1. Introduction. The study of linear inverse problems gives rise to a linear equation of the form

\[ Ax = b, \]

where \( A : \mathcal{X} \to \mathcal{H} \) is a bounded linear operator between two Hilbert spaces \( \mathcal{X} \) and \( \mathcal{H} \).

In applications, we will not simply consider (1.1) alone, instead we will incorporate a priori available information on solutions into the problem. Assume that we have a priori information on a feature, such as sparsity, of the sought solution under a suitable transform \( W \) from \( \mathcal{X} \) to another Hilbert space \( \mathcal{Y} \) with domain \( \mathcal{D}(W) \). We may then take a convex function \( f : \mathcal{Y} \to (-\infty, \infty) \) to capture such a feature. This leads us to consider the convex minimization problem

\[
\begin{aligned}
\text{minimize} & \quad f(Wx) \\
\text{subject to} & \quad Ax = b, \quad x \in \mathcal{D}(W),
\end{aligned}
\]

where \( f \) and \( W \) should be specified during applications. For inverse problems, the operator \( A \) usually is either noninvertible or ill conditioned with a huge condition number. Thus, a small perturbation of the data may lead the problem (1.2) to have no solution; even if it has a solution, this solution may not depend continuously on the data due to the uncontrollable amplification of noise. In order to overcome such
ill-posedness, regularization techniques should be taken into account to produce a reasonable approximate solution from noisy data. One may refer to [17, 30, 45] for comprehensive accounts on the variational regularization methods as well as iterative regularization methods.

Variational regularization methods typically consider (1.2) by solving a family of well-posed minimization problems,

\[
\min_{x \in \mathcal{D}(W)} \left\{ \frac{1}{2} \|Ax - b\|^2 + \alpha f(Wx) \right\},
\]

where \( \alpha > 0 \) is the so called regularization parameter whose choice crucially affects the performance of the method. When the regularization parameter \( \alpha \) is given, many efficient solvers have been developed to solve (1.3) when \( f \) are sparsity promoting functions. However, to find a good approximate solution, the regularization parameter \( \alpha \) should be carefully chosen; consequently one has to solve (1.3) for many different values of \( \alpha \), which can be time consuming.

Among algorithms for solving (1.3), the alternating direction method of multipliers (ADMM) is a favorable one. The ADMM was proposed in [22, 24] around the mid-1970s and was analyzed in [16, 21, 36, 43]. It has been widely used in solving structured optimization problems due to its decomposability and superior flexibility. Recently, it has been revisited and popularized in modern signal/image processing, statistics, machine learning, and so on; see the recent survey paper [4] and the references therein. Due to the popularity of ADMM and its variants, new and refined convergence results have been obtained from several different perspectives; see, for example, [9, 12, 13, 25, 26, 27, 29, 35, 48], just to name a few of them. To the best of our knowledge, the existing convergence analyses of ADMM depend on the solvability of the dual problem or the existence of saddle points for the corresponding Lagrangian function, which might not be true for inverse problems (1.2).

In this paper we propose an ADMM algorithm in the framework of iterative regularization methods. By introducing an additional variable \( y = Wx \), we can reformulate (1.2) into the equivalent form

\[
\begin{align*}
\text{minimize} & \quad f(y) \\
\text{subject to} & \quad Ax = b, \quad Wx = y, \quad x \in \mathcal{D}(W).
\end{align*}
\]

The corresponding augmented Lagrangian function is

\[
L_{\rho_1, \rho_2}(x, y; \lambda, \mu) = f(y) + \langle \lambda, Ax - b \rangle + \langle \mu, Wx - y \rangle + \frac{\rho_1}{2} \|Ax - b\|^2 + \frac{\rho_2}{2} \|Wx - y\|^2,
\]

where \( \rho_1 \) and \( \rho_2 \) are two positive constants. Our ADMM algorithm then reads

\[
\begin{align*}
x_{k+1} &= \arg\min_{x \in \mathcal{D}(W)} L_{\rho_1, \rho_2}(x, y_k; \lambda_k, \mu_k), \\
y_{k+1} &= \arg\min_{y \in \mathcal{Y}} L_{\rho_1, \rho_2}(x_{k+1}, y; \lambda_k, \mu_k), \\
\lambda_{k+1} &= \lambda_k + \rho_1 (Ax_{k+1} - b), \\
\mu_{k+1} &= \mu_k + \rho_2 (Wx_{k+1} - y_{k+1}).
\end{align*}
\]

The \( x \) subproblem in (1.6) is a quadratical minimization problem, which can be solved by many methods, and the \( y \) subproblem can be solved explicitly when \( f \) is properly chosen, e.g., \( f \) are certain sparsity promoting functions, Thus, our ADMM algorithm
can be efficiently implemented. When the datum $b$ in (1.2) is consistent in the sense that $b = Ax$ for some $x \in \mathcal{D}(W)$ with $Wx \in \mathcal{D}(f)$, and when $f$ is strongly convex, we give a convergence analysis of our ADMM algorithm by using tools from convex analysis. The proof is based on some crucial monotonicity results on the residual and the Bregman distance. The whole proof, however, does not require the existence of a Lagrange multiplier to (1.4). When the data contain noise, similarly to other iterative regularization methods, our ADMM algorithm shows the semiconvergence property: i.e., the iterate becomes close to the sought solution at the beginning; however, after a critical number of iterations, the iterate leaves the sought solution far away as the iteration proceeds. By proposing a suitable stopping rule, we establish the regularization property of our algorithm. To the best of our knowledge, this is the first time that ADMM has been used to solve inverse problems directly as an iterative regularization method.

There are several different iterative regularization methods proposed for solving (1.2). The augmented Lagrangian method (ALM) is a popular and efficient algorithm which takes the form

\[
\begin{align*}
    x_{k+1} &= \arg \min_{x \in \mathcal{D}(W)} \left\{ f(Wx) + \langle \lambda_k, Ax - b \rangle + \frac{\rho_k}{2} \| Ax - b \|^2 \right\}, \\
    \lambda_{k+1} &= \lambda_k + \rho_k (Ax_{k+1} - b),
\end{align*}
\]

where $\{\rho_k\}$ is a sequence of positive numbers satisfying suitable properties. ALM was originally proposed by Hestenes [28] and Powell [41] independently; see [2, 42] for its convergence analysis for well-posed optimization problems. Recently, ALM has been applied to solve ill-posed inverse problems in [18, 19, 20, 34, 40]. It has been shown that ALM can produce a satisfactory approximate solution within a few iterations if $\{\rho_k\}$ is chosen to be geometrically increasing. However, since $f(Wx)$ and $A$ are coupled, solving the $x$ subproblem in (1.7) is highly nontrivial, and an additional inner solver should be incorporated into the algorithm. To remedy this drawback, a linearization step can be introduced to modify the $x$ subproblem in (1.7), which leads to the Uzawa-type iteration

\[
\begin{align*}
    \lambda_{k+1} &= \lambda_k + \rho_k (Ax_k - b), \\
    x_{k+1} &= \arg \min_{x \in \mathcal{D}(W)} \left\{ f(Wx) + \langle \lambda_{k+1}, Ax - b \rangle \right\}.
\end{align*}
\]

This method and its variants have been analyzed in [3, 31, 32] for ill-posed inverse problems. Unlike (1.7), the resolution of the $x$ subproblem in (1.8) only relies on $f(Wx)$ and hence is much easier to implement. For instance, if the sought solution is sparse, one may take $f(x) = \| x \|_\ell^1 + \frac{\mu}{2} \| x \|^2$ and the identity $W = I$, then the $x$ subproblem in (1.8) can be solved explicitly by the soft thresholding. However, if the sought solution is sparse under a transform $W$ that is not identity, which can occur when using the total variation [44], the wavelet frame [15, 46], and so on, the $x$ subproblem in (1.8) does not have a closed form solution and an inner solver is needed. In contrast to (1.7) and (1.8), our ADMM algorithm (1.6) can admit closed form solutions for each subproblem for many important applications.

The rest of this paper is organized as follows. In section 2, we give conditions to guarantee that (1.2) has a unique solution, show that our ADMM (1.6) is well defined, and establish an important monotonicity result. When the data are given exactly, we provide various convergence results of (1.6). When the data contain noise, we propose a stopping rule to terminate the iteration and show that our ADMM becomes a regularization method. In section 3 we report various numerical results to test the efficiency of our ADMM algorithm. Finally, we draw conclusions in section 4.
2. The method and its convergence analysis.

2.1. Preliminary. We consider the convex minimization problem (1.2) arising from linear inverse problems, where \( A : \mathcal{X} \to \mathcal{H} \) is a bounded linear operator between two Hilbert spaces \( \mathcal{X} \) and \( \mathcal{H} \), \( W \) is a linear operator from \( \mathcal{X} \) to another Hilbert space \( \mathcal{Y} \) with domain \( \mathcal{D}(W) \), and \( f : \mathcal{Y} \to (-\infty, \infty] \) is a convex function. The inner products and norms on \( \mathcal{X}, \mathcal{Y} \), and \( \mathcal{H} \) will be simply denoted by \( \langle \cdot, \cdot \rangle \) and \( \| \cdot \| \), respectively, which should be clear from the context. Throughout the paper we will make the following assumptions on the operators \( W, A \) and the function \( f \):

- **(A1)** \( A : \mathcal{X} \to \mathcal{H} \) is a bounded linear operator. We use \( A^* : \mathcal{H} \to \mathcal{X} \) to denote its adjoint.
- **(A2)** \( f : \mathcal{Y} \to (-\infty, \infty] \) is a proper, lower semicontinuous, strongly convex function in the sense that there is a constant \( c_0 > 0 \) such that
  \[
  f(ty_1 + (1-t)y_2) + c_0 t(1-t)\|y_1 - y_2\|^2 \leq tf(y_1) + (1-t)f(y_2)
  \]
  for all \( y_1, y_2 \in \mathcal{Y} \) and \( 0 \leq t \leq 1 \).
- **(A3)** \( W : \mathcal{X} \to \mathcal{Y} \) is a densely defined, closed, linear operator with domain \( \mathcal{D}(W) \).
- **(A4)** There is a constant \( c_1 > 0 \) such that
  \[
  \|Ax\|^2 + \|Wx\|^2 \geq c_1 \|x\|^2 \quad \forall x \in \mathcal{D}(W).
  \]

The assumptions \((A1)\) and \((A2)\) are standard. We will use \( \partial f(y) \) to denote the subdifferential of \( f \) at \( y \), i.e.,

\[
\partial f(y) = \{ \mu \in \mathcal{Y} : \langle \mu, \bar{y} - y \rangle \geq f(y) - f(\bar{y}) \text{ for all } \bar{y} \in \mathcal{Y} \}.
\]

Let \( \mathcal{D}(\partial f) = \{ y \in \mathcal{Y} : \partial f(y) \neq \emptyset \} \). Then for \( y \in \mathcal{D}(\partial f) \) and \( \mu \in \partial f(y) \) we can introduce

\[
D_\mu f(\bar{y}, y) = f(\bar{y}) - f(y) - \langle \mu, \bar{y} - y \rangle \quad \forall \bar{y} \in \mathcal{Y},
\]

which is called the Bregman distance induced by \( f \) at \( y \) in the direction \( \mu \); see [5]. When \( f \) is strongly convex in the sense of (2.1), by definition one can show that

\[
D_\mu f(\bar{y}, y) \geq c_0 \|\bar{y} - y\|^2
\]

for all \( \bar{y} \in \mathcal{Y}, y \in \mathcal{D}(\partial f) \), and \( \mu \in \partial f(y) \). Moreover

\[
\langle \mu - \bar{\mu}, y - \bar{y} \rangle \geq 2c_0 \|y - \bar{y}\|^2
\]

for all \( y, \bar{y} \in \mathcal{D}(\partial f), \mu \in \partial f(y) \), and \( \bar{\mu} \in \partial f(\bar{y}) \).

The assumptions \((A3)\) and \((A4)\) are standard conditions used in the literature on regularization methods with differential operators; see [17, 37, 38]. They will be used to show that our ADMM (1.6) is well defined. The closedness of \( W \) in \((A3)\) implies that \( W \) is also weakly closed. We can define the adjoint \( W^* \) of \( W \) which is also closed and densely defined. Moreover, \( z \in \mathcal{D}(W^*) \) if and only if \( \langle W^* z, x \rangle = \langle z, Wx \rangle \) for all \( x \in \mathcal{D}(W) \). A sufficient condition which guarantees \((A4)\) is that

\[
W \text{ has a closed range in } \mathcal{Y}, \quad \dim \ker W < +\infty, \quad \ker W \cap \ker A = \{0\};
\]

see [17, Chapter 8]. This sufficient condition is important in practice as it is satisfied by many interesting examples. For instance, consider the following three examples:
(i) \( W = I \) the identity operator with \( \mathcal{D}(W) = \mathcal{X} = \mathcal{Y} \).

(ii) \( W \) is a frame transform, i.e., there exist \( 0 < c \leq C < +\infty \) such that

\[
\|c\varphi\|_{2}^{2} \leq \|W\varphi\|_{2}^{2} \leq C\|\varphi\|_{2}^{2} \quad \forall \varphi \in \mathcal{X}
\]

with \( \mathcal{D}(W) = \mathcal{X} = L^{2}(\Omega) \) and \( \mathcal{Y} = l^{2}(\mathbb{N}) \).

(iii) The constant function 1 is not in the kernel of \( A,W = \nabla \), the gradient operator with \( \mathcal{X} = L^{2}(\Omega), \mathcal{D}(W) = H^{1}(\Omega), \) and \( \mathcal{Y} = [L^{2}(\Omega)]^{d} \).

For (i), the conditions (A3) and (2.4) hold trivially. For (ii), (A3) follows from [10, Proposition 12.7] and (2.4) follows from the coercivity of \( W \) (see, e.g., [10, p. 107]). For (iii), when \( \Omega \subset \mathbb{R}^{d} \) with \( d \leq 3 \) is an open bounded domain with Lipschitz boundary, (A3) follows from the definition of weak derivatives. Note that \( \text{Ker} W \) is a one dimension subspace. This fact together with the Helmholtz-Hodge decomposition (see [23, Theorem 3.4]) implies (2.4).

Under the assumptions (A1)–(A4), the following result shows that the minimization problem (1.2) admits a unique solution whenever \( b \) is consistent in the sense that \( b = Ax \) for some \( x \in \mathcal{D}(W) \) with \( Wx \in \mathcal{D}(f) \).

**Theorem 2.1.** Let \( b \) in (1.2) be consistent and let (A1)–(A4) hold. Then the optimization problem (1.2) admits a unique solution \( x* \in \mathcal{D}(W) \) with \( Wx* \in \mathcal{D}(f) \).

**Proof.** Let \( f_{\ast} := \inf \{ f(Wz) : Az = b, z \in \mathcal{D}(W) \} \). Since \( b \) is consistent, we have \( f_{\ast} < \infty \). Let \( \{ z_{k} \} \) be the minimizing sequence such that

\[
z_{k} \in \mathcal{D}(W), \quad Az_{k} = b, \quad \text{and} \quad \lim_{k \to \infty} f(Wz_{k}) = f_{\ast}.
\]

By (A2), \( f \) is strongly convex and hence is coercive; see, e.g., [1, Proposition 11.16]. Thus \( \{ Wz_{k} \} \) is bounded in \( \mathcal{Y} \). In view of (A4), \( \{ z_{k} \} \) is bounded in \( \mathcal{X} \). Therefore, \( \{ z_{k} \} \) has a subsequence, which is denoted by the same notation, such that

\[
z_{k} \to x* \text{ weakly in } \mathcal{X}, \quad Wz_{k} \to y* \text{ weakly in } \mathcal{Y}.
\]

By using (A3) and \( \{ z_{k} \} \subset \mathcal{D}(W) \), we have \( x* \in \mathcal{D}(W) \) and \( y* = Wx* \). Since \( f \) is convex and lower semicontinuous, \( f \) is also weakly lower semicontinuous (see [1, Proposition 10.23]). Thus \( f(Wx*) \leq \liminf_{k \to \infty} f(Wz_{k}) = f_{\ast} \) and hence \( x* \) is an optimal solution of (1.2). The uniqueness follows by the strong convexity of \( f \) and (A4). \( \square \)

2.2. ADMM algorithm and basic estimates. As described in the introduction, by introducing an additional variable \( y = Wx \), we can reformulate (1.2) into the equivalent form (1.4). Recalling the augmented Lagrangian function (1.5), our ADMM (1.6) starts from some initial guess \( y_{0} \in \mathcal{Y}, \lambda_{0} \in \mathcal{H}, \mu_{0} \in \mathcal{Y} \) and defines

\[
(5.1) \quad x_{k+1} = \arg \min_{x \in \mathcal{D}(W)}\left\{ \langle \lambda_{k}, Ax \rangle + \langle \mu_{k}, Wx \rangle + \frac{\rho_{1}}{2}\|Ax - b\|^{2} + \frac{\rho_{2}}{2}\|Wx - y_{k}\|^{2} \right\},
\]

\[
(5.2) \quad y_{k+1} = \arg \min_{y \in \mathcal{Y}}\left\{ f(y) - \langle \mu_{k}, y \rangle + \frac{\rho_{2}}{2}\|Wx_{k+1} - y\|^{2} \right\},
\]

\[
(5.3) \quad \lambda_{k+1} = \lambda_{k} + \rho_{1}(Ax_{k+1} - b),
\]

\[
(5.4) \quad \mu_{k+1} = \mu_{k} + \rho_{2}(Wx_{k+1} - y_{k+1})
\]

for \( k = 0,1,\ldots \), where \( \rho_{1} \) and \( \rho_{2} \) are two fixed positive constants.
We need to show that \( x_{k+1} \) and \( y_{k+1} \) are well defined. Note that (2.5) and (2.6) can be written as

\[
x_{k+1} = \arg \min_{x \in \mathcal{D}(W)} \left\{ \frac{\rho_1}{2} \|Ax - b + \lambda_k + \rho_1 x\|_2^2 + \frac{\rho_2}{2} \|Wx - y_k - \mu_k\|_2^2 \right\},
\]

\[
y_{k+1} = \arg \min_{y \in \mathcal{Y}} \left\{ f(y) + \frac{\rho_2}{2} \|y - Wx_{k+1} - \mu_k\|_2^2 \right\}.
\]

Therefore, the well-posedness of \( x_{k+1} \) and \( y_{k+1} \) follows from the following result.

**Lemma 2.2.** Let assumptions (A1)–(A4) hold.

(i) For any \( h \in \mathcal{H} \) and \( v \in \mathcal{Y} \), the minimization problem

\[
(2.9) \quad \min_{z \in \mathcal{D}(W)} \left\{ \frac{\rho_1}{2} \|Az - h\|_2^2 + \frac{\rho_2}{2} \|Wz - v\|_2^2 \right\}
\]

admits a unique solution \( z \). Moreover, \( z \) and \( Wz \) depend continuously on \( h \) and \( v \).

(ii) For any \( v \in \mathcal{Y} \) the minimization problem

\[
(2.10) \quad \min_{y \in \mathcal{Y}} \left\{ f(y) + \frac{\rho_2}{2} \|y - v\|_2^2 \right\}
\]

admits a unique solution \( y \). Moreover, \( y \) and \( f(y) \) depend continuously on \( v \).

**Proof.** (i) follows from [39, p. 23 Theorem 4 and p. 26 Theorem 6] and (ii) follows from [33, Lemma 2.2]. \( \square \)

We now take a closer look at the ADMM algorithm (2.5)–(2.8). From (2.5) it follows that \( x_{k+1} \in \mathcal{D}(W) \) satisfies the optimality condition

\[
(A^*\lambda_k + \rho_1 A^*(Ax_{k+1} - b), x) + (\mu_k + \rho_2(Wx_{k+1} - y_k), Wx) = 0, \quad x \in \mathcal{D}(W).
\]

This implies that \( \mu_k + \rho_2(Wx_{k+1} - y_k) \in \mathcal{D}(W^*) \) and

\[
(2.11) \quad A^*\lambda_k + \rho_1 A^*(Ax_{k+1} - b) + W^*[\mu_k + \rho_2(Wx_{k+1} - y_k)] = 0.
\]

From (2.6) we can obtain that

\[
(2.12) \quad 0 \in \partial f(y_{k+1}) - \mu_k - \rho_2(Wx_{k+1} - y_{k+1}).
\]

For simplicity of exposition, we introduce the residuals

\[
r_k = Ax_k - b \quad \text{and} \quad s_k = Wx_k - y_k, \quad k = 1, 2, \ldots.
\]

It then follows from (2.7), (2.8), (2.11), and (2.12) that

\[
(2.13) \quad \lambda_{k+1} - \lambda_k = \rho_1 r_{k+1},
\]

\[
(2.14) \quad \mu_{k+1} - \mu_k = \rho_2 s_{k+1},
\]

\[
(2.15) \quad \mu_{k+1} \in \partial f(y_{k+1}),
\]

\[
(2.16) \quad A^*\lambda_k + \rho_1 A^*r_{k+1} = -W^*[\mu_k + \rho_2(Wx_{k+1} - y_k)]
\]

for \( k = 0, 1, \ldots \).

**Lemma 2.3.** There holds \( A^*\lambda_1 = W^*[\rho_2(y_0 - y_1) - \mu_1] \). Moreover, for \( k \geq 1 \) there holds \( \rho_1 A^*r_{k+1} = \rho_2 W^*[y_k - y_{k+1}] - (y_{k-1} - y_k) - s_{k+1} \), that is

\[
\rho_1(r_{k+1}, Ax) = \rho_2((y_k - y_{k+1}) - (y_{k-1} - y_k) - s_{k+1}, Wx)
\]

for all \( x \in \mathcal{D}(W) \).
In view of (2.19) and the Cauchy–Schwarz inequality, we have

This shows the desired inequality. 

By virtue of the identity

we have

which in particular implies

Further, by using (2.13), (2.17) and (2.14) we have

for \( k \geq 1 \), which completes the proof. 

The following monotonicity result plays an essential role in the forthcoming convergence analysis.

**Lemma 2.4.** Let \( E_k = \rho_1 \|r_k\|^2 + \rho_2 \|s_k\|^2 + \rho_2 \|y_k - y_{k-1}\|^2 \). Then

\[
E_{k+1} - E_k \leq -\rho_1 \|r_{k+1} - r_k\|^2 - 4\rho_0 \|y_{k+1} - y_k\|^2
\]

for all \( k \geq 1 \). In particular, \( E_k \) is monotonically decreasing along the iteration and \( \sum_{k=1}^{\infty} \|y_{k+1} - y_k\|^2 < \infty \).

**Proof.** By the definition of \( r_k \) and \( s_k \) we have

\[
\begin{align*}
(2.18) & \quad r_{k+1} - r_k = A(x_{k+1} - x_k), \\
(2.19) & \quad s_{k+1} - s_k = W(x_{k+1} - x_k) + (y_k - y_{k+1}).
\end{align*}
\]

Therefore

\[
\begin{align*}
\rho_1 \langle r_{k+1} - r_k, r_{k+1} \rangle + \rho_2 \langle s_{k+1} - s_k, s_{k+1} \rangle \\
= \rho_1 \langle A(x_{k+1} - x_k), r_{k+1} \rangle + \rho_2 \langle W(x_{k+1} - x_k) + (y_k - y_{k+1}), s_{k+1} \rangle.
\end{align*}
\]

Recall that \( x_{k+1} - x_k \in \mathcal{P}(W) \). We may use Lemma 2.3, (2.14), (2.15), and (2.3) to derive that

\[
\begin{align*}
& \rho_1 \langle r_{k+1} - r_k, r_{k+1} \rangle + \rho_2 \langle s_{k+1} - s_k, s_{k+1} \rangle \\
& = \rho_2 \langle W(x_{k+1} - x_k), (y_k - y_{k+1}) - (y_{k-1} - y_k) \rangle - \langle y_{k+1} - y_k, \mu_{k+1} - \mu_k \rangle \\
& \leq \rho_2 \langle W(x_{k+1} - x_k), (y_k - y_{k+1}) - (y_{k-1} - y_k) \rangle - 2\rho_0 \|y_{k+1} - y_k\|^2.
\end{align*}
\]

In view of (2.19) and the Cauchy–Schwarz inequality, we have

\[
\begin{align*}
& \rho_1 \langle r_{k+1} - r_k, r_{k+1} \rangle + \rho_2 \langle s_{k+1} - s_k, s_{k+1} \rangle \\
& \leq \rho_2 \langle s_{k+1} - s_k \rangle + (y_{k+1} - y_k), (y_k - y_{k+1}) - (y_{k-1} - y_k) \rangle - 2\rho_0 \|y_{k+1} - y_k\|^2, \\
& \leq \rho_2 \frac{1}{2} \langle s_{k+1} - s_k \rangle^2 - \|y_{k+1} - y_k\|^2 + \|y_{k-1} - y_k\|^2 \rangle - 2\rho_0 \|y_{k+1} - y_k\|^2.
\end{align*}
\]

By virtue of the identity

we therefore obtain

\[
\begin{align*}
& \rho_1 \|r_{k+1}\|^2 + \rho_2 \|s_{k+1}\|^2 \leq \rho_1 \|r_k\|^2 + \rho_2 \|s_k\|^2 - \rho_2 \|y_{k+1} - y_k\|^2 \\
& \quad + \rho_2 \|y_{k-1} - y_k\|^2 - 4\rho_0 \|y_{k+1} - y_k\|^2.
\end{align*}
\]

This shows the desired inequality. 

\[ \square \]
2.3. Exact data case. In this subsection we will give the convergence analysis of the ADMM algorithm (2.5)–(2.8) under the condition that the datum \( b \) is consistent so that (1.2) has a unique solution. We will always use \((\hat{x}, \hat{y})\) to represent any feasible point of (1.4), i.e., \( \hat{x} \in \mathcal{D}(W) \) and \( \hat{y} \in \mathcal{D}(f) \) such that \( Ax = b \) and \( W\hat{x} = \hat{y} \).

**Lemma 2.5.** The sequences \( \{x_k\} \) and \( \{y_k\} \) are bounded and

\[
\sum_{k=1}^{\infty} \{D_{\mu_k} f(y_{k+1}, y_k) + E_k\} < \infty.
\]

In particular, \( Ax_k \to b \), \( Wx_k - y_k \to 0 \), and \( y_{k+1} - y_k \to 0 \) as \( k \to \infty \).

**Proof.** Let \((\hat{x}, \hat{y})\) be any feasible point of (1.4). By using (2.14) and Lemma 2.3 we have

\[
D_{\mu_{k+1}} f(\hat{y}, y_{k+1}) - D_{\mu_k} f(\hat{y}, y_k) + D_{\mu_k} f(y_{k+1}, y_k) = \langle \mu_k - \mu_{k+1}, \hat{y} - y_{k+1} \rangle = -\rho_2 \langle s_{k+1}, W(\hat{x} - x_{k+1}) + s_{k+1} \rangle
\]

\[
= -\rho_2 \|s_{k+1}\|^2 + \rho_2 \langle (y_{k-1} - y_k) - (y_k - y_{k+1}), W(\hat{x} - x_{k+1}) \rangle
\]

\[
+ \rho_1 \langle r_{k+1}, A(\hat{x} - x_{k+1}) \rangle
\]

\[
= -\rho_1 \|r_{k+1}\|^2 - \rho_2 \|s_{k+1}\|^2 + \rho_2 \langle y_{k-1} - y_k, W(\hat{x} - x_{k+1}) \rangle
\]

\[
- \rho_2 \langle y_k - y_{k+1}, W(\hat{x} - x_{k+1}) \rangle.
\]

(2.20)

For any positive integers \( m < n \), by summing the above inequality over \( k \) from \( k = m \) to \( k = n - 1 \) we can obtain

\[
D_{\mu_n} f(\hat{y}, y_n) - D_{\mu_m} f(\hat{y}, y_m) + \sum_{k=m}^{n-1} D_{\mu_k} f(y_{k+1}, y_k) = \sum_{k=m+1}^{\infty} (\rho_1 \|r_k\|^2 + \rho_2 \|s_k\|^2) + \rho_2 \langle y_{m-1} - y_m, W(\hat{x} - x_{m+1}) \rangle + \rho_2 \sum_{k=m}^{n-2} (\langle y_k - y_{k+1}, W(x_{k+1} - x_{k+2}) \rangle - \langle y_{n-1} - y_n, W(\hat{x} - x_n) \rangle).
\]

(2.21)

By taking \( m = 1 \) in the above equation, it follows that

\[
D_{\mu_n} f(\hat{y}, y_n) + \sum_{k=1}^{n-1} D_{\mu_k} f(y_{k+1}, y_k) = D_{\mu_1} f(\hat{y}, y_1) - \sum_{k=2}^{n} (\rho_1 \|r_k\|^2 + \rho_2 \|s_k\|^2) + \rho_2 \langle y_0 - y_1, W(\hat{x} - x_2) \rangle
\]

\[
+ \rho_2 \sum_{k=1}^{n-2} (\langle y_k - y_{k+1}, W(x_{k+1} - x_{k+2}) \rangle - \langle y_{n-1} - y_n, W(\hat{x} - x_n) \rangle).
\]
We need to estimate the last two terms. By the Cauchy–Schwarz inequality, we have
\[
\sum_{k=1}^{n-2} (y_k - y_{k+1}, W(x_{k+1} - x_{k+2})) = \sum_{k=1}^{n-2} (y_k - y_{k+1}, s_{k+1} + (y_{k+1} - y_{k+2}) - s_{k+2})
\]
\[
\leq \sum_{k=1}^{n-2} \left( \frac{1}{8} ||s_{k+1}||^2 + \frac{1}{8} ||s_{k+2}||^2 + \frac{9}{2} ||y_k - y_{k+1}||^2 + \frac{1}{2} ||y_{k+1} - y_{k+2}||^2 \right)
\]
(2.22) \[
\leq \frac{1}{4} \sum_{k=2}^{n} ||s_k||^2 + 5 \sum_{k=1}^{n-1} ||y_k - y_{k+1}||^2.
\]

Similarly we have
\[
-\langle y_{n-1} - y_n, W(\hat{x} - x_n) \rangle
\]
\[
= -\langle y_{n-1} - y_n, W(\hat{x} - x_1) \rangle - \sum_{k=1}^{n-1} (y_{n-1} - y_n, s_k + (y_k - y_{k+1}) - s_{k+1})
\]
\[
\leq \frac{1}{4} ||W(\hat{x} - x_1)||^2 - \langle y_{n-1} - y_n, s_1 - s_n \rangle - \sum_{k=1}^{n-2} \langle y_{n-1} - y_n, y_k - y_{k+1} \rangle
\]
\[
\leq \frac{1}{4} ||W(\hat{x} - x_1)||^2 + \frac{1}{4} (||s_1||^2 + ||s_n||^2) + \frac{n}{4} ||y_{n-1} - y_n||^2 + 2 \sum_{k=1}^{n-2} ||y_k - y_{k+1}||^2.
\]
(2.23)

Therefore, we can conclude that there is a constant \( C \) independent of \( n \) such that
\[
D_{\mu_n} f(\hat{y}, y_n) + \sum_{k=1}^{n-1} D_{\mu_k} f(y_{k+1}, y_k)
\]
\[
\leq C - \rho_1 \sum_{k=2}^{n} ||r_k||^2 - \rho_2 \sum_{k=2}^{n} ||s_k||^2 + \frac{1}{4} \rho_2 n ||y_{n-1} - y_n||^2
\]
\[
+ 7 \rho_2 \sum_{k=1}^{n-1} ||y_k - y_{k+1}||^2.
\]
(2.24)

From Lemma 2.4 it follows that \( \sum_{n=1}^{\infty} ||y_n - y_{n+1}||^2 < \infty \). Thus, we can find a subsequence of integers \( \{n_j\} \) with \( n_j \to \infty \) such that \( n_j ||y_{n_j} - y_{n_j+1}||^2 \to 0 \) as \( j \to \infty \). Consequently, it follows from (2.24) that
\[
\sum_{k=1}^{n_{j-1}} D_{\mu_k} f(y_{k+1}, y_k) + \rho_1 \sum_{k=2}^{n_j} ||r_k||^2 + \frac{\rho_2}{2} \sum_{k=2}^{n_j} ||s_k||^2 \leq C.
\]

Letting \( j \to \infty \) gives
\[
\sum_{k=1}^{\infty} (D_{\mu_k} f(y_{k+1}, y_k) + \rho_1 ||r_k||^2 + \rho_2 ||s_k||^2) < \infty.
\]

We therefore obtain \( \sum_{k=1}^{\infty} E_k < \infty \). By Lemma 2.4, \( \{ E_k \} \) is monotonically decreasing. Thus \( n E_n \leq \sum_{k=1}^{n} E_k \leq C \), and \( n \rho_2 ||y_n - y_{n+1}||^2 \leq n E_n \leq C \). Consequently, from
(2.24) it follows that $D_{\mu_n} f(\hat{y}, y_n) \leq C$. By the strong convexity of $f$, we can conclude that $\{y_n\}$ is bounded. Furthermore, using $\sum_{n=1}^{\infty} E_n < \infty$, we can conclude that $Ax_n \to b$, $Wx_n - y_n \to 0$, and $y_n - y_{n+1} \to 0$ as $n \to \infty$. In view of the boundedness of $\{Ax_n\}$ and $\{Wx_n\}$, we can use (A4) to conclude that $\{x_n\}$ is bounded.

**Lemma 2.6.** Let $(\hat{x}, \hat{y})$ be any feasible point of (1.4). Then $\{D_{\mu_n} f(\hat{y}, y_k)\}$ is a convergent sequence.

**Proof.** Let $m < n$ be any two positive integers. By using (2.21) we have

\[
|D_{\mu_n} f(\hat{y}, y_n) - D_{\mu_m} f(\hat{y}, y_m)| \leq \sum_{k=m}^{n-1} D_{\mu_k} f(y_{k+1}, y_k) + \sum_{k=m+1}^{n} (\rho_1\|r_k\|^2 + \rho_2\|s_k\|^2) + \rho_2 \langle y_m - y_n, W(\hat{x} - x_{m+1}) \rangle + \rho_2 \sum_{k=m}^{n-2} \|y_k - y_{k+1}, W(x_{k+1} - x_{k+2}) \rangle + \rho_2 \|y_n - y_n, W(\hat{x} - x_n) \rangle.
\]

By the same argument for deriving (2.22), we have

\[
\left| \sum_{k=m}^{n-2} \langle y_k - y_{k+1}, W(x_{k+1} - x_{k+2}) \rangle \right| \leq \frac{1}{4} \sum_{k=m+1}^{n} \|s_k\|^2 + 5 \sum_{k=m}^{n-1} \|y_k - y_{k+1}\|^2.
\]

Therefore

\[
|D_{\mu_n} f(\hat{y}, y_n) - D_{\mu_m} f(\hat{y}, y_m)| \leq \sum_{k=m}^{\infty} \left( D_{\mu_k} f(y_{k+1}, y_k) + \rho_1\|r_k\|^2 + \frac{5}{4}\rho_2\|s_k\|^2 + 5\rho_2\|y_k - y_{k+1}\|^2 \right) + \rho_2 \|y_m - y_n\||W(\hat{x} - x_{m+1})\| + \rho_2 \|y_n - y_n\||W(\hat{x} - x_n)\|.
\]

In view of Lemma 2.5, we can conclude that

\[
|D_{\mu_n} f(\hat{y}, y_n) - D_{\mu_m} f(\hat{y}, y_m)| \to 0 \quad \text{as } m, n \to \infty.
\]

This shows that $\{D_{\mu_k} f(\hat{y}, y_k)\}$ is a Cauchy sequence and hence is convergent.

Now we are ready to give the main convergence result concerning the ADMM algorithm (2.5)–(2.8) with exact data.

**Theorem 2.7.** Let (A1)–(A4) hold and let $b$ be consistent. Let $x^*$ be the unique solution of (1.2) and let $y^* = Wx^*$. Then for the ADMM (2.5)–(2.8) there hold

\[
x_k \to x^*, \quad y_k \to y^*, \quad Wx_k \to y^*, \quad f(y_k) \to f(y^*), \quad \text{and} \quad D_{\mu_k} f(y^*, y_k) \to 0
\]

as $k \to \infty$.

**Proof.** We first show that $\{y_k\}$ is a Cauchy sequence. Let $(\hat{x}, \hat{y})$ be any feasible point of (1.4). We will use the identity

\[
(2.25) \quad D_{\mu_n} f(y_n, y_m) = D_{\mu_n} f(\hat{y}, y_m) - D_{\mu_n} f(\hat{y}, y_n) + \langle \mu_n - \mu_m, y_n - \hat{y} \rangle.
\]
In view of (2.14) and lemma (2.3), we can write
\[
\langle \mu_n - \mu_m, y_n - \hat{y} \rangle = \sum_{k=m}^{n-1} \langle \mu_{k+1} - \mu_k, y_n - \hat{y} \rangle = \rho_2 \sum_{k=m}^{n-1} \langle s_{k+1}, y_n - \hat{y} \rangle
\]
\[
= -\rho_2 \sum_{k=m}^{n-1} \langle s_{k+1}, s_n \rangle + \rho_2 \sum_{k=m}^{n-1} \langle s_{k+1}, W(x_n - \hat{x}) \rangle
\]
\[
= -\sum_{k=m}^{n-1} (\rho_2 s_{k+1, s_n} + \rho_1 r_{k+1, r_n}) + \rho_2 (y_{n-1} - y_n, W(x_n - \hat{x}))
\]
\[-\rho_2 \langle y_{n-1} - y_n, W(x_n - \hat{x}) \rangle.
\]
By the Cauchy–Schwarz inequality and the monotonicity of \( \{E_k\} \), we have
\[
|\langle \mu_n - \mu_m, y_n - \hat{y} \rangle| \leq \frac{1}{2} \sum_{k=m+1}^{n} (\rho_1 \|r_k\|^2 + \rho_2 \|s_k\|^2) + \frac{n-m}{2} (\rho_1 \|r_n\|^2 + \rho_2 \|s_n\|^2)
\]
\[+ \rho_2 \langle y_{n-1} - y_n, W(x_n - \hat{x}) \rangle + \rho_2 \langle y_{n-1} - y_n, W(x_n - \hat{x}) \rangle
\leq \sum_{k=m+1}^{n} E_k + \rho_2 \langle y_{n-1} - y_n, W(x_n - \hat{x}) \rangle
+ \rho_2 \langle y_{n-1} - y_n, W(x_n - \hat{x}) \rangle.
\]
(2.26)

This together with Lemma 2.5 implies that
\[
(2.27) \quad \langle \mu_n - \mu_m, y_n - \hat{y} \rangle \to 0 \quad \text{as } m, n \to \infty.
\]
Combining this with (2.25) and using Lemma 2.6, we obtain that \( D_{\mu_n} f(y_n, y_m) \to 0 \) as \( m, n \to \infty \). By the strong convexity of \( f \) we have \( \|y_n - y_m\| \to 0 \) as \( m, n \to \infty \). Thus \( \{y_k\} \) is a Cauchy sequence in \( Y \). Consequently, there is \( \hat{y} \in Y \) such that \( y_k \to \hat{y} \) as \( k \to \infty \).

We will show that there is \( \hat{x} \in \mathcal{D}(W) \) such that \( x_k \to \hat{x}, A\hat{x} = b, \) and \( W\hat{x} = \hat{y} \). By virtue of Lemma 2.5 and \( y_k \to \hat{y} \), we have \( Wx_k \to \hat{y} \) and \( Ax_k \to b \). From (A4) it follows that
\[
c_1 \|x_n - x_k\|^2 \leq \|Ax_n - Ax_k\|^2 + \|Wx_n - Wx_k\|^2
\]
for any integers \( n, k \). Therefore \( \|x_n - x_k\| \to 0 \) as \( n, k \to \infty \) which shows that \( \{x_k\} \) is a Cauchy sequence in \( X \). Thus, there is \( \hat{x} \in X \) such that \( x_k \to \hat{x} \) as \( k \to \infty \). Clearly \( b = \lim_{k \to \infty} Ax_k = A\hat{x} \). By using the closedness of \( W \) and \( \{x_k\} \subset \mathcal{D}(W) \), we can further conclude that \( \hat{x} \in \mathcal{D}(W) \) and \( W\hat{x} = \hat{y} \).

Next we will show that
\[
\hat{y} \in \partial f, \quad \lim_{k \to \infty} f(y_k) = f(\hat{y}), \quad \text{and} \quad \lim_{k \to \infty} D_{\mu_k} f(\hat{y}, y_k) = 0.
\]
Recalling that \( \mu_k \in \partial f(y_k) \), we have
\[
(2.28) \quad f(y_k) \leq f(\hat{y}) + \langle \mu_k, y_k - \hat{y} \rangle.
\]
By using (2.26), it yields
\[
\langle \mu_k, y_k - \hat{y} \rangle \leq \langle \mu_1, y_k - \hat{y} \rangle + \sum_{i=2}^{k} E_i + \rho_2 \|y_{k-1} - y_k\||W(x_k - \hat{x})|
+ \rho_2 \|y_0 - y_1\||W(x_1 - \hat{x})|,
\]
which together with (2.28) and Lemma 2.5 implies that \( f(y_k) \leq C < \infty \) for some constant \( C \) independent of \( k \). By the lower semicontinuity of \( f \) we have

\[
(2.29) \quad f(\tilde{y}) \leq \liminf_{k \to \infty} f(y_k) \leq C < \infty.
\]

Thus \( \tilde{y} \in \mathcal{D}(f) \). Since the above argument shows that \((\tilde{x}, \tilde{y})\) is a feasible point of (1.4), we may replace \((\tilde{x}, \tilde{y})\) in (2.26) by \((\tilde{x}, \tilde{y})\) and use \( y_k \to \tilde{y} \) and \( Wx_k \to W\tilde{x} \) to obtain

\[
\limsup_{k \to \infty} |\langle \mu_k, y_k - \tilde{y} \rangle| \leq \sum_{i=m+1} E_i
\]

for all integers \( m \). This together with Lemma 2.5 implies that \( \langle \mu_k, y_k - \tilde{y} \rangle \to 0 \) as \( k \to \infty \). Now we can use (2.28) with \( \tilde{y} \) replaced by \( \tilde{y} \) to obtain \( \limsup_{k \to \infty} f(y_k) \leq f(\tilde{y}) \). This together with (2.29) gives \( \lim_{k \to \infty} f(y_k) = f(\tilde{y}) \). It is now straightforward to show that \( \lim_{k \to \infty} D_{\mu_k} f(\tilde{y}, y_k) = 0 \).

Finally we show that \( \tilde{x} = x^* \) and \( \tilde{y} = y^* \). To see this, we first prove that \( f(\tilde{y}) \leq f(\tilde{y}) \) for any feasible point \((\tilde{x}, \tilde{y})\) of (1.4). We will use (2.28). Let \( \varepsilon > 0 \) be any small number. By using (2.27) and Lemma 2.5, we can find \( k_0 \) such that

\[
(2.30) \quad |\langle \mu_k - \mu_{k_0}, y_k - \tilde{y} \rangle| \leq \varepsilon \quad \text{and} \quad \rho_2 |\langle y_{k_0-1} - y_{k_0}, W(x_k - \tilde{x}) \rangle| \leq \varepsilon
\]

for all \( k \geq k_0 \). Thus, it follows from (2.28) that

\[
(2.31) \quad f(y_k) \leq f(\tilde{y}) + \varepsilon + |\langle \mu_{k_0}, y_k - \tilde{y} \rangle|.
\]

In view of (2.14) and Lemma 2.3 we have

\[
\langle \mu_{k_0}, y_k - \tilde{y} \rangle = -\langle \mu_{k_0}, s_k \rangle + \langle \mu_{k_0}, W(x_k - \tilde{x}) \rangle
\]

\[
= -\langle \mu_{k_0}, s_k \rangle + \langle \mu_1, W(x_k - \tilde{x}) \rangle + \rho_2 \sum_{i=2}^{k_0} \langle s_i, W(x_k - \tilde{x}) \rangle
\]

\[
= -\langle \mu_{k_0}, s_k \rangle - \langle \lambda_1, r_k \rangle + \rho_2 \langle y_0 - y_1, W(x_k - \tilde{x}) \rangle
\]

\[
- \sum_{i=2}^{k_0} \langle \rho_1 (r_i, r_k) - \rho_2 ((y_{i-1} - y_i) - (y_{i-2} - y_{i-1}), W(x_k - \tilde{x})) \rangle
\]

\[
= -\langle \mu_{k_0}, s_k \rangle - \langle \lambda_1, r_k \rangle - \rho_1 \sum_{i=2}^{k_0} \langle r_i, r_k \rangle + \rho_2 \langle y_{k_0-1} - y_{k_0}, W(x_k - \tilde{x}) \rangle.
\]

Thus, by using the second equation in (2.30) we can derive that

\[
(2.32) \quad |\langle \mu_{k_0}, y_k - \tilde{y} \rangle| \leq \|\mu_{k_0}\| \|s_k\| + \|\lambda_1\| \|r_k\| + \rho_1 \left( \sum_{i=2}^{k_0} \| r_i \| \right) \| r_k \| + \varepsilon.
\]

Combining (2.31) and (2.32), we obtain

\[
f(y_k) \leq f(\tilde{y}) + 2\varepsilon + \|\mu_{k_0}\| \|s_k\| + \|\lambda_1\| \|r_k\| + \rho_1 \left( \sum_{i=2}^{k_0} \| r_i \| \right) \| r_k \|.
\]

In view of Lemma 2.5, this implies that \( \limsup_{k \to \infty} f(y_k) \leq f(\tilde{y}) + 2\varepsilon \). By the lower semicontinuity of \( f \) and the fact \( y_k \to \tilde{y} \) we can derive that

\[
f(\tilde{y}) \leq \liminf_{k \to \infty} f(y_k) \leq f(\tilde{y}) + 2\varepsilon.
\]
Because \( \varepsilon > 0 \) can be arbitrarily small, we must have \( f(\tilde{y}) \leq f(\hat{y}) \) for any feasible point \((\tilde{x}, \tilde{y})\) of (1.4). Since \((\tilde{x}, \tilde{y})\) is a feasible point of (1.4), it follows that

\[
 f(W\tilde{x}) = f(\tilde{y}) = f(y^*) = f(Wx^*) = \min \{ f(Wx) : x \in \mathcal{D}(W) \text{ and } Ax = b \}.
\]

From the uniqueness of \( x^* \), see Theorem 2.1, we can conclude that \( \tilde{x} = x^* \) and hence \( \hat{y} = y^* \). The proof is therefore complete.

2.4. Noisy data case. In practical applications, the data are usually obtained by measurement and unavoidably contain error. Thus, instead of \( b \), usually we only have noisy data \( b^{\delta} \) satisfying

\[
\|b^{\delta} - b\| \leq \delta
\]

for a small noise level \( \delta > 0 \). In this situation, we need to replace \( b \) in our ADMM algorithm (2.5)–(2.8) by the noisy data \( b^{\delta} \) for numerical computation. For inverse problems, an iterative method using noisy data usually exhibits the semiconvergence property; i.e., the iterate converges toward the sought solution at the beginning, and, after a critical number of iterations, the iterate eventually diverges from the sought solution due to the amplification of noise. The iteration should be terminated properly in order to produce a reasonable approximate solution for (1.2). Incorporating a stopping criterion into the iteration leads us to propose Algorithm 1 for solving inverse problems with noisy data.

Algorithm 1 (ADMM with Noisy Data).

1. Input: initial guess \( y_0 \in \mathcal{Y}, \lambda_0 \in \mathcal{H}, \) and \( \mu_0 \in \mathcal{Y} \), Constants \( \rho_1 > 0, \rho_2 > 0, \) and \( \tau > 1 \), noise level \( \delta > 0 \).
2. Let \( y_0^{\delta} = y_0, \lambda_0^{\delta} = \lambda_0, \) and \( \mu_0^{\delta} = \mu_0 \).
3. for \( k = 0, 1, \ldots \) do
4. update \( x, y \) and the Lagrange multipliers \( \lambda, \mu \) as follows:

\[
 x_{k+1}^{\delta} = \arg \min_{x \in \mathcal{D}(W)} \left\{ \langle \lambda_k^{\delta}, Ax \rangle + \langle \mu_k^{\delta}, Wx \rangle + \frac{\rho_1}{2}\|Ax - b^{\delta}\|^2 + \frac{\rho_2}{2}\|Wx - y_k^{\delta}\|^2 \right\},
\]

\[
 y_{k+1}^{\delta} = \arg \min_{y \in \mathcal{Y}} \left\{ f(y) - \langle \mu_k^{\delta}, y \rangle + \frac{\rho_2}{2}\|Wx_{k+1}^{\delta} - y\|^2 \right\},
\]

\[
 \lambda_{k+1}^{\delta} = \lambda_k^{\delta} + \rho_1(Ax_{k+1}^{\delta} - b^{\delta}),
\]

\[
 \mu_{k+1}^{\delta} = \mu_k^{\delta} + \rho_2(Wx_{k+1}^{\delta} - y_{k+1}^{\delta}).
\]

5. check the stopping criterion:

\[
(2.33) \quad \rho_1^2\|Ax_k^{\delta} - b^{\delta}\|^2 + \rho_2^2\|Wx_k^{\delta} - y_k^{\delta}\|^2 \leq \max(\rho_1^2, \rho_2^2)\tau^2\delta^2.
\]

6. end for

Under (A1)–(A4), we may use Lemma 2.2 to conclude that \( x_k^{\delta}, y_k^{\delta}, \lambda_k^{\delta}, \) and \( \mu_k^{\delta} \) in Algorithm 1 are well defined for \( k \geq 1 \). Furthermore, we have the following stability result in which we take \( x_0^{\delta} = x_0 \) to be any element in \( \mathcal{D}(W) \).

Lemma 2.8. Consider Algorithm 1 without (2.33). Then for each fixed \( k \geq 0 \) there hold

\[
 x_k^{\delta} \to x_k, \quad y_k^{\delta} \to y_k, \quad Wx_k^{\delta} \to Wx_k, \quad \lambda_k^{\delta} \to \lambda_k, \quad \mu_k^{\delta} \to \mu_k, \quad f(y_k^{\delta}) \to f(y_k)
\]
as \( \delta \to 0 \), where \((x_k, y_k, \lambda_k, \mu_k)\) are defined by the ADMM algorithm (2.5)–(2.8) with exact data.

**Proof.** We use an induction argument. The result is trivial when \( k = 0 \). Assuming that the result is true for some \( k = n \), we show that it is also true for \( k = n + 1 \). From Lemma 2.2 and the induction hypothesis we can obtain that

\[
x_{n+1}^{\delta} \to x_{n+1}, \quad y_{n+1}^{\delta} \to y_{n+1}, \quad W x_{n+1}^{\delta} \to W x_{n+1}, \quad \text{and} \quad f(y_{n+1}^{\delta}) \to f(y_{n+1})
\]

as \( \delta \to 0 \). Now we can obtain \( \lambda_{n+1}^{\delta} \to \lambda_{n+1} \) and \( \mu_{n+1}^{\delta} \to \mu_{n+1} \) as \( \delta \to 0 \) from their definition.

In the following we will show that Algorithm 1 terminates after a finite number of iterations and defines a regularization method. For simplicity of exposition, we use the notation

\[
r_k^{\delta} = A x_k^{\delta} - b \quad \text{and} \quad s_k^{\delta} = W x_k^{\delta} - y_k^{\delta}.
\]

From the description of Algorithm 1, one can easily see that

\[
\begin{align*}
\lambda_{k+1}^{\delta} - \lambda_k^{\delta} &= \rho_1 s_k^{\delta}, \\
\mu_{k+1}^{\delta} - \mu_k^{\delta} &= \rho_2 s_k^{\delta}, \\
A^* \lambda_k^{\delta} + \rho_1 A^* s_k^{\delta} &= -W^* \left[ s_k^{\delta} + \rho_2 W x_k^{\delta} - y_k^{\delta} \right]
\end{align*}
\]

for \( k \geq 0 \). By the same argument in the proof of Lemma 2.3 one can derive that

\[
(2.34) \quad \rho_1 \langle r_k^{\delta}, Ax \rangle = \rho_2 \left( \langle y_k^{\delta} - y_{k+1}^{\delta} \rangle - \langle y_{k-1}^{\delta} - y_k^{\delta} \rangle \right) - s_k^{\delta}, \quad x \in \mathcal{D}(W)
\]

for \( k \geq 1 \). Furthermore, by using the same argument in the proof of Lemma 2.4 we can show the following result.

**Lemma 2.9.** Let \( E_k^{\delta} = \rho_1 \|Ax_k^{\delta} - b\|^2 + \rho_2 \|W x_k^{\delta} - y_k^{\delta}\|^2 + \rho_2 \|y_k^{\delta} - y_{k-1}^{\delta}\|^2 \) for \( k \geq 1 \). Then

\[
E_{k+1}^{\delta} - E_k^{\delta} \leq -\rho_1 \|A(x_{k+1}^{\delta} - x_k^{\delta})\|^2 - 4\rho_2 \|y_k^{\delta} - y_{k-1}^{\delta}\|^2.
\]

Consequently \( \{E_k^{\delta}\} \) is monotonically decreasing along the iteration and there hold

\[
\sum_{k=m}^{n-1} \|y_k^{\delta} - y_{k+1}^{\delta}\|^2 \leq \frac{1}{4c_0} E_m^{\delta} \quad \text{and} \quad (n-m)\rho_2 \|y_n^{\delta} - y_{n-1}^{\delta}\|^2 \leq \sum_{k=m+1}^{n} E_k^{\delta}
\]

for any integers \( 1 \leq m < n \).

The following result shows that the stopping criterion (2.33) is satisfied for some finite integer, hence Algorithm 1 terminates after a finite number of iterations.

**Lemma 2.10.** There exists a finite integer \( k_\delta \) such that the stop condition (2.33) is satisfied for the first time. Moreover, there exist positive constants \( c \) and \( C \) depending only on \( \rho_2, \tau \), and \( c_0 \) such that

\[
(2.35) \quad D_{\mu_2} f(\hat{y}, \hat{y}_n^{\delta}) + c \sum_{k=m}^{n} E_k^{\delta} \leq D_{\mu_2} f(\hat{y}, y_n^{\delta}) + \rho_2 \langle y_{n-1}^{\delta} - y_m^{\delta}, W(\hat{x} - x_{m+1}^{\delta}) \rangle + C \left( \|W(\hat{x} - x_m^{\delta})\|^2 + \|s_m^{\delta}\|^2 + E_m^{\delta} \right)
\]

for any integers \( 1 \leq m < n < k_\delta \) and any feasible point \((\hat{x}, \hat{y})\) of (1.4).
Proof. Similarly to the derivation of (2.20) and using (2.34) we can obtain for \( k \geq 1 \) that

\[
D_{\mu_k^+} f(\hat{y}, y_{k+1}^\delta) - D_{\mu_k^-} f(\hat{y}, y_k^\delta) + D_{\mu_k^+} f(y_{k+1}^\delta, y_k^\delta)
\]

\[
= -\rho_2 \|s_{k+1}^\delta\|^2 - \rho_1 \|r_{k+1}^\delta\|^2 + \rho_1 \langle r_{k+1}^\delta, b - b^\delta \rangle
\]

\[
+ \rho_2 \langle y_{k-1}^\delta - y_k^\delta, W(x - x_{k+1}^\delta) \rangle
\]

\[
- \rho_2 \langle y_k^\delta - y_{k+1}^\delta, W(x - x_{k+1}^\delta) \rangle
\]

\[
\leq -\rho_2 \|s_{k+1}^\delta\|^2 - \rho_1 \|r_{k+1}^\delta\|^2 + \rho_1 \delta \|r_{k+1}^\delta\| + \rho_2 \langle y_{k-1}^\delta - y_k^\delta, W(x - x_{k+1}^\delta) \rangle
\]

\[
- \rho_2 \langle y_k^\delta - y_{k+1}^\delta, W(x - x_{k+1}^\delta) \rangle.
\]

By using the formulation of the stop criterion (2.33), we can see that for any \( k \) satisfying \( 1 \leq k < k_3 - 1 \) there holds

\[
\rho_1 \delta \|r_{k+1}^\delta\| \leq \frac{\rho_1^2 \|r_{k+1}^\delta\|^2 + \rho_2^2 \|s_{k+1}^\delta\|^2}{\tau \max(\rho_1, \rho_2)} \leq \frac{1}{\tau} (\rho_1 \|r_{k+1}^\delta\|^2 + \rho_2 \|s_{k+1}^\delta\|^2).
\]

By the strong convexity of \( f \) we have \( D_{\mu_k^+} f(y_{k+1}^\delta, y_k^\delta) \geq c_0 \|y_{k+1}^\delta - y_k^\delta\|^2 \). Therefore, by setting \( c_2 := \min(1 - 1/\tau, c_0/\rho_2) > 0 \) and using the definition of \( E_k^\delta \) we have

\[
D_{\mu_k^+} f(\hat{y}, y_{k+1}^\delta) - D_{\mu_k^-} f(\hat{y}, y_k^\delta) + c_2 E_k^\delta
\]

\[
\leq \rho_2 \langle y_{k-1}^\delta - y_k^\delta, W(x - x_{k+1}^\delta) \rangle - \rho_2 \langle y_k^\delta - y_{k+1}^\delta, W(x - x_{k+1}^\delta) \rangle.
\]

For any two integers \( 1 \leq m < n < k_3 \), we sum (2.37) over \( k \) from \( k = m \) to \( k = n - 1 \) to derive that

\[
D_{\mu_k^+} f(\hat{y}, y_n^\delta) + c_2 \sum_{k=m+1}^{n} E_k^\delta
\]

\[
\leq D_{\mu_k^+} f(\hat{y}, y_m^\delta) + \rho_2 \langle y_{m-1}^\delta - y_m^\delta, W(x - x_{m+1}^\delta) \rangle
\]

\[
- \rho_2 \langle y_{m-1}^\delta - y_m^\delta, W(x - x_m^\delta) \rangle
\]

\[
+ \rho_2 \sum_{k=m}^{n-2} \langle y_k^\delta - y_{k+1}^\delta, W(x_{k+1}^\delta - x_{k+2}^\delta) \rangle.
\]

Let \( \varepsilon > 0 \) be a small number which will be specified later. By using the Cauchy–Schwarz inequality and the similar arguments for deriving (2.22) and (2.23), we can show that there is constant \( C_\varepsilon > 0 \) depending only on \( \varepsilon \) such that

\[
\sum_{k=m}^{n-2} \langle y_k^\delta - y_{k+1}^\delta, W(x_{k+1}^\delta - x_{k+2}^\delta) \rangle \leq \varepsilon \sum_{k=m+1}^{n} \|s_k^\delta\|^2 + C_\varepsilon \sum_{k=m}^{n-1} \|y_k^\delta - y_{k+1}^\delta\|^2
\]

and

\[
- \langle y_{n-1}^\delta - y_n^\delta, W(x - x_n^\delta) \rangle \leq \varepsilon (\|s_n^\delta\|^2 + \|s_n^\delta\|^2) + \varepsilon (n - m) \|y_{n-1}^\delta - y_n^\delta\|^2
\]

\[
+ \frac{1}{4} \|W(x - x_m^\delta)\|^2 + C_\varepsilon \sum_{k=m}^{n-1} \|y_k^\delta - y_{k+1}^\delta\|^2.
\]
Combining the above two equations with (2.38), we obtain

\[ D_{\mu^k_n} f(\hat{y}, y^\delta_m) + c_2 \sum_{k=m+1}^n E^\delta_k \]
\[ \leq D_{\mu^k_n} f(\hat{y}, y^\delta_m) + \rho_2 \langle y^\delta_{m-1} - y^\delta_m, W(\hat{x} - x^\delta_m) \rangle + \frac{\rho_2^2}{4} \|W(\hat{x} - x^\delta_m)\|^2 + \varepsilon \rho_2 \|s^\delta_m\|^2 \]
\[ + 2\varepsilon \rho_2 \sum_{k=m+1}^n \|s^\delta_k\|^2 + \varepsilon(n - m)\rho_2 \|y^\delta_{m-1} - y^\delta_m\|^2 + 2\rho_2 C_\varepsilon \sum_{k=m}^{n-1} \|y^\delta_k - y^\delta_{k+1}\|^2. \]

By using Lemma 2.9 we further obtain

\[ D_{\mu^k_n} f(\hat{y}, y^\delta_m) + c_2 \sum_{k=m+1}^n E^\delta_k \leq D_{\mu^k_n} f(\hat{y}, y^\delta_m) + \rho_2 \langle y^\delta_{m-1} - y^\delta_m, W(\hat{x} - x^\delta_m) \rangle \]
\[ + \frac{\rho_2^2}{4} \|W(\hat{x} - x^\delta_m)\|^2 + \varepsilon \rho_2 \|s^\delta_m\|^2 + 3\varepsilon \sum_{k=m+1}^n E^\delta_k + \frac{\rho_2 C_\varepsilon}{2c_0} E^\delta_m. \]

Now we take \( \varepsilon = c_2/6. \) Then

\[ D_{\mu^k_n} f(\hat{y}, y^\delta_m) + \frac{c_2}{2} \sum_{k=m+1}^n E^\delta_k \leq D_{\mu^k_n} f(\hat{y}, y^\delta_m) + \rho_2 \langle y^\delta_{m-1} - y^\delta_m, W(\hat{x} - x^\delta_m) \rangle \]
\[ + \frac{\rho_2^2}{4} \|W(\hat{x} - x^\delta_m)\|^2 + \varepsilon \rho_2 \|s^\delta_m\|^2 + \frac{\rho_2 C_\varepsilon}{2c_0} E^\delta_m. \]

This shows (2.35) immediately.

Finally we show that there is a finite integer \( k_\delta \) such that (2.33) is satisfied. If not, then for any \( k \geq 1 \) there holds

\[ \rho_1^2 \|x^\delta_k\|^2 + \rho_2^2 \|s^\delta_k\|^2 > \max(\rho_1^2, \rho_2^2) \tau^2 \delta^2 \]

It then follows from (2.35) with \( m = 1 \) that

\[ c(n - 1) \max(\rho_1, \rho_2) \tau^2 \delta^2 \leq c \sum_{k=2}^n E^\delta_k \leq D_{\mu^k_n} f(\hat{y}, y^\delta_1) + \rho_2 \langle y^\delta_0 - y^\delta_1, W(\hat{x} - x^\delta_1) \rangle \]
\[ + C \left( \|W(\hat{x} - x^\delta_1)\|^2 + \|s^\delta_1\|^2 + E^\delta_1 \right). \]

for any integer \( n \geq 1. \) Letting \( n \to \infty \) yields a contradiction. \( \square \)

Remark 2.11. Let \( k_\delta \) denote the first integer such that (2.33) is satisfied. Then (2.41) holds for all \( n < k_\delta. \) According to Lemma 2.8, the right-hand side of (2.41) can be bounded by a constant independent of \( \delta. \) Thus, we may use it to conclude that \( k_\delta = O(\delta^{-2}). \)

We next derive some estimates which will be crucially used in the proof of regularization property of Algorithm 1.

Lemma 2.12. There exist positive constants \( c \) and \( C \) depending only on \( \rho_2, \tau, \) and \( c_0 \) such that for any integer \( m < k_\delta - 1 \) the hold

\[ D_{\mu^k_n} f(\hat{y}, y^\delta_m) + cE^\delta_k \leq D_{\mu^k_n} f(\hat{y}, y^\delta_m) + \max\{\rho_1, \rho_2\} \tau^2 \delta^2 + C\|W(\hat{x} - x^\delta_m)\|^2 + CE^\delta_m \]
\[ + C\|s^\delta_m\|^2 + C \left| \langle y^\delta_{m-1} - y^\delta_m, W(\hat{x} - x^\delta_{m+1}) \rangle \right| \]
\[ + c \sum_{k=m+1}^n E^\delta_k. \]
and

\[
|\langle \mu_{k_3}^\delta, y_{k_3}^\delta - \tilde{y} \rangle| \leq |\langle \mu_m^\delta, y_{k_3}^\delta - \tilde{y} \rangle| + \max\{\rho_1, \rho_2\} \tau \delta^2 + C \sum_{k=m}^{k_3} E_k^\delta + C \|W(x_{k_3}^\delta - \hat{x})\|^2,
\]

where \((\hat{x}, \tilde{y})\) denotes any feasible point of \((1.4)\).

Proof. By using (2.36) with \(k = k_3 - 1\), the strong convexity of \(f\), and the fact \(\rho_1\|r_{k_3}^\delta\| \leq \max(\rho_1, \rho_2)\tau \delta\), we have

\[
D_{\nu_{k_3}^\delta} f(\tilde{y}, y_{k_3}^\delta) + c_0 \|y_{k_3}^\delta - y_{k_3-1}^\delta\|^2 \leq D_{\nu_{k_3-1}^\delta} f(\tilde{y}, y_{k_3-1}^\delta) - \rho_2 \|s_{k_3}^\delta\|^2
\]

\[
- \rho_1 \|r_{k_3}^\delta\|^2 + \max\{\rho_1, \rho_2\} \tau \delta^2
\]

\[
+ \rho_2 \langle y_{k_3-2}^\delta - y_{k_3-1}^\delta, W(\hat{x} - x_{k_3}^\delta) \rangle
\]

\[
- \rho_2 \langle y_{k_3-1}^\delta - y_{k_3}^\delta, W(\hat{x} - x_{k_3}^\delta) \rangle.
\]

Let \(\varepsilon > 0\) be a small number specified later. Similar to (2.23) and (2.39) we can derive for \(m < k_3 - 1\) that

\[
\langle y_{k_3-2}^\delta - y_{k_3-1}^\delta, W(\hat{x} - x_{k_3}^\delta) \rangle - \langle y_{k_3-1}^\delta - y_{k_3}^\delta, W(\hat{x} - x_{k_3}^\delta) \rangle
\]

\[
\leq \frac{1}{2} \|W(\hat{x} - x_m^\delta)\|^2 + \varepsilon \|s_m^\delta\|^2 + \frac{\varepsilon}{\rho_2} \|E_m^\delta\|^2 + \frac{2\varepsilon}{\rho_2} \sum_{k=m+1}^{k_3-1} E_k^\delta + C_\varepsilon E_m^\delta.
\]

Therefore

\[
D_{\nu_{k_3}^\delta} f(\tilde{y}, y_{k_3}^\delta) + c_0 \|y_{k_3}^\delta - y_{k_3-1}^\delta\|^2 + \rho_1 \|r_{k_3}^\delta\|^2 + \rho_2 \|s_{k_3}^\delta\|^2
\]

\[
\leq D_{\nu_{k_3-1}^\delta} f(\tilde{y}, y_{k_3-1}^\delta) + \max\{\rho_1, \rho_2\} \tau \delta^2 + \frac{\rho_2}{2} \|W(\hat{x} - x_m^\delta)\|^2 + \varepsilon \|s_m^\delta\|^2
\]

\[
+ \varepsilon \|E_m^\delta\|^2 + 2\varepsilon \rho_2 C_\varepsilon E_m^\delta.
\]

By taking \(\varepsilon = \min\{1, c_0/\rho_2\}/2\), we obtain with a constant \(c_3 = \varepsilon\) that

\[
D_{\nu_{k_3}^\delta} f(\tilde{y}, y_{k_3}^\delta) + c_3 E_{k_3}^\delta \leq D_{\nu_{k_3-1}^\delta} f(\tilde{y}, y_{k_3-1}^\delta) + \max\{\rho_1, \rho_2\} \tau \delta^2 + \frac{\rho_2}{2} \|W(\hat{x} - x_m^\delta)\|^2
\]

\[
+ \varepsilon \rho_2 \|s_m^\delta\|^2 + 2\varepsilon \rho_2 C_\varepsilon E_m^\delta.
\]

(2.42)

An application of (2.35) with \(n = k_3 - 1\) then gives the first estimate.

To see the second one, we apply similar argument for deriving (2.26) and the Cauchy–Schwarz inequality to obtain

\[
|\langle \mu_{k_3}^\delta - \mu_m^\delta, y_{k_3}^\delta - \tilde{y} \rangle|
\]

\[
\leq \sum_{k=m+1}^{k_3} E_k^\delta + \rho_1 \|r_{k_3}^\delta - b_{k_3}^\delta\| + \rho_2 \|r_{k_3}^\delta - b_{k_3}^\delta - b_{k_3-1}^\delta\|
\]

\[
+ \rho_2 \|y_{k_3-1}^\delta - y_{k_3}^\delta, W(\hat{x}_{k_3} - \hat{x})\|
\]

\[
\leq \sum_{k=m+1}^{k_3} E_k^\delta + \rho_1 \|r_{k_3}^\delta\| + E_{k_3}^\delta + E_m^\delta + \frac{\rho_2}{2} \|W(\hat{x}_{k_3} - \hat{x})\|^2.
\]
Note that $\rho_1\|r^k_k\| \leq \max(\rho_1, \rho_2)\tau\delta$ and $\max(\rho_1^2, \rho_2^2)\tau^2\delta^2 \leq \rho_1^2\|r^k_k\|^2 + \rho_2^2\|s^k_k\|^2$ for $k < k_3$. We thus obtain

$$\rho_1\delta \sum_{k=m+1}^{k_3} \|r^k_k\| = \rho_1\delta\|r^k_k\| + \sum_{k=m+1}^{k_3-1} \rho_1\delta\|r^k_k\| \leq \max\{\rho_1, \rho_2\}\tau\delta^2 + \frac{1}{\tau^2} \sum_{k=m+1}^{k_3-1} E^k_k.$$ 

Therefore

$$\langle \mu^{\delta}_k - \mu_m^{\delta}, y^\delta_k - \hat{y} \rangle \leq \left(2 + \frac{1}{\tau}\right) \sum_{k=m}^{k_3} E^k_k + \max\{\rho_1, \rho_2\}\tau\delta^2 + \frac{\rho_2}{2}\|W(x^*_k - \hat{x})\|^2$$

which gives the desired estimate.

**Theorem 2.13.** Let (A1)-(A4) hold and let $b$ be consistent. Let $x^*$ be the unique solution of (1.2) and let $y^* = Wx^*$. Let $k_3$ denote the first integer such that (2.33) is satisfied. Then for Algorithm 1 there hold

$$x^k_k \to x^*, \quad y^k_k \to y^*, \quad Wx^k_k \to y^*, \quad f(y^k_k) \to f(y^*), \quad D_{\mu^{\delta}_k} f(y^*, y^\delta_k) \to 0$$

as $\delta \to 0$.

**Proof.** We show the convergence result by considering two cases via a subsequence-subsequence argument.

Assume first that $\{b^{\delta_i}\}$ is a sequence satisfying $\|b^{\delta_i} - b\| \leq \delta_i$ with $\delta_i \to 0$ such that $k_{\delta_i} = k_0$ for all $i$, where $k_0$ is a finite integer. By the definition of $k_{\delta_i}$, we have

$$\rho_1^2\|Ax^{\delta_i}_k - b^{\delta_i}\|^2 + \rho_2^2\|Wx^{\delta_i}_k - y^{\delta_i}_k\|^2 \leq \max(\rho_1^2, \rho_2^2)\tau^2\delta_i^2.$$ 

Letting $i \to \infty$ and using Lemma 2.8, we can obtain $Ax_{k_0} = b$ and $Wx_{k_0} = y_{k_0}$. This together with the definition of $\lambda_k$ and $\mu_k$ implies that $\lambda_{k_0} = \lambda_{k_0-1}$ and $\mu_{k_0} = \mu_{k_0-1}$.

Recall that $\mu_k \in \partial f(y_k)$, we may use (2.3) to obtain

$$0 = \langle \mu_k - \mu_{k-1}, y_k - y_{k-1} \rangle \geq 2C_0\|y_k - y_{k-1}\|^2$$

which implies $y_{k_0} = y_{k_0-1}$. Now we can use (2.5) and (2.6) to conclude that $x_{k_0+1} = x_{k_0}$ and $y_{k_0+1} = y_{k_0}$. Repeating this argument we can derive that $x_k = x_{k_0}$, $y_k = y_{k_0}$, $\lambda_k = \lambda_{k_0}$, and $\mu_k = \mu_{k_0}$ for all $k \geq k_0$. In view of Theorem 2.7, we must have $x_{k_0} = x^*$ and $y_{k_0} = y^*$. With the help of Lemma 2.8, the desired conclusion then follows.

Assume next that $\{b^{\delta_i}\}$ is a sequence satisfying $\|b^{\delta_i} - b\| \leq \delta_i$ with $\delta_i \to 0$ such that $k_{\delta_i} := k_{\delta_i} \to \infty$ as $i \to \infty$. We first show that

$$D_{\mu^{\delta}_k} f(y^*, y^{\delta_k}_k) \to 0 \quad \text{as } i \to \infty.$$ 

Let $m \geq 1$ be any integer. Then $k_i > m + 1$ for large $i$. Thus we may use Lemma 2.12 to conclude that

$$D_{\mu^{\delta}_k} f(y^*, y^{\delta_k}_k) \leq D_{\mu^{\delta}_k} f(y^*, y^{\delta_k}_m) + C\|W(x^* - x^{\delta_m}_m)\|^2 + C\|s^{\delta_k}_m\|^2$$

$$+ CE^{\delta_k}_m + C\left(\langle y^{\delta_k}_m - y^{\delta_k}_{m-1}, W(x^* - x^{\delta_k}_{m+1}) \rangle\right).$$

By virtue of Lemma 2.8, we have

$$\limsup_{i \to \infty} D_{\mu^{\delta}_k} f(y^*, y^{\delta_k}_k) \leq D_{\mu} f(y^*, y_m) + C\|W(x^* - x_m)\|^2 + C\|s_m\|^2 + CE_m$$

$$+ C\|y_{m-1} - y_m, W(x^* - x_{m+1})\|.$$
Letting $m \to \infty$ and using Theorem 2.7, we obtain
\[
\limsup_{i \to \infty} D_{\mu_i} f(y^i, y^\delta_{k_i}) \leq 0
\]
which shows (2.43). Now by using the strong convexity of $f$ we can conclude that $y^\delta_{k_i} \to y^*$ as $i \to \infty$. Since $\rho_1^2\|Ax^\delta_{k_i} - b\| + \rho_2^2\|Wx^\delta_{k_i} - y^\delta_{k_i}\|^2 \leq \max(\rho_1^2, \rho_2^2)\tau^2\delta_i^2$, we also have $Ax^\delta_{k_i} \to b$ and $Wx^\delta_{k_i} \to y^*$ as $i \to \infty$. In view of (A4), we have
\[
c_i \|x^\delta_{k_i} - x^*\| \leq \|Ax^\delta_{k_i} - b\|^2 + \|Wx^\delta_{k_i} - y^*\|^2
\]
which implies that $x^\delta_{k_i} \to x^*$ as $i \to \infty$.

Finally, we show that $f(y^\delta_{k_i}) \to f(y^*)$ as $i \to \infty$. According to (2.43), it suffices to show that
\[
(\mu^\delta_{k_i}, y^* - y^\delta_{k_i}) \to 0 \quad \text{as} \quad i \to \infty.
\]
By virtue of Lemma 2.12, $y^\delta_{k_i} \to y^*$, and $Wx^\delta_{k_i} \to Wx^*$, we have
\[
\limsup_{i \to \infty} \left|\langle \mu^\delta_{k_i}, y^* - y^\delta_{k_i} \rangle \right| \leq C \limsup_{i \to \infty} \sum_{k=m}^{k_i} E^\delta_{k_i}.
\]
In view of Lemmas 2.10, 2.12, and 2.8, we can obtain
\[
\limsup_{i \to \infty} \left|\langle \mu^\delta_{k_i}, y^* - y^\delta_{k_i} \rangle \right| \leq C \left(D_{\mu_m} f(y^*, y_m) + |\langle y_m-1 - y_m, W(x^* - x_{m+1}) \rangle| + \|W(x^* - x_m)\|^2 + \|s_m\|^2 + E_m \right)
\]
for any integer $m$. Letting $m \to \infty$ and using Theorem 2.7, it follows that
\[
\limsup_{i \to \infty} \left|\langle \mu^\delta_{k_i}, y^* - y^\delta_{k_i} \rangle \right| \leq 0
\]
which shows (2.44). The proof is therefore complete.

3. Numerical experiments. In this section we will present various numerical results for one-dimensional as well as two-dimensional problems to show the efficiency of Algorithm 1. All the experiments are done on a four-core laptop with 1.90 GHz and 8 GB RAM. First we give the setup for the data generation, the choice of parameters, and the stopping rule. In all numerical examples the sought solutions $x$ and using Theorem 2.7, we obtain
\[
\|\cdot\|_\ast
\]
for other cases. We also take the initial guess $y_0, \lambda_0, \mu_0$ to be the zero elements and fix $\tau = 1.0001$ in all experiments. The numerical results are not sensitive to the parameters $\rho_1$ and $\rho_2$, so we fix them as $(\rho_1, \rho_2) = (1000, 10)$ in subsections 3.1–3.4. The operators $A, W$ and the noise level $\delta$ will be specified in each example. The ADMM codes can be found at http://xllv.whu.edu.cn/.
ADMM FOR LINEAR INVERSE PROBLEMS

3.1. One-dimensional deconvolution. In this subsection we consider the one-dimensional deconvolution problem of the form

\[ b^\delta(s) = \int_0^1 k(s,t)x(t)dt + \eta(s) := (Ax)(s) + \eta(s) \quad \text{on} \quad [0,1], \]

where \( k(s,t) = \frac{\gamma}{\sqrt{2\pi}} \exp\left(-\frac{(s-t)^2}{2\gamma^2}\right) \) with \( \gamma = 0.01 \). This problem arises from an inverse heat conduction. To find the sought solution \( x^* \) from \( b^\delta \) numerically, we divide \([0,1]\) into \( N = 400 \) subintervals of equal length and approximate integrals by the midpoint rule. Let \( x^* \) be sparse, we take \( W = I \) the identity. Numerical results are reported in Figure 1 which shows that Algorithm 1 can capture the features of solutions as the function \( f \) is properly chosen. Moreover, when the noise level decreases, more iterations are needed and more accurate approximate solutions can be obtained.

3.2. Two-dimensional TV deblurring. In this and next subsections we test the performance of Algorithm 1 on image deblurring problems whose objective is to reconstruct the unknown true image \( x^* \in \mathbb{R}^{M \times N} \) from an observed image \( b^\delta = Ax^* + \eta \) degraded by a linear blurring operator \( A \) and a Gaussian noise \( \eta \). We consider the case that the blurring operator is shift invariant so that \( A \) is a convolution operator whose kernel is a point spread function. This subsection concerns total variation (TV) deblurring [44], assuming periodic boundary conditions on images. To apply Algorithm 1, we take \( W = \nabla \) to be the discrete gradient operator as used in [47, 48]. Correspondingly, the \( x \) subproblem can be solved efficiently by the fast Fourier transform (FFT) and the \( y \) subproblem has an explicit solution given by the soft thresholding [47]. Therefore, Algorithm 1 can be efficiently implemented.

Figure 2 shows numerical results by Algorithm 1 on test images Cameraman (256 × 256) and Pirate (512 × 512) with motion blur (fspecial('motion',35,50)) and Gaussian blur (fspecial('gaussian',[20 20], 20)) respectively. The noise levels are \( \delta = 0.256 \) and 0.511 respectively. As comparisons, we also include the results obtained by FTVd v4.1 in [47] which is a state-of-art algorithm for image deblurring. We can see that the images reconstructed by our proposed ADMM have quality comparable to the ones obtained by FTVd v4.1 with similar PSNR (peak signal-to-noise ratio), while the choice of the regularization parameter is not needed in our algorithm. Here the PSNR is defined by

\[ \text{PSNR} = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}} \quad [\text{dB}], \]

where MSE stands for the mean-squared error per pixel.
3.3. Two-dimensional framelet deblurring. In this subsection we show the performance of Algorithm 1 for image deblurring using wavelet frames [6, 7, 14, 15, 46].

In our numerical simulations on grayscale digital images represented by $M \times N$ arrays, we will use the two-dimensional Haar with three-level decomposition and piece-wise linear $B$-spline framelets with one-level decomposition, which can be constructed by taking tensor products of univariate ones [8]. The action of the discrete framelet transform and its adjoint on images can be implemented implicitly by multi-resolution analysis (MRA)-based algorithms [11]. Assuming periodic boundary condition on the images, the $x$ subproblem in Algorithm 1 then can be solved by FFT. The $y$ subproblem can be solved by the soft thresholding. Thus, Algorithm 1 can be efficiently implemented. Figure 3 shows the reconstruction results using the test images Phantom (256 × 256) and Peppers (256 × 256) with motion blur (fspecial('motion',50,90)) and Gaussian blur (fspecial('gaussian',[20 20], [30])) respectively. The noise level is $\delta = 0.256$ for the two examples. These results indicate the satisfactory performance of our proposed ADMM.

3.4. Semiconvergence. In this subsection we investigate the numerical performance of our ADMM if it is not terminated properly. We use the two-dimensional Cameraman TV deblurring as an example and run our ADMM until a preassigned maximum number of iterations (500) is achieved. In Figure 4 we plot the corresponding results on the PSNR values and the values of $E_k^x$ versus the number of iterations. It turns out that the PSNR increases first and then decreases after a critical number of iterations. This illustrates the semiconvergence property of our proposed ADMM in the framework of iterative regularization methods and indicates the importance of terminating the iteration by a suitable stop rule.

3.5. Sensitivity to $\rho_1$ and $\rho_2$. In this subsection we illustrate that the performance of Algorithm 1 is not quite sensitive to the two parameters $\rho_1$ and $\rho_2$. To this end we run our ADMM on the Cameraman TV deblurring problem with same parameters as in section 3.2 with different $\rho_1$ and $\rho_2$. The PSNR and the number of iterations are given in Table 1, from which we can see that the reconstructions remain stable for a large range of values of $\rho_1$ and $\rho_2$. 

Fig. 2. ADMM TV debluring.
ADMM FOR LINEAR INVERSE PROBLEMS

Fig. 3. ADMM framelet.

Fig. 4. Semiconvergence test.

Table 1
PSNR and number of iteration for different $\rho_1$ and $\rho_2$.

<table>
<thead>
<tr>
<th>$\rho_2$</th>
<th>$\rho_1$</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>(31.9, 61)</td>
<td>(32.1, 30)</td>
<td>(32.7, 15)</td>
<td>(31.7, 9)</td>
<td>(30.6, 5)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(32.0, 64)</td>
<td>(32.9, 33)</td>
<td>(32.4, 17)</td>
<td>(31.6, 10)</td>
<td>(31.5, 6)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(32.5, 68)</td>
<td>(33.0, 36)</td>
<td>(32.1, 20)</td>
<td>(32.2, 11)</td>
<td>(32.0, 7)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>(33.1, 74)</td>
<td>(32.3, 40)</td>
<td>(32.5, 23)</td>
<td>(32.7, 14)</td>
<td>(32.4, 9)</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>(32.8, 82)</td>
<td>(32.6, 47)</td>
<td>(32.8, 28)</td>
<td>(32.5, 17)</td>
<td>(32.0, 11)</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>(32.7, 97)</td>
<td>(32.9, 58)</td>
<td>(32.8, 36)</td>
<td>(32.3, 23)</td>
<td>(31.6, 15)</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>(32.9, 120)</td>
<td>(32.9, 76)</td>
<td>(32.5, 50)</td>
<td>(31.9, 34)</td>
<td>(31.3, 24)</td>
<td></td>
</tr>
</tbody>
</table>

4. Conclusion. In this work we propose an alternating direction method of multiplies to solve inverse problems. When the data is given exactly, we prove the convergence of the algorithm without using the existence of Lagrange multipliers. When the data contain noise, we propose a stop rule and show that our ADMM becomes a regularization method. Numerical simulations are given to show the efficiency of the proposed algorithm.

There are several possible extensions for this work. First, in our ADMM for solving inverse problems, we used two parameters $\rho_1$ and $\rho_2$ which are fixed during iterations. It is natural to consider the situation that $\rho_1$ and $\rho_2$ change dynamically.
Variable step sizes have been used in the augmented Lagrangian method to reduce the number of iterations (see [20, 19, 34]). It would be interesting to investigate what will happen if dynamically changing parameters $\rho_1$ and $\rho_2$ are used in our ADMM. Second, the $x$ subproblem in our ADMM requires one to solve linear systems related to $\rho_1 A^* A + \rho_2 W^* W$. In general, solving such linear systems is very expensive. It might be possible to remedy this drawback by applying linearization and/or precondition strategies. Finally, in applications where the sought solution is a priori known to satisfy certain constraints, it is of interest to consider how to incorporate such constraints into our ADMM in an easily implementable way and to prove some convergence results.

REFERENCES

ADMM FOR LINEAR INVERSE PROBLEMS