.1 we have $z^* = x^*$. Moreover by [9, p. 170] we have the following characterization of (x^*, z^*),

$$(3.10) \qquad q(x) - q(x^*) + \langle -\lambda^*, x - x^* \rangle \geq 0, \ \ \forall\, x \in R^n$$
$$(3.11) \qquad \langle \lambda^*, z - z^* \rangle \geq 0, \ \ \forall z \in [l, u]$$

Taking $x = x^k$ in (3.10), $z = z^k$ in (3.11), $x = x^*$ in (3.7) and $z = z^*$ in (3.8), we obtain by addition

$$\langle \bar{\lambda}^k, z^k - x^k \rangle + \beta\|x^k - z^k\|^2 + \beta\langle z^k - z^{k-1}, \bar{x}^k \rangle \leq 0.$$

Then by (3.9)

$$(3.12) \qquad \|\bar{\lambda}^k\|^2 - \|\bar{\lambda}^{k+1}\|^2 \geq \beta^2\|z^k - x^k\|^2 + 2\beta^2\langle z^k - z^{k-1}, \bar{x}^k \rangle.$$

We now provide a lower bound for $\langle z^k - z^{k-1}, \bar{x}^k \rangle$ in (3.12) using the equation

$$(3.13) \qquad \langle z^k - z^{k-1}, \bar{x}^k \rangle = \langle z^k - z^{k-1}, x^k - z^{k-1} \rangle + \langle z^k - z^{k-1}, \bar{z}^{k-1} \rangle.$$

First we note that by using (2.7) we get

$$\langle z^k - z^{k-1}, x^k - z^{k-1} \rangle = \langle z^k - z^{k-1}, (x^k - \frac{\lambda^k}{\beta}) - (x^{k-1} - \frac{\lambda^{k-1}}{\beta}) \rangle.$$

By the definition of z^k in (2.9) and of the projection map P, we have

$$\langle z^{k-1} - z^k, x^k - \frac{\lambda^k}{\beta} - z^k \rangle \leq 0,$$

$$\langle z^{k-1} - z^k, z^{k-1} - x^{k-1} + \frac{\lambda^{k-1}}{\beta} \rangle \leq 0,$$

and summing these inequalities, we have

$$\|z^{k-1} - z^k\|^2 \leq \langle z^k - z^{k-1}, (x^k - \frac{\lambda^k}{\beta}) - (x^{k-1} - \frac{\lambda^{k-1}}{\beta}) \rangle.$$

Thus, (3.13) takes the form

$$\langle z^k - z^{k-1}, \bar{x}^k \rangle \geq \|z^{k-1} - z^k\|^2 + \langle z^k - z^{k-1}, \bar{z}^{k-1} \rangle$$

$$= \|\bar{z}^{k-1} - \bar{z}^k\|^2 + \frac{1}{2}(\|\bar{z}^k\|^2 - \|\bar{z}^{k-1}\|^2 - \|\bar{z}^k - \bar{z}^{k-1}\|^2)$$

$$(3.14) \qquad = \frac{1}{2}(\|\bar{z}^k\|^2 - \|\bar{z}^{k-1}\|^2 + \|z^k - z^{k-1}\|^2),$$

and (3.12) and (3.14) yield

$$(\|\bar{\lambda}^k\|^2 + \beta^2\|\bar{z}^{k-1}\|^2) - (\|\bar{\lambda}^{k+1}\|^2 + \beta^2\|\bar{z}^k\|^2) \geq \beta^2\|z^k - x^k\|^2 + \beta^2\|z^k - z^{k-1}\|^2.$$

Lemma 3.2 indicates that $\|z^k - x^k\|^2 + \|z^k - z^{k-1}\|^2 > 0$, unless $x^k = z^k = x^*$. Then the sequence $\{\|\bar{\lambda}^k\|^2 + \beta^2\|\bar{z}^{k-1}\|^2\}$ is monotonically decreasing and bounded below and we can conclude that it is convergent. Moreover, the sequences $\{\lambda^k\}$ and $\{z^k\}$ are bounded and

$$(3.15) \qquad \lim_{k\to\infty} \|z^k - x^k\| = 0, \qquad \lim_{k\to\infty} \|z^k - z^{k-1}\| = 0.$$

Now since (x^*, z^*, λ^*) is a saddle-point of $L(x, z, \lambda)$, by the second inequality in (3.1) we have

$$(3.16) \qquad q(x^*) \leq q(x^k) + \langle \lambda^*, z^k - x^k \rangle + \frac{\beta}{2}\|z^k - x^k\|^2.$$

Further, summing (3.7) with $x = x^*$ and (3.8) with $z = z^*$ we obtain

$$(3.17) \qquad q(x^*) \geq q(x^k) + \langle \lambda^k, z^k - x^k \rangle + \beta\|z^k - x^k\|^2 + \beta\langle z^k - z^{k-1}, \bar{x}^k \rangle.$$

Hence, by taking $\liminf$ in (3.16), $\limsup$ in (3.17) and using (3.15) we have

$$\liminf_{k\to\infty} q(x^k) \geq q(x^*) \geq \limsup_{k\to\infty} q(x^k),$$

which leads to $\lim_{k\to\infty} q(x^k) = q(x^*)$. Since $q(x)$ is continuous and has a unique minimizer in $[l, u]$, we have $\lim_{k\to\infty} x^k = x^*$. Then (3.15) gives $\lim_{k\to\infty} z^k = z^*$ which completes the proof.
☐

4. Overview of the algorithms used for the comparison. The remainder of this paper is devoted to test Algorithms 2.1 and 2.2 and to consider their effectiveness with respect to algorithms from other classes of procedures for bound-constrained optimization.

The alternating direction method was compared with three algorithms. The first, denoted Projection (P) method, solves the least-square problem dropping the constraints and then

projects the unconstrained solution onto the box $[l, u]$. The second solver is the Reduced Newton (RN) method proposed and tested on image restoration problems in [20]. The third method is the Affine Scaling Cyclic Barzilai-Borwain (AS_CBB) method proposed and tested on image restoration problems in [11].

RN and AS_CBB methods belong to the framework of the affine scaling methods for bound-constrained optimization and generate strictly feasible iterates throughout the process. For simplicity, we will review the methods when applied to Problem (1.1) but the application to (1.2) can be easily derived. A detailed description of RN and AS_CBB and of their theoretical properties can be found in [20] and [11].

Let $g(x) = A^T(Ax - b) + \alpha^2 B^T Bx$ be the gradient of the objective function in (1.1), $g_i(x)$ and x_i represent the ith component of g and x, respectively. The first-order optimality condition for (1.1) can be formulated as

$$D(x)g(x) = 0, \tag{4.1}$$

where $D(x) = \mathrm{diag}(d_1(x), \ldots, d_n(x))$ has entries

$$d_i(x) = u_i - x_i, \quad \text{if } g_i(x) < 0,$$
$$d_i(x) = x_i - l_i, \quad \text{if } g_i(x) > 0,$$
$$d_i(x) = \min\{x_i - l_i, u_i - x_i\} \text{ otherwise.}$$

Let x^k be a strictly feasible iterate; i.e., $l < x^k < u$, $k \geq 0$. Applying the Newton method to (4.1) requires solving one linear system at each iteration. At the kth iteration the linear system takes the form

$$(D(x^k)(A^T A + \alpha^2 B^T B) + E(x^k))p^k = -D(x^k)g(x^k), \tag{4.2}$$

where the coefficient matrix is obtained by formal application of the product rule and E is a diagonal positive semidefinite matrix.

Clearly, handling the bounds makes the solution of (4.2) difficult. In fact, the matrix $D(x^k)(A^T A + \alpha^2 B^T B) + E(x^k)$ does not preserve the structure of A and B. Moreover, if x^k approaches a degenerate solution of the problem, this matrix tends to become singular. On the other hand, the following linear system equivalent to (4.2),

$$M(x^k)p^k = -g(x^k), \quad M(x^k) = A^T A + \alpha^2 B^T B + D(x^k)^{-1}E(x^k), \tag{4.3}$$

tends to become singular as x^k approaches a nondegenerate solution of (1.1).

The RN method considers (4.3) and exploits an active set strategy to overcome the above mentioned problems and to reduce the dimension of the system. In particular, at each iteration RN identifies the active set by checking the closeness of x^k to the boundary. Then, for the component of x^k in the active set, the step to the nearest bound is taken while the linear system (4.3) restricted to the inactive components of x^k is solved. Given the inactive set $\mathcal{I}_k$ at x^k, the coefficient matrix $(M(x^k))_{\mathcal{I}_k \mathcal{I}_k}$ of the system to be solved is the submatrix of $M(x^k)$ with elements having row and column index in $\mathcal{I}_k$. This submatrix is symmetric and positive definite, it is better conditioned than $M(x^k)$, its inverse is uniformly bounded for any k, and its dimension may be considerably smaller than the original system. Finally, strict feasibility of the iterates is enforced projecting the step onto the box and taking a large fraction of it. The RN method is locally quadratic convergent to the solution x^* of (1.1) even in the presence of degeneracy and shown to be reliable in the solution of image restoration methods.

Basically the computational cost of RN in each iteration amounts to the solution of one linear system of dimension equal to the cardinality of $\mathcal{I}_k$. The matrices $(M(x^k))_{\mathcal{I}_k \mathcal{I}_k}$ are

typically smaller than the matrices used in ADM but they do not preserve their structure. Such loss of structure may result in loss of efficiency in the solution of the linear systems with respect to ADM.

The AS_CBB method is a first-order method and avoids the solution of linear systems. In (4.2) it replaces the coefficient matrix by the diagonal matrix $(\mu_k D(x^k) + E(x^k))$, where μ_k is a positive scalar computed by a Quasi-Newton rule. Thus, the ith component $(p^k)_i$ of the step p^k is given by

$$(p^k)_i = - \left(\frac{1}{\mu_k + |g_i(x^k)|/d_i(x^k)} \right) g_i(x^k).$$

The positive scalar μ_k is computed using a cyclic version of the Barzilai-Borwein (BB) step-size rule. Specifically, let $\mu_0^{BB} = \max(\bar{\mu}, \|g(x^0)\|_\infty)$ and

$$\mu_k^{BB} = \arg \min_{\mu \geq \bar{\mu}} \|\mu s^{k-1} - y^{k-1}\| = \max \left\{ \bar{\mu}, \frac{(s^{k-1})^T y^{k-1}}{(s^{k-1})^T s^{k-1}}, \right\}, \quad k \geq 1$$

where $\bar{\mu}$ is a fixed positive parameter, $s^{k-1} = x^k - x^{k-1}$, $y^{k-1} = g(x^k) - g(x^{k-1})$. The cyclic BB strategy consists of re-using the BB stepsize for several iterations. Namely, letting $c \geq 1$ be the cycle length and $l \geq 0$ be the cycle number, the value of the scalars μ_k is assigned by the rule $\mu_{cl+i} = \mu_{cl+1}^{BB}$, $i = 1, \ldots, c$.

Once p^k is computed, the AS_CBB method generates a new iterate of the form $x^{k+1} = x^k + \zeta_k p^k$, where the stepsize $\zeta_k \in (0, 1]$ is computed by a non-monotone line-search strategy. The generated sequence is strictly feasible and converges R-linearly to a nondegenerate solution of (1.1).

The computational effort of AS_CBB at the kth iteration amounts to one gradient evaluation and to a number of objective function evaluations equal to the number of backtracks performed. Hence the computational cost can be monitored by the total number of iterations and backtracks performed.

5. Experimental results. In this section, we show the performance of procedures P, RN, AS_CBB and Algorithms 2.1, 2.2. Computations were performed in double precision using MATLAB 7.12.0.635 (R2011a) on an Intel(R) Core(TM) i7-2600 CPU @3.40 GHz, 4.00 GB RAM.

Eight 256-by-256 gray-scale images shown in Figure 5.1 were considered.. The Satellite image is from the US Air Force Phillips Laboratory and it is contained in the image restoration software package [22]; the Clock image is taken from the USC-SIPI image database [29], the remaining images are from the Berkeley Segmentation Dataset [18]. The dimensions of the least-squares problems are $m = n = 65536$ and the constraints are $l = (0, \ldots, 0)^T$ and $u = (255, \ldots, 255)^T$.

We choose the blurring matrix A to be the out-of-focus blur with radius 3 and the regularization matrix B to be the gradient matrix. Thus, $B^T B$ is the two-dimensional discrete Laplacian matrix. For both matrices, we employed Neumann boundary conditions, which usually gives less artifacts at the boundary. Hence, $A^T A + \alpha^2 B^T B$ (or $A^T A$) is a block-Toeplitz-plus-Hankel matrix with Toeplitz-plus-Hankel blocks and the multiplication of this structured matrix times a vector can be done via fast Cosine transform in $O(n \log n)$ operations [23]. The observed image was defined as $b = Ax + \eta r$, where x is the true image, r is a random vector with elements distributed as standard normal and η is the level of noise. Four levels of noise, $\eta = 1, 3, 5, 7$ were tested.

FIG. 5.1. *The true images.*

The peak signal-to-noise ratio (PSNR) [6] is used to give quantitative performance measures

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{\dfrac{1}{n} \sum_{i=1}^{n} (\tilde{x}_i - x_i)^2},$$

where $\tilde{x}_i$ and x_i denote the pixel values of the restored image and the original image, respectively. The regularization term α in (1.1) and (1.2) is fixed by trial and error along with the parameter β in ADM, which is kept constant through the iteration process.

Concerning the implementation of the algorithms compared, the parameters in RN and AS_CBB methods are set as suggested in [20] and [11]. The initial guess x^0 for RN and AS_CBB is the same; since x^0 must be strictly feasible, it is formed by projecting the noise image b onto the box $[e, 254e]$. For a fair comparison, in ADM we set $z^{-1} = x^0$, $\lambda^0 = (0, \ldots, 0)^T$. The linear systems in RN are solved by the Conjugate Gradient method and the structure of A and B is exploited in the computation of matrix-vector products. On the other hand, the ADM implementation fully exploits the properties of the coefficient matrices; the matrix $A^T A + B^T B$ (or $A^T A$) is diagonalized by the discrete cosine transform matrix and this diagonalization is reused through the iterations.

A maximum number of 500 iterations are allowed for RN, AS_CBB and ADM. Further, the iterations of RN and AS_CBB are terminated when the distance between two successive iterations is below the fixed relative tolerance $\tau = 10^{-4}$; i.e.,

$$\|x^{k+1} - x^k\| \le \tau \|x^{k+1}\|.$$

The accuracy requirement on the feasible iterate z^k generated by ADM is

$$\|x^k - z^k\| \le \tau \|z^k\|.$$

A first set of experiments was conducted solving (1.1) by the P, RN, AS_CBB and ADM algorithms. In Table 5.1 we report the PSNR values of the observed images and of the images recovered by the four algorithms tested. In Table 5.2 we show the elapsed times in seconds required by the procedures compared.

TABLE 5.1

PSNR values of the observed images and of the images recovered by P, RN, AS_CBB and ADM algorithms applied to Problem (1.1).

		PSNR				
Image	η	Observed Image	P	RN	AS_CBB	ADM
Satellite	1	25.24	30.38	34.58	34.75	34.71
	3	25.06	27.78	30.05	30.06	30.06
	5	24.73	26.92	28.37	28.38	28.38
	7	24.29	26.62	27.68	27.69	27.69
Church	1	25.18	30.06	30.64	30.87	31.01
	3	25.01	27.41	27.41	27.96	28.28
	5	24.69	26.51	26.98	26.99	27.46
	7	24.23	26.06	26.50	26.52	26.93
Eagle	1	29.17	32.50	33.73	34.19	34.70
	3	28.76	30.09	31.14	31.15	32.55
	5	28.02	29.45	30.43	30.45	31.68
	7	27.09	28.71	29.51	29.53	30.98
Bridge	1	21.15	27.29	28.08	28.02	28.31
	3	21.07	23.79	24.44	24.49	24.60
	5	20.93	23.07	23.52	23.52	23.70
	7	20.74	22.37	22.69	22.68	22.91
Clock	1	25.36	29.53	29.54	29.93	30.23
	3	25.19	26.89	26.90	26.96	27.59
	5	24.84	26.04	26.05	26.05	26.99
	7	22.37	25.62	25.62	25.64	26.50
Surf	1	22.70	28.41	29.53	29.48	29.71
	3	22.59	25.36	25.88	25.89	26.13
	5	22.42	24.29	24.79	24.81	25.02
	7	22.16	23.99	24.28	24.29	24.56
Zebra	1	19.64	26.30	26.51	26.21	26.59
	3	19.59	22.80	22.87	22.88	22.96
	5	19.49	21.93	21.96	21.96	22.06
	7	19.36	21.10	21.12	21.12	21.31
Bear	1	20.39	27.20	27.69	27.37	27.83
	3	20.34	23.67	23.88	23.88	24.07
	5	20.21	22.88	23.08	23.09	23.23
	7	20.05	22.10	22.24	22.45	22.45

The results in Table 5.1 show that RN, AS_CBB and ADM algorithms produce images of higher quality than those obtained by using the P method. Enforcing the bounds throughout the iterations offers a significant advantage over the P method since an increase of 1dB in the PSNR value translates roughly to 10% reduction in the relative error between the true and restored images. ADM achieves the highest restored quality in 27 tests out of 32 but in 17 runs the gain over RN and AS_CBB is within 0.3dB.

From Table 5.2 we observe that ADM is much faster than RN in all runs. Moreover, comparing to AS_CBB the execution time of ADM is more than halved for 27 problems while AS_CBB is the winner in two runs.

Let us now make some comments on the good performance of ADM. ADM shows a slower convergence rate than RN and consequently requires a higher number of linear sys-

TABLE 5.2
Execution times of RN, AS_CBB and ADM algorithms applied to Problem (1.1).

			Algorithm	
Image	η	RN	AS_CBB	ADM
Satellite	1	27.67	10.28	0.80
	3	19.25	6.76	1.29
	5	14.87	3.49	1.58
	7	15.97	4.56	0.42
Church	1	28.83	4.32	0.21
	3	12.58	3.19	0.17
	5	8.65	1.58	0.19
	7	8.84	1.16	0.32
Eagle	1	16.89	2.93	0.32
	3	8.43	1.38	0.19
	5	8.40	1.42	0.19
	7	9.66	1.34	1.18
Bridge	1	29.06	6.93	0.40
	3	21.38	4.82	3.23
	5	12.67	6.78	0.34
	7	14.13	5.55	0.89
Clock	1	20.75	3.00	0.04
	3	11.80	3.36	0.07
	5	7.39	1.83	3.31
	7	7.87	1.87	0.09
Surf	1	29.52	5.76	0.39
	3	12.64	3.17	0.34
	5	13.52	3.35	0.30
	7	9.30	1.52	0.34
Zebra	1	22.22	7.11	0.34
	3	20.42	6.25	0.30
	5	12.00	3.45	0.52
	7	12.15	4.06	0.34
Bear	1	32.82	7.30	0.35
	3	13.29	3.06	7.35
	5	14.68	3.81	3.28
	7	9.40	1.88	0.26

tems solves. However, this disadvantage is alleviated by the fact that the diagonalization of the coefficient matrix in ADM is reused through different iterations and the numerical results confirm that the overhead of RN in the linear algebra phase is not compensated by fast convergence.

The computational overhead of AS_CBB depends on the number of iterations and function evaluations required. In order to compare the number of iterations of AS_CBB and ADM, in Figure 5.2 we display the performance profile [8]. The performance profile is defined as follows. Consider the 32 tests performed and the two solvers ADM and AS_CBB. For each test t solved by the solver s, let $I_{s,t}$ denote the number of iterations required and let $\tilde{I}_t$ be the smallest number of iterations required by the two solvers in the solution of test t. Then, the ratio $i_{s,t} = \dfrac{I_{s,t}}{\tilde{I}_t}$, measures the performance on test t by solver s with respect to the better
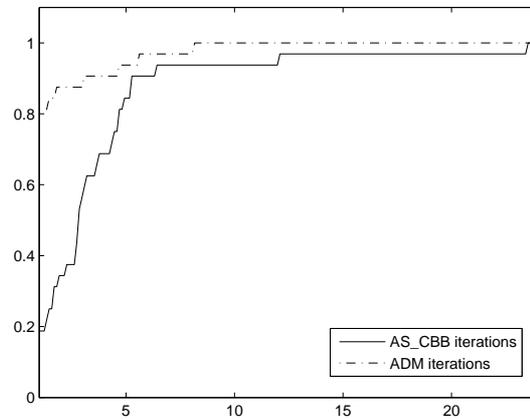
FIG. 5.2. *Performance profile: number of AS_CBB and ADM iterations for Problem (1.1).*

performance on such test and the performance profile of solver s is defined as

$$\pi_s(\tau) = \frac{no.\ of\ tests\ s.t.\ i_{s,t} \le \tau}{total\ no.\ of\ tests}, \qquad \tau \ge 1.$$

The performance profile is given in Figure 5.2; the left side of the plot gives the percentage of test problems for which the solver is the more efficient. It shows that ADM requires a lower number of iterations than AS_CBB in 80% of the runs and that AS_CBB is within a factor 5 with respect to ADM for 90% of the tests.

Besides the fact the ADM outperforms AS_CBB in terms of iterations, we also point out that, in all runs, the number of quadratic function evaluations required by AS_CBB varies between a factor 1.2 and 2.2 with respect to the number of iterations. Since performing one iteration of ADM is very cheap, the above analysis supports the effectiveness of ADM in terms of computational time.

We conclude giving results obtained by using the regularized Problem (1.2) and ADM on images: Satellite, Eagle, and Clock. The results reported in Table 5.3 show that ADM solves (1.2) efficiently. In Figure 5.3 we show the satellite images recovered by ADM applied to Problems (1.1) and (1.2).

6. Conclusions. We have proposed the solution of ℓ_1 and ℓ_2 as bound-constrained linear least-squares problems by ADM. The procedures proposed allow us to exploit the specific structure of the matrices appearing in the problems and are suitable for recovering images from noisy and blurry observations in image processing. Experiments on image deblurring problems show that ADM is effective in terms of quality of the restored images and speed, and compares favorably to existing procedures for large bound-constrained linear least-squares problems.

TABLE 5.3
Computational results for ADM applied to Problem (1.2).

Image	η	ADM		
		PSNR Observed Image	PSNR	Time
Satellite	1	25.24	34.75	0.82
	3	25.06	30.28	0.99
	5	24.73	28.50	1.11
	7	24.29	27.63	1.11
Eagle	1	29.17	34.33	0.29
	3	28.76	31.72	0.28
	5	28.02	30.25	0.31
	7	27.09	28.84	0.33
Clock	1	25.36	29.55	0.09
	3	25.19	27.14	0.11
	5	24.84	26.23	0.11
	7	22.37	25.41	0.15

REFERENCES

[1] M. AFONSO, J. BIOUCAS-DIAS, AND M. FIGUEIREDO, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Image Process., 19 (2010), pp. 2345–2356.
[2] S. BELLAVIA, M. MACCONI, AND B. MORINI, *An interior point Newton-like method for nonnegative least-squares problems with degenerate solution*, Numer. Linear Algebra Appl., 13 (2006) pp. 825–846.
[3] S. BELLAVIA, V. DE SIMONE, D. DI SERAFINO, AND B. MORINI, *Efficient preconditioner updates for shifted linear systems*, SIAM J. Sci. Comput., 33 (2011), pp. 1785–1809.
[4] ———, *A preconditioning framework for sequences of diagonally modified linear systems arising in optimization*, SIAM J. Numer. Anal., 50 (2012), pp. 3280–3302.
[5] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2010), pp. 1–122.
[6] A. BOVIK, *Handbook of Image and Video Processing*, Academic Press, Orlando, 2000.
[7] R. H. CHAN, M. TAO, AND X. YUAN, *Linearized alternating method for constrained least-squares problem*, East Asian J. Appl. Math., to appear, 2013.
[8] E. D. DOLAN AND J. J. MORÉ, *Benchmarking optimization software with performance profiles*, Math. Program., 91 (2002), pp. 201–213.
[9] I. EKELAND AND R. TÉMAM, *Convex Analysis and Variational Problems*, SIAM, Philadelphia, 1999.
[10] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.
[11] W. HAGER, B. MAIR, AND H. ZHANG, *An affine-scaling interior-point CBB method for box-constrained optimization*, Math. Program., 119 (2009), pp. 1–32.
[12] W. HAGER, D. PHAN, AND H. ZHANG, *Gradient-based methods for sparse recovery*, SIAM J. Imaging Sci., 4 (2011), pp. 146–165.
[13] B. HE AND X. YUAN, *On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.
[14] M. R. HESTENES, *Multiplier and gradient methods*, J. Optimization Theory Appl., 4 (1969), pp. 303–320.
[15] M. R. HESTENES AND E. L. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand., 49 (1952), pp. 409–436.
[16] Y. HUANG, M. NG, AND Y. WEN, *A fast total variation minimization method for image restoration*, Multiscale Model. Simul., 7 (2008), pp. 774–795.
[17] S. J. KIM, K. KOH, M. LUSTIG, S. BOYD, AND D. GORINEVSKY, *An interior-point method for large-scale l^1-regularized least squares*, IEEE J. Select. Top. Signal Process., 1 (2007), pp. 606–617.
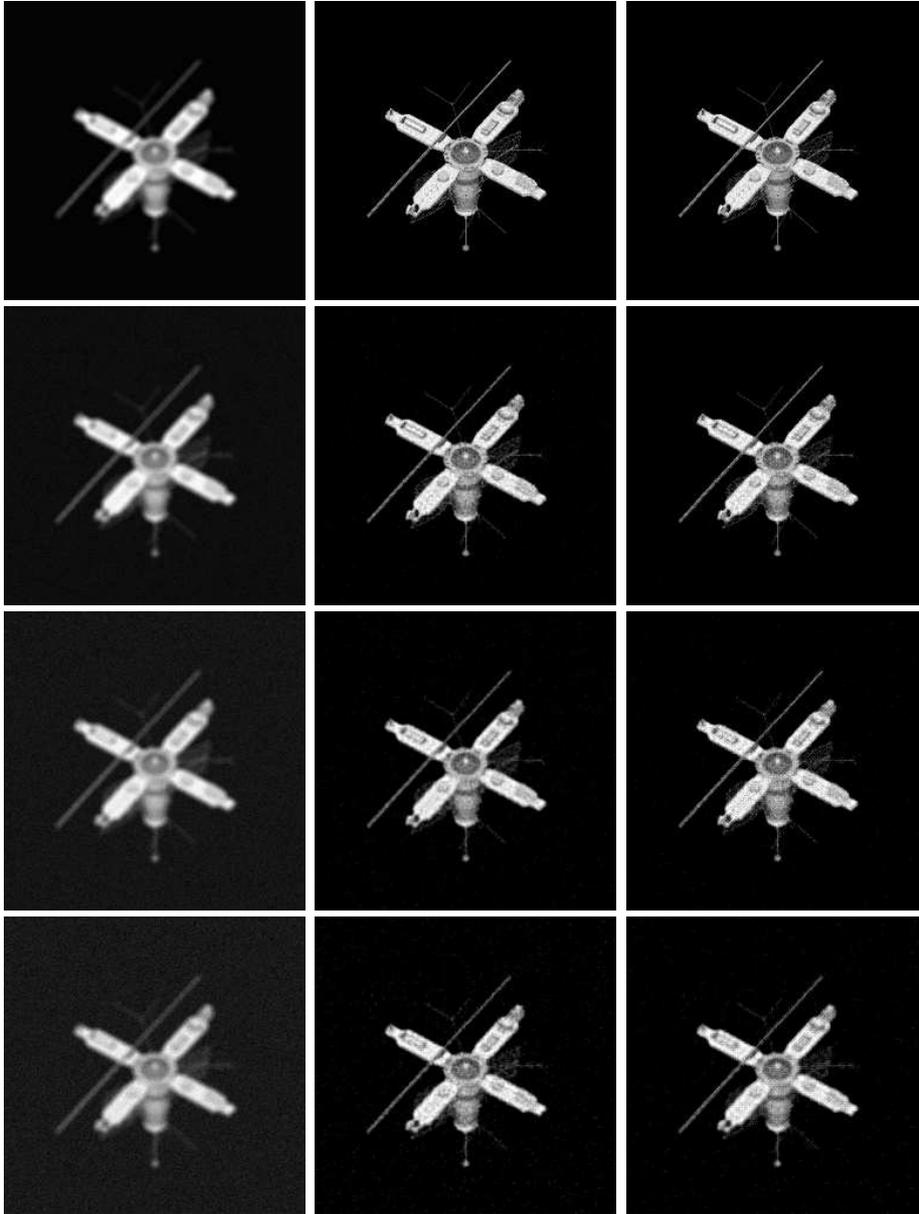
FIG. 5.3. *Observed satellite image and images recovered by ADM. From top row to bottom: level of noise* $\eta = 1, 3, 5, 7$. *On each row: observed image (left), image recovered by ADM applied to the* ℓ_2 *regularized least-squares problem (middle), image recovered by ADM applied to the* ℓ_1 *regularized least-squares problem (right).*

[18] D. MARTIN, C. FOWLKES, D. TAL, AND J. MALIK, *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*, in Proceedings of the Eighth IEEE International Conference on Computer Vision ICCV 2001, IEEE Conference Proceedings, Los Alamitos, CA, 2001, pp. 416–423.

[19] J. J. MORÉ AND G. TORALDO, *On the solution of large quadratic programming problems with bound constraints*, SIAM J. Optim., 1 (1991), pp. 93–113.

[20] B. MORINI, M. PORCELLI, AND R. H. CHAN, *A reduced Newton method for constrained linear least-squares problems*, J. Comput. Appl. Math., 233 (2010), pp. 2200–2212.

[21] S. MORIGI, L. REICHEL, F. SGALLARI, AND F. ZAMA, *An iterative method for linear discrete ill-posed problems with box constraints*, J. Comput. Appl. Math., 198 (2007) pp. 505–520.

[22] J. G. NAGY, K. PALMER, AND L. PERRONE, *Iterative methods for image deblurring: a Matlab object oriented approach*, Numer. Algorithms, 36 (2004), pp. 73–93.

[23] M. NG, R. CHAN, AND W. TANG, *A fast algorithm for deblurring models with Neumann boundary condition*, SIAM J. Sci. Comput., 21 (1999), pp. 851–866.

[24] M. NG, P. WEISS, AND X. YUAN, *Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods*, SIAM J. Sci. Comput., 32 (2010), pp. 2710–2736.

[25] M. J. D. POWELL, *A method for nonlinear constraints in minimization problems*, in Optimization: Symposium of the Institute of Mathematics and Its Applications, University of Keele, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.

[26] M. A. SAUNDERS, *PDCO, Primal-Dual interior method for Convex Objectives*, Stanford Dept. of Management Science and Engineering, Stanford.
http://www.stanford.edu/group/SOL/software/pdco.html

[27] T. SERAFINI, G. ZANGHIRATI, AND L. ZANNI, *Gradient projection methods for quadratic programs and applications in training support vector machines*, Optim. Methods Softw., 20 (2005), pp. 353–378.

[28] R. TIBSHIRANI, *Regression selection and shrinkage via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267–288.

[29] USC, *The USC-SIPI Image Database*, Signal and Image Processing Institute, USC, Los Angeles.
http://sipi.usc.edu/database/

[30] Y. WANG, J. YANG, W. YIN, AND Y. ZHANG, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imaging Sci., 1 (2008), pp. 248–272.

[31] S. WRIGHT, R. NOWAK, AND M. FIGUEIREDO, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493.

[32] C. L. WU AND X.C. TAI, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM J. Imaging Sci., 3 (2010), pp. 300–339.

[33] J. YAN, W. YIN, Y. ZHANG, AND Y. WANG, *Fast algorithm for edge-preserving variational multichannel image restoration*, SIAM J. Imaging Sci., 2 (2009), pp. 569–592.

[34] H. ZOU, T. HASTIE, AND R. TIBSHIRANI, *Sparse principal component analysis*, J. Comput. Graph. Statist., 15 (2006), pp. 265–286.