SIAM J. OPTIM.
Vol. 24, No. 3, pp. 1011–1040
© 2014 Society for Industrial and Applied Mathematics

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

A STRICTLY CONTRACTIVE PEACEMAN–RACHFORD SPLITTING METHOD FOR CONVEX PROGRAMMING*

BINGSHENG HE†, HAN LIU‡, ZHAORAN WANG§, AND XIAOMING YUAN¶

Abstract. In this paper, we focus on the application of the Peaceman–Rachford splitting method (PRSM) to a convex minimization model with linear constraints and a separable objective function. Compared to the Douglas–Rachford splitting method (DRSM), another splitting method from which the alternating direction method of multipliers originates, PRSM requires more restrictive assumptions to ensure its convergence, while it is always faster whenever it is convergent. We first illustrate that the reason for this difference is that the iterative sequence generated by DRSM is strictly contractive, while that generated by PRSM is only contractive with respect to the solution set of the model. With only the convexity assumption on the objective function of the model under consideration, the convergence of PRSM is not guaranteed. But for this case, we show that the first t iterations of PRSM still enable us to find an approximate solution with an accuracy of $O(1/t)$. A worst-case $O(1/t)$ convergence rate of PRSM in the ergodic sense is thus established under mild assumptions. After that, we suggest attaching an underdetermined relaxation factor with PRSM to guarantee the strict contraction of its iterative sequence and thus propose a strictly contractive PRSM. A worst-case $O(1/t)$ convergence rate of this strictly contractive PRSM in a nonergodic sense is established. We show the numerical efficiency of the strictly contractive PRSM by some applications in statistical learning and image processing.

Key words. convex programming, Peaceman–Rachford splitting method, convergence rate, contraction

AMS subject classifications. 90C25, 90C30, 94A08

DOI. 10.1137/13090849X

1. Introduction.

We consider the following convex minimization model with linear constraints and a separable objective function:

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\, x \in \mathcal{X}, y \in \mathcal{Y}\}, \tag{1.1}$$

where $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$, $b \in \Re^m$, $\mathcal{X} \subset \Re^{n_1}$ and $\mathcal{Y} \subset \Re^{n_2}$ are closed convex sets, and $\theta_1 : \mathcal{X} \to \Re$ and $\theta_2 : \mathcal{Y} \to \Re$ are convex functions. Note that both θ_1 and θ_2 could be nonsmooth functions. Throughout, the solution set of (1.1) (denoted by S^*) is assumed to be nonempty.

When $A = I_m$, $B = -I_m$, $b = 0$, $X = \Re^m$, and $Y = \Re^m$, the model (1.1) reduces to

$$\min\{\theta_1(x) + \theta_2(x) \mid x \in \Re^m\}, \tag{1.2}$$

*Received by the editors February 4, 2013; accepted for publication (in revised form) March 17, 2014; published electronically July 17, 2014.
http://www.siam.org/journals/siopt/24-3/90849.html

†International Centre of Management Science and Engineering, and Department of Mathematics, Nanjing University, Nanjing, 200093, China (hebma@nju.edu.cn). This author was supported by NSFC grant 91130007 and MOEC fund 20110091110004.

‡Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544 (hanliu@princeton.edu). This author was supported by NSF grant III–1116730.

§Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 (zhaoran@princeton.edu).

¶Department of Mathematics, Hong Kong Baptist University, Hong Kong (xmyuan@hkbu.edu.hk). This author was partially supported by the General Research Fund from Hong Kong Research Grants Council: 12302514.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where θ_1 and θ_2 can be explained, respectively, as data-fidelity and regularization terms for some ill-posed inverse problems arising widely in statistical learning and image processing areas. Because of the separable structure in its objective function, the majority of efficient solvers for (1.1) in the literature belong to the category of the splitting method, which is effective for taking advantage of the properties of the functions θ_1 and θ_2 individually in the algorithmic design. The resulting subproblems are usually easy enough to have closed-form solutions or can be solved easily up to high precisions; therefore, the implementation of such an algorithm is extremely easy and its convergence is fast. A benchmark is the alternating direction method of multipliers (ADMM) proposed in [24] (see also [9, 21]), which has received tremendous attention from a number of areas recently (see, e.g., [6, 17, 22] for review papers). As analyzed in [20], ADMM is just an application of the Douglas–Rachford splitting method (DRSM) in [15, 39] to the dual problem of (1.1), and its iterative scheme reads as

(1.3)
$$\left\{\begin{array}{l} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \;\big|\; x \in \mathcal{X}\big\}, \\ y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \;\big|\; y \in \mathcal{Y}\big\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), \end{array}\right.$$

where $\lambda \in \Re^m$ is the Lagrange multiplier associated with the linear constraints in (1.1) and $\beta > 0$ is a penalty parameter.

In this paper, we focus on the application of the Peaceman–Rachford splitting method (PRSM) in [39, 46] to (1.1). As elaborated on in [20], applying PRSM to the dual of (1.1), we obtain the iterative scheme of PRSM for (1.1),

(1.4)
$$\left\{\begin{array}{l} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \;\big|\; x \in \mathcal{X}\big\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \;\big|\; y \in \mathcal{Y}\big\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), \end{array}\right.$$

where $\lambda \in \Re^m$ and β have the same meaning as (1.3). As analyzed in [20], the PRSM scheme (1.4) differs from ADMM "only through the addition of the intermediate update of the multipliers (i.e., $\lambda^{k+\frac{1}{2}}$); it thus offers the same set of advantages." The PRSM scheme (1.4), however, according to [20] again (see also [23]), "is less 'robust' in that it converges under more restrictive assumptions than ADMM." Also, it was remarked in [20] that the PRSM scheme (1.4) with optimal parameters converges on the linear rate if Lipschitz continuity and coercivity of $\partial\theta_2^*$ (where θ_2^* denotes the conjugate function of θ_2) are assumed. We refer the reader to [3, 25] for some numerical verification of the efficiency of PRSM.

We first show that the difference between DRSM and PRSM in convergence can be illustrated by the contraction property of their iterative sequences.[1] More specifically, the iterative sequence generated by DRSM is strictly contractive with respect to the solution set of (1.1), as proved in [29], while this does not hold for the iterative sequence generated by PRSM (see (3.20)). This difference is also the reason a worst-case $O(1/t)$ convergence rate of ADMM for (1.1) in a nonergodic sense can be established in [32], while we can only establish the same convergence rate in the

[1]Fort the definition of a contractive sequence we refer the reader to [5].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ergodic sense for the PRSM scheme (1.4);[2] see Theorem 4.1. Also inspired by the failure of strict contraction of PRSM's iterative sequence, we find that when an underdetermined relaxation factor $\alpha \in (0,1)$ is attached to the penalty parameter β in the steps of Lagrange multiplier updating in (1.4), the resulting sequence becomes strictly contractive with respect to the solution set of (1.1). This strict contraction property makes it possible to establish a worst-case $O(1/t)$ convergence rate in a nonergodic sense for the PRSM (1.4) with an underdetermined relaxation factor, which will be named as a strictly contractive PRSM from now on. Let us specify the iterative scheme of the strictly contractive PRSM for (1.1):

(1.5)
$$\left\{\begin{array}{l} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \;\big|\; x \in \mathcal{X}\big\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \;\big|\; y \in \mathcal{Y}\big\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b), \end{array}\right.$$

where $\alpha \in (0,1)$. Note that we follow the standard terminology in numerical linear algebra and call $\alpha \in (0,1)$ an underdetermined relaxation factor; see also [18, 26]. As we shall show, the consideration of an additional relaxation factor in the PRSM scheme (1.5) ensures the sequence generated by (1.5) to be strictly contractive with respect to the solution set of (1.1). Thus we can establish some worst-case convergence rates for (1.5) without any further assumption on the model (1.1). Numerically, we can simply choose α close to 1.

The rest of this paper is organized as follows. In section 2, we summarize some useful preliminary results and prove some simple assertions for further analysis. Then, we prove some properties for the sequence generated by the strictly contractive PRSM (1.5) in section 3. In section 4, we establish a worst-case $O(1/t)$ convergence rate in the ergodic sense for the PRSM scheme (1.4); and in section 5, we establish a worst-case $O(1/t)$ convergence rate in a nonergodic sense for the strictly contractive PRSM scheme (1.5). Then, we show the numerical efficiency of the strictly contractive PRSM in section 6 by some applications to statistical learning and image processing. Some comparisons with existing efficient methods are also reported. Finally, some conclusions are drawn in section 7.

2. Preliminaries.

In this section, we summarize some useful preliminaries known in the literature and prove some simple conclusions for further analysis.

2.1. Variational reformulation of (1.1).

First, as in the work [31, 32] for analyzing the convergence rate of ADMM, we need a variational inequality (VI) reformulation of the model (1.1) and a characterization of its solution set. More specifically, solving (1.1) is equivalent to finding $w^* = (x^*, y^*, \lambda^*) \in \Omega := \mathcal{X} \times \mathcal{Y} \times \Re^m$ such that

$$\text{VI}(\Omega, F, \theta): \qquad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0 \quad \forall\, w \in \Omega, \tag{2.1a}$$

[2] We follow [42, 43] and many others to measure the worst-case convergence rate in terms of the iteration complexity. That is, a worst-case $O(1/t)$ convergence rate means the accuracy of a solution under certain criteria is of the order $O(1/t)$ after t iterations of an iterative scheme; or, equivalently, it requires at most $O(1/\epsilon)$ iterations to achieve an approximate solution with an accuracy of ϵ.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where
(2.1b)

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \; w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \; F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \text{ and } \theta(u) = \theta_1(x) + \theta_2(y).$$

Since the mapping $F(w)$ defined in (2.1b) is affine with a skew-symmetric matrix, it is monotone. We denote by Ω^* the solution set of $\mathrm{VI}(\Omega, F, \theta)$, and it is not nonempty under our nonempty assumption onto S^*.

According to Theorem 2.3.5 in [19], a very useful characterization of Ω^* can be summarized in the following theorem. Its proof can be found in [19, 31].

THEOREM 2.1. *The solution set of VI(Ω, F, θ) is closed and convex, and it can be characterized as*

$$\Omega^* = \bigcap_{w\in\Omega} \big\{\tilde{w} \in \Omega : \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0\big\}. \tag{2.2}$$

Theorem 2.1 thus implies that $\tilde{w} \in \Omega$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with an accuracy of $O(1/t)$ if it satisfies

$$\sup_{w\in\mathcal{D}} \{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\} \leq \epsilon, \tag{2.3}$$

with $\epsilon = O(1/t)$ and $\mathcal{D} \subset \Omega$ a compact set.[3] In fact, this characterization makes it possible to analyze the convergence rate of ADMM and other splitting methods via the VI approach rather than the conventional approach based on the functional values in the literature. In the following, we shall show that either sequence (1.4) or sequence (1.5) enables us to find an approximate solution of (1.1) in the sense of (2.3) after t iterations.

2.2. Some notation. As mentioned in [6] for ADMM, the variable x is an intermediate variable during the PRSM iteration since it essentially requires only (y^k, λ^k) in (1.4) or (1.5) to generate the $(k+1)$th iterate. For this reason, we define the notation $v^k = (y^k, \lambda^k)$, $\mathcal{V} = \mathcal{Y} \times \Re^m$,

$$\mathcal{V}^* := \{v^* = (y^*, \lambda^*) \,|\, w^* = (x^*, y^*, \lambda^*) \in \Omega^*\};$$

it suffices to analyze the convergence rate of the sequence $\{v^k\}$ to the set $\mathcal{V}^*$ in order to study the convergence rate of the sequence $\{w^k\}$ generated by (1.4) or (1.5). Note that $\mathcal{V}^*$ is also closed and convex.

Then, we define some matrices in order to present our analysis in a compact way. Let

$$M = \begin{pmatrix} I_{n_2} & 0 \\ -\alpha\beta B & 2\alpha I_m \end{pmatrix} \tag{2.4}$$

and

$$Q_0 = \begin{pmatrix} 0 & 0 \\ \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.5}$$

[3] As in [44], the compact set $\mathcal{D}$ can be chosen as $\mathcal{D}_{\mathcal{W}}(\tilde{w}) = \big\{w \in \mathcal{W} \mid \|w - \tilde{w}\| \leq 1\big\}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In (2.5), "0" is a matrix with all zero entries in appropriate dimensionality. We further define

$$Q = \begin{pmatrix} \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \tag{2.6}$$

as the submatrix of Q_0 excluding all the first zero rows. The matrices Q_0 and Q are associated with the analysis for the sequences $\{w^k\}$ and $\{v^k\}$, respectively. Last, for $\alpha \in (0, 1]$ we define a symmetric matrix

$$H = \frac{1}{2} \begin{pmatrix} (2-\alpha)\beta B^T B & -B^T \\ -B & \frac{1}{\alpha\beta} I_m \end{pmatrix}. \tag{2.7}$$

Below we prove some assertions regarding the matrices just defined. These assertions will be used in our theoretical analysis about the convergence rate of (1.4) and (1.5); their role is to make our proof presentable in compact notation.

LEMMA 2.2. *The matrix H defined in* (2.7) *is positive definite (if B is a full column rank matrix) for $\alpha \in (0, 1)$ and positive semidefinite for $\alpha = 1$.*

Proof. We have

$$H = \frac{1}{2} \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I_m \end{pmatrix} \begin{pmatrix} (2-\alpha) I_m & -I_m \\ -I_m & \frac{1}{\alpha} I_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I_m \end{pmatrix}.$$

Note that the matrix

$$\begin{pmatrix} (2-\alpha) & -1 \\ -1 & \frac{1}{\alpha} \end{pmatrix}$$

is positive definite if $\alpha \in (0, 1)$ and positive semidefinite if $\alpha = 1$. The assertion of this lemma is thus proved. ∎

LEMMA 2.3. *The matrices M, Q, and H defined, respectively, in* (2.4), (2.6), *and* (2.7) *have the following relationships:*

$$HM = Q \tag{2.8}$$

and

$$Q^T + Q - M^T H M \succeq \frac{(1-\alpha)}{2(1+\alpha)} M^T H M. \tag{2.9}$$

Proof. Using the definitions of the matrices M, Q, and H, by a simple manipulation, we obtain

$$\begin{aligned} HM &= \frac{1}{2} \begin{pmatrix} (2-\alpha)\beta B^T B & -B^T \\ -B & \frac{1}{\alpha\beta} I_m \end{pmatrix} \begin{pmatrix} I_{n_2} & 0 \\ -\alpha\beta B & 2\alpha I_m \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 2\beta B^T B & -2\alpha B^T \\ -2B & \frac{2}{\beta} I_m \end{pmatrix} = Q. \end{aligned}$$

The first assertion is proved. Consequently, we get

$$\begin{aligned} M^T H M = M^T Q &= \begin{pmatrix} I_{n_2} & -\alpha\beta B^T \\ 0 & 2\alpha I_m \end{pmatrix} \begin{pmatrix} \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} (1+\alpha)\beta B^T B & -2\alpha B^T \\ -2\alpha B & \frac{2\alpha}{\beta} I_m \end{pmatrix}. \end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Using (2.6) and the above equation, we have

$$(Q^T + Q) - M^T H M = (1-\alpha)\begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{2.10}$$

Note that

$$2(1+\alpha)\begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix} - M^T H M = \begin{pmatrix} (1+\alpha)\beta B^T B & -2B^T \\ -2B & \frac{4+2\alpha}{\beta} I_m \end{pmatrix}. \tag{2.11}$$

Because

$$\begin{pmatrix} (1+\alpha) & -2 \\ -2 & 4+2\alpha \end{pmatrix} \succeq 0 \quad \forall\, \alpha \geq 0,$$

the right-hand side of (2.11) is positive semidefinite. Thus, it follows that

$$\begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix} \succeq \frac{1}{2(1+\alpha)} M^T H M. \tag{2.12}$$

Substituting (2.12) into (2.10), we obtain (2.9), and the lemma is proved. ∎

Remark 2.1. When $\alpha = 1$, the matrices H defined in (2.7) and $Q^T + Q - M^T H M$ are both positive semidefinite. However, in the following analysis we still use $\|v - \tilde{v}\|_H$ and $\|v - \tilde{v}\|_{(Q^T+Q-M^T HM)}$ to denote, respectively,

$$\|v - \tilde{v}\|_H = \big((v-\tilde{v})^T H (v - \tilde{v})\big)^{1/2}$$

and

$$\|v - \tilde{v}\|_{(Q^T+Q-M^T HM)} = \big((v-\tilde{v})^T (Q^T + Q - M^T H M)(v - \tilde{v})\big)^{1/2}$$

for v, $\tilde{v} \in \mathcal{Y} \times \Re^m$. This slight abuse of notation will simplify the notation in our analysis greatly.

3. Contraction analysis. In this section, we analyze the contraction property for the sequence $\{v^k\}$ generated by the PRSM scheme (1.4) or the strictly contractive PRSM scheme (1.5) with respect to the set $\mathcal{V}^*$. The convergence rate analysis for (1.4) and (1.5) to be presented is based on this analysis of contraction property. Since (1.4) can be included by the strictly contractive PRSM scheme (1.5) if we extend the value of $\alpha = 1$ and if the algebra of convergence analysis for these two schemes are of the same framework, below we only present the contraction analysis for (1.5); the analysis for (1.4) is readily obtained by taking $\alpha = 1$ in our analysis.

First, to further simplify the notation in our analysis, we need to define an auxiliary sequence $\{\tilde{w}^k\}$ as

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{3.1}$$

where (x^{k+1}, y^{k+1}) is generated by (1.4) or (1.5). Note that with the notation of $\tilde{w}^k$, we immediately have

$$x^{k+1} = \tilde{x}^k, \qquad y^{k+1} = \tilde{y}^k, \qquad \text{and} \qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k). \tag{3.2}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The strictly contractive PRSM (1.5) can be written as

$$\left\{\begin{array}{l} \tilde{x}^k = \arg\min\left\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\right\}, \\ \tilde{y}^k = \arg\min\left\{\theta_2(y) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)]^T By + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\right\}. \end{array}\right. \tag{3.3}$$

Then, based on (1.5) and (3.2), we immediately get

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \alpha\beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\big[\beta(A\tilde{x}^k + By^k - b) - \beta B(y^k - \tilde{y}^k)\big] \\ &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\big[(\lambda^k - \tilde{\lambda}^k) - \beta B(y^k - \tilde{y}^k)\big] \\ &= \lambda^k - \big[2\alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\beta B(y^k - \tilde{y}^k)\big]. \end{aligned} \tag{3.4}$$

Furthermore, together with $y^{k+1} = \tilde{y}^k$, we have the relationship

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_2} & 0 \\ -\alpha\beta B & 2\alpha I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix},$$

which can be rewritten in a compact form by using the notation of v^k and $\tilde{v}^k$:

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{3.5}$$

where M is as defined in (2.4).

Now, we start to prove some properties for the sequence $\{\tilde{w}^k\}$ defined in (3.1). Recall that our primary purpose is to analyze the convergence rate for the sequences (1.4) and (1.5) based on the solution characterization (2.2), and the accuracy of an approximate solution $\tilde{w} \in \Omega$ is measured by an upper bound of the quantity of $\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)$ for all $w \in \Omega$ (see (2.3)). Hence, we are interested in estimating how accurate the point $\tilde{w}^k$ defined in (3.1) is to a solution point of $\mathrm{VI}(\Omega, F, \theta)$. The main result is proved in Theorem 3.4. But before that, we first show some lemmas. The first lemma presents an upper bound of $\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)$ for all $w \in \Omega$ in terms of a quadratic term involving the matrix Q.

LEMMA 3.1. *For given $v^k \in \mathcal{Y} \times \Re^m$, let w^{k+1} be generated by the strictly contractive PRSM scheme* (1.5)*, and let $\tilde{w}^k$ be as defined in* (3.1)*. Then, we have $\tilde{w}^k \in \Omega$ and*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) \quad \forall\, w \in \Omega, \tag{3.6}$$

where the matrix Q is as defined in (2.6)*.*

Proof. Since $x^{k+1} = \tilde{x}^k$, by deriving the first-order optimality condition of the x-minimization problem in (3.3), we have

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{A^T[\beta(A\tilde{x}^k + By^k - b) - \lambda^k]\} \geq 0 \quad \forall\, x \in \mathcal{X}. \tag{3.7}$$

According to the definition (3.1), we have

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \tag{3.8}$$

Using (3.8), the inequality (3.7) can be written as

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \geq 0 \quad \forall\, x \in \mathcal{X}. \tag{3.9}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Similarly, by deriving the first-order optimality condition of the y-minimization problem in (3.3), we get

$$\theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T B^T\big\{\beta(A\tilde{x}^k+B\tilde{y}^k-b)-[\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)]\big\} \geq 0 \;\forall\, y \in \mathcal{Y}. \tag{3.10}$$

Again, using (3.8), we have

$$\begin{aligned}&\beta(A\tilde{x}^k + B\tilde{y}^k - b) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)]\\ &= -(\tilde{\lambda}^k - \lambda^k) + \beta B(\tilde{y}^k - y^k) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)]\\ &= -\tilde{\lambda}^k + \beta B(\tilde{y}^k - y^k) - \alpha(\tilde{\lambda}^k - \lambda^k).\end{aligned}$$

Consequently, it follows from (3.10) that

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - \alpha B^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0 \;\forall\, y \in \mathcal{Y}. \tag{3.11}$$

In addition, based on (3.1) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \tag{3.12}$$

Combining (3.9), (3.11), and (3.12), we get $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$; and for any $w = (x, y, \lambda) \in \Omega$, it holds that

$$\theta(u)-\theta(\tilde{u}^k)+\begin{pmatrix} x-\tilde{x}^k \\ y-\tilde{y}^k \\ \lambda-\tilde{\lambda}^k \end{pmatrix}^T \left\{\begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k+B\tilde{y}^k-b \end{pmatrix}+\begin{pmatrix} 0 \\ \beta B^T B(\tilde{y}^k-y^k)-\alpha B^T(\tilde{\lambda}^k-\lambda^k) \\ -B(\tilde{y}^k-y^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k) \end{pmatrix}\right\} \geq 0.$$

The assertion (3.6) is only a compact form of the above inequality by using the notation of Q in (2.6), w and F in (2.1b), and v. The proof is complete. ∎

Based on the optimality condition (2.1) and Lemma 3.1, we can prove the following lemma, which makes it possible to measure the accuracy of $\tilde{w}^k$ to a solution point in $\mathcal{W}^*$ by the quantity $\|v^k - v^{k+1}\|_H^2$. This is also an important assertion to establish a nonergodic convergence rate for the proposed strictly contractive PRSM in section 5.

LEMMA 3.2. *Let $\{w^k\}$ be generated by the strictly contractive PRSM scheme* (1.5)*, and let $\{\tilde{w}^k\}$ be as defined in* (3.1)*; let M, Q, and H be as defined in* (2.4), (2.6), *and* (2.7)*, respectively. Then, $\tilde{w}^k$ is a solution of VI(Ω, F, θ) if $\|v^k - v^{k+1}\|_H^2 = 0$.*

Proof. By using $Q = HM$ and $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (see (2.8) and (3.5)), it follows that

$$(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) = (v - \tilde{v}^k)^T H(v^k - v^{k+1}). \tag{3.13}$$

Substituting this into (3.6), we get

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}) \quad \forall\, w \in \Omega. \tag{3.14}$$

Note that $\tilde{w}^k \in \Omega$. Since H is positive semidefinite, in the case $\|v^k - v^{k+1}\|_H^2 = 0$, we have $H(v^k - v^{k+1}) = 0$, and thus

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq 0 \quad \forall\, w \in \Omega.$$

According to (2.1), $\tilde{w}^k$ is a solution of VI(Ω, F, θ). ∎

In the next lemma, we aim at further bounding the term $(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k)$ found in Lemma 3.1 by the difference of two quadratic terms involving two consecutive

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

iterates of the sequence $\{v^k\}$ and a quadratic term involving v^k and the auxiliary iterate $\tilde{v}^k$. This refined bound is convenient for the manipulation over the whole sequence $\{v^k\}$ recursively and thus for establishing the convergence rate of $\{v^k\}$ in either the ergodic or a nonergodic sense.

LEMMA 3.3. *Let $\{w^k\}$ be generated by the strictly contractive PRSM scheme* (1.5)*, and let $\{\tilde{w}^k\}$ be as defined in* (3.1)*; let M, Q, and H be as defined in* (2.4), (2.6)*, and* (2.7)*, respectively. Then we have*

$$(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) \geq \frac{1}{2}\big(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2\big) + \frac{(1-\alpha)}{4(1+\alpha)}\|v^k-v^{k+1}\|_H^2 \ \forall v \in \mathcal{V}. \tag{3.15}$$

Proof. For the vectors a, b, c, d in the same space and a matrix H with appropriate dimensionality, we have the identity

$$(a-b)^T H(c-d) = \frac{1}{2}\{\|a-d\|_H^2 - \|a-c\|_H^2\} + \frac{1}{2}\{\|c-b\|_H^2 - \|d-b\|_H^2\}.$$

In this identity, we take

$$a = v, \quad b = \tilde{v}^k, \quad c = v^k, \quad \text{and} \quad d = v^{k+1}$$

and substitute it into the right-hand side of (3.13). The resulting equation is

$$(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) = \frac{1}{2}\big(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2\big) + \frac{1}{2}(\|v^k-\tilde{v}^k\|_H^2 - \|v^{k+1}-\tilde{v}^k\|_H^2). \tag{3.16}$$

Now, we deal with the last term of the right-hand side of (3.16). By using (3.5) and (2.8), we get

$$\begin{aligned}
&\|v^k-\tilde{v}^k\|_H^2 - \|v^{k+1}-\tilde{v}^k\|_H^2 \\
&\quad = \|v^k-\tilde{v}^k\|_H^2 - \|(v^k-\tilde{v}^k)-(v^k-v^{k+1})\|_H^2 \\
&\quad \overset{(3.5)}{=} \|v^k-\tilde{v}^k\|_H^2 - \|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2 \\
&\quad = 2(v^k-\tilde{v}^k)^T HM(v^k-\tilde{v}^k) - (v^k-\tilde{v}^k)^T M^T HM(v^k-\tilde{v}^k) \\
&\quad \overset{(2.8)}{=} (v^k-\tilde{v}^k)^T(Q^T+Q-M^T HM)(v^k-\tilde{v}^k) \\
&\quad \overset{(2.9)}{\geq} \frac{(1-\alpha)}{2(1+\alpha)}(v^k-\tilde{v}^k)^T M^T HM(v^k-\tilde{v}^k) \\
&\quad \overset{(3.5)}{=} \frac{(1-\alpha)}{2(1+\alpha)}\|v^k-v^{k+1}\|_H^2.
\end{aligned}$$

Substituting into (3.16), we obtain the assertion (3.15). The proof is complete. ∎

Now we are ready to present an inequality where an upper bound of $\theta(\tilde{u}^k)-\theta(u)+(\tilde{w}^k-w)^T F(w)$ is found for all $w \in \Omega$. This inequality is also crucial for analyzing the contraction property and the convergence rate for the iterative sequence generated by either (1.4) or (1.5).

THEOREM 3.4. *For given $v^k \in \mathcal{Y} \times \Re^m$, let w^{k+1} be generated by the strictly contractive PRSM scheme* (1.5)*, and let $\tilde{w}^k$ be as defined in* (3.1)*; let M and H be as defined in* (2.4) *and* (2.7)*, respectively. Then, we have $\tilde{w}^k \in \Omega$ and*

$$\begin{aligned}
&\theta(u) - \theta(\tilde{u}^k) + (w-\tilde{w}^k)^T F(w) \\
&\geq \frac{1}{2}\big(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2\big) + \frac{1-\alpha}{4(1+\alpha)}\|v^k-v^{k+1}\|_H^2 \quad \forall w \in \Omega.
\end{aligned} \tag{3.17}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. First, because of the monotonicity of $F(w)$, we have

$$(w-\tilde{w})^T(F(w)-F(\tilde{w}))\geq 0 \quad \forall w,\ \tilde{w}\in\Omega.$$

Then, using the above inequality and replacing the right-hand side term in (3.6) with the inequality (3.15), we obtain the assertion (3.17). The proof is complete. ∎

The assertion (3.17) also enables us to study the contraction property of the sequence $\{v^k\}$ generated by (1.4) or (1.5). In fact, setting $w=w^*$ in (3.17) where w^* is an arbitrary solution point in Ω^*, we get

(3.18)
$$\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2\geq\frac{1-\alpha}{2(1+\alpha)}\|v^k-v^{k+1}\|_H^2+2\{\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^TF(w^*)\}.$$

Recall the optimality in (2.1). We thus have

$$\|v^{k+1}-v^*\|_H^2\leq\|v^k-v^*\|_H^2-\frac{1-\alpha}{2(1+\alpha)}\|v^k-v^{k+1}\|_H^2 \quad \forall v^*\in\mathcal{V}^*. \tag{3.19}$$

Therefore, when $\alpha=1$, i.e., for the PRSM scheme (1.4), we have

$$\|v^{k+1}-v^*\|_H^2\leq\|v^k-v^*\|_H^2, \tag{3.20}$$

which means the sequence $\{v^k\}$ generated by (1.4) is contractive, but not strictly, to the set $\mathcal{V}^*$. In fact, it is possible that the sequence $\{v^k\}$ stays away from the solution set with a constant distance (i.e., the equivalence in (3.20) holds for any k); hence no convergence of (1.4) is guaranteed under our assumption on (1.1). In [12] and [16], such an example was shown. On the other hand, when $\alpha\in(0,1)$, the inequality (3.19) ensures a reduction of $\frac{1-\alpha}{2(1+\alpha)}\|v^k-v^{k+1}\|_H^2$ to the set $\mathcal{V}^*$ at the $(k+1)$th iteration; i.e., the strict contraction of $\{v^k\}$ is guaranteed for the sequence generated by (1.5). Recall Lemma 3.2, which indicates that $\|v^k-v^{k+1}\|_H^2\neq 0$ whenever a solution is not yet found. Thus, the inequality (3.19) implies that the sequence $\{v^k\}$ generated by the proposed strictly contractive PRSM (1.5) converges to $\mathcal{V}^*$ with a guaranteed reduction of proximity to the solution set. As we have mentioned, the difference of contraction between (3.19) and (3.20) is also the reason we can establish a nonergodic convergence rate for the strictly contractive PRSM (1.5) in section 5 while only the ergodic convergence rate can be established for the original PRSM (1.4) in section 4.

4. Convergence rate of (1.4) in the ergodic sense. In this section, we show that although the original PRSM (1.4) might not be convergent to a solution point of the model (1.1), it is still possible to find an approximate solution of $\mathrm{VI}(\Omega,F,\theta)$ with an accuracy of $O(1/t)$ based on the first t iterations of the PRSM scheme (1.4). This estimate helps us better understand the convergence property of the original PRSM (1.4).

THEOREM 4.1. *Let $\{w^k\}$ be generated by PRSM* (1.4) *and $\{\tilde{w}^k\}$ be defined by* (3.1). *Let $\tilde{w}_t$ be defined as*

$$\tilde{w}_t=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{4.1}$$

Then, for any integer number $t>0$, $\tilde{w}_t\in\Omega$ and

$$\theta(\tilde{u}_t)-\theta(u)+(\tilde{w}_t-w)^TF(w)\leq\frac{1}{2(t+1)}\|v-v^0\|_H^2 \quad \forall w\in\Omega, \tag{4.2}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where H is as defined in (2.7).

Proof. First, because of (3.1), it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (4.1) implies that $\tilde{w}_t \in \Omega$. Second, by taking $\alpha = 1$ in (3.17) we have

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_H^2 \geq \frac{1}{2}\|v - v^{k+1}\|_H^2 \quad \forall w \in \Omega. \tag{4.3}$$

Summing the inequality (4.3) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\Big)^T F(w) + \frac{1}{2}\|v - v^0\|_H^2 \geq 0 \quad \forall w \in \Omega.$$

Using the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)}\|v - v^0\|_H^2 \quad \forall w \in \Omega. \tag{4.4}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \leq \frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k).$$

Substituting this into (4.4), the assertion of this theorem follows directly. ∎

Let $v^0 = (y^0, \lambda^0)$ be the initial iterate. For a given compact set $\mathcal{D} \subset \mathcal{Y} \times \Re^m$, let $d = \sup\{\|v - v^0\|_H \,|\, v \in \mathcal{D}\}$. Then, after t iterations of the PRSM (1.4), the point $\tilde{w}_t \in \Omega$ defined in (4.1) satisfies

$$\sup_{w \in \mathcal{D}} \big\{\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w)\big\} \leq \frac{d^2}{2(t+1)},$$

which means $\tilde{w}_t$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with an accuracy of $O(1/t)$ (recall (2.3)).

Remark 4.1. In the proof of Theorem 4.1, we take $\alpha = 1$ in (4.3). Obviously, the proof is still valid if we take $\alpha \in (0, 1)$. Thus, a worst-case $O(1/t)$ convergence rate in the ergodic sense can be established easily for the strictly contractive PRSM (1.5). As we shall show in section 5, this is less interesting because a nonergodic worst-case $O(1/t)$ convergence rate can be established for (1.5). We thus omit the details.

5. Convergence rate of (1.5) in a nonergodic sense. In this section, we show that the sequence $\{v^k\}$ generated by the strictly contractive PRSM scheme (1.5) is convergent to a point in $\mathcal{V}^*$, and its worst-case convergence rate is $O(1/t)$ in a nonergodic sense. Our starting point for the analysis is the inequality (3.19), and a crucial property is the monotonicity of the sequence $\{\|v^k - v^{k+1}\|_H^2\}$. That is, we will prove that

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2 \quad \forall\, k \geq 0.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We first take a closer look at the assertion (3.6) in Lemma 3.1.

LEMMA 5.1. *Let $\{w^k\}$ be the sequence generated by the strictly contractive PRSM* (1.5)*, let $\tilde{w}^k$ be as defined in* (3.1)*, and let the matrix Q be as defined in* (2.6)*. Then, we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0. \tag{5.1}$$

Proof. Setting $w = \tilde{w}^{k+1}$ in (3.6), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{5.2}$$

Note that (3.6) is also true for $k := k+1$, and thus

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (v - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}) \quad \forall\, w \in \Omega.$$

Setting $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{5.3}$$

Adding (5.2) and (5.3) and using the monotonicity of F, we get (5.1) immediately. ∎

LEMMA 5.2. *Let $\{w^k\}$ be the sequence generated by the strictly contractive PRSM* (1.5)*; let the matrices M, Q, and H be as defined in* (2.4)*,* (2.6)*, and* (2.7)*, respectively. Then, we have*

$$(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \geq \frac{3+\alpha}{4(1+\alpha)} \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2. \tag{5.4}$$

Proof. Adding the equation

$$\begin{aligned} &\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ &\quad = \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 \end{aligned}$$

to both sides of (5.1), we get

$$(v^k - v^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2. \tag{5.5}$$

By using $Q = HM$ and $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (see (2.8) and (3.5)) in the left-hand side of (5.5), we obtain

$$(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2. \tag{5.6}$$

Due to (2.9) we have

$$Q^T + Q \succeq \frac{(3+\alpha)}{2(1+\alpha)} M^T H M.$$

Substituting this into the right-hand side of (5.6) and using $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ again, we obtain (5.4), and the lemma is proved. ∎

Now, we are ready to prove the monotonicity of the sequence $\{\|v^k - v^{k+1}\|_H^2\}$.

THEOREM 5.3. *Let $\{w^k\}$ be the sequence generated by the strictly contractive PRSM* (1.5)*, let $\tilde{w}^k$ be as defined in* (3.1)*, and let the matrix H be as defined in* (2.7)*. Then, we have*

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2 - \frac{1-\alpha}{2(1+\alpha)}\|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2. \tag{5.7}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. Setting $a = (v^k - v^{k+1})$ and $b = (v^{k+1} - v^{k+2})$ in the identity

$$\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2,$$

we obtain

$$\begin{aligned}&\|v^k - v^{k+1}\|_H^2 - \|v^{k+1} - v^{k+2}\|_H^2 \\ &= 2(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2.\end{aligned}$$

Inserting (5.4) into the first term of the right-hand side of the last equality, we obtain

$$\|v^k - v^{k+1}\|_H^2 - \|v^{k+1} - v^{k+2}\|_H^2 \geq \frac{1-\alpha}{2(1+\alpha)}\|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2.$$

The assertion (5.7) follows from the above inequality directly, and the proof is complete. ∎

Now, we can establish a worst-case $O(1/t)$ convergence rate in a nonergodic sense for the strictly contractive PRSM scheme (1.5).

THEOREM 5.4. *Let $\{w^t\}$ be the sequence generated by the strictly contractive PRSM scheme* (1.5). *For any $v^* \in \mathcal{V}^*$, we have*

$$\|v^t - v^{t+1}\|_H^2 \leq \frac{2(1+\alpha)}{(t+1)(1-\alpha)}\|v^0 - v^*\|_H^2. \tag{5.8}$$

Proof. First, it follows from (3.19) that

$$\frac{1-\alpha}{2(1+\alpha)}\sum_{k=0}^{\infty}\|v^k - v^{k+1}\|_H^2 \leq \|v^0 - v^*\|_H^2 \quad \forall\, v^* \in \mathcal{V}^*. \tag{5.9}$$

According to Theorem 5.3, the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ is monotonically nonincreasing. Therefore, we have

$$(t+1)\|v^t - v^{t+1}\|_H^2 \leq \sum_{k=0}^{t}\|v^k - v^{k+1}\|_H^2. \tag{5.10}$$

The assertion (5.8) follows from (5.9) and (5.10) immediately. The proof is complete. ∎

Notice that $\mathcal{V}^*$ is convex and closed. Let $v^0 = (y^0, \lambda^0)$ be the initial iterate and $d := \inf\{\|v^0 - v^*\|_H \,|\, v^* \in \mathcal{V}^*\}$. Then, for any given $\epsilon > 0$, Theorem 5.4 shows that the strictly contractive PRSM scheme (1.5) needs at most $k = \lfloor \frac{2(1+\alpha)d^2}{(1-\alpha)\epsilon} \rfloor$ iterations to ensure that $\|v^k - v^{k+1}\|_H^2 \leq \epsilon$. It follows from Lemma 3.2 that w^{k+1} is a solution of $\mathrm{VI}(\Omega, F, \theta)$ if $\|v^k - v^{k+1}\|_H^2 = 0$. A worst-case $O(1/t)$ convergence rate in a nonergodic sense for the strictly contractive PRSM scheme (1.5) is thus established in Theorem 5.4.

6. Numerical results. In this section, we verify the theoretical assertions analyzed in previous sections by some numerical experiments. We focus on some applications of the abstract model (1.1) in statistical learning and image processing, including the least absolute shrinkage and selection operator (LASSO) model, the group LASSO model, the sparse logistic regression model, the image deblurring problem, the image inpainting problem, and the magnetic resonance imaging (MRI) problem. We shall verify the following assertions.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(1) The original PRSM (1.4) is indeed fast if it is convergent; the assertions in [20, 23] are thus further backed up.
(2) For some scenarios of (1.1), the original PRSM (1.4) might fail to converge while the proposed strictly contractive PRSM (1.5) is convergent; the theoretical significance of the underdetermined relaxation factor α is thus verified.
(3) For the proposed strictly contractive PRSM (1.5), the underdetermined relaxation factor α can be easily determined. In fact, empirically, $\alpha \in [0.8, 0.9]$ is preferred for all tested cases.
(4) The proposed strictly contractive PRSM (1.5) converges quickly for a wide range of applications, and it numerically outperforms some efficient methods in the literature.

We shall use the previously mentioned statistical leaning models to illustrate the first three assertions, and the last assertion will be verified by the previously mentioned imaging models. Our code was written by MATLAB 2012a, and all the numerical experiments were conducted on a laptop computer with a 2.9GHz i7 processor and an 8GB memory.

6.1. Statistical learning problems. In this subsection, we test some popular sparse learning models in the area of statistical learning. As we have mentioned, via these models our purpose is to justify the advantages of the proposed strictly contractive PRSM (1.5) itself. We thus choose the ADMM (1.3) as the only benchmark for numerical comparison in this subsection. Note that the ADMM (1.3) has been shown to be a widely applicable efficient solver for some popular statistical learning problems; see, e.g., [6] for a review.

6.1.1. Models and iterative schemes. We first introduce the sparse learning models to be studied.

(1) The LASSO model proposed in [51],

$$\min\left\{\frac{1}{2}\|Dx - r\|_2^2 + \gamma\|x\|_1 \mid x \in \Re^d\right\}, \tag{6.1}$$

where $r \in \Re^n$ is the response vector, $D \in \Re^{n\times d}$ is the design matrix, n is the number of data points, d is the number of features, $\gamma > 0$ is a regularization parameter, and $\|x\|_1 := \sum_{i=1}^n |x_i|$. The LASSO model provides a sparse estimation of x when there are more features than data points (i.e., $d > n$), and it has been very influential in several areas (e.g., bioinformatics [53, 40, 41], econometrics [7], and climate analysis [10]).

(2) The group LASSO model proposed in [55],

$$\min\left\{\frac{1}{2}\|Dx - r\|_2^2 + \gamma\sum_{i=1}^N \|x_i\|_2 \mid x \in \Re^d,\ x_i \in \Re^{d_i}\right\}, \tag{6.2}$$

where $\sum_{i=1}^N d_i = d$, N represents the number of disjointed groups partitioned among the variable x, and all other settings are the same as in (6.1). The model (6.2) promotes selecting the grouped variables (factors), which are common in multifactor analysis-of-variance (ANOVA) problems or additive models with polynomial or nonparametric components. In these problems, important factors are groups of variables rather than the individual derived variables. Group LASSO reduces to LASSO when $d_i = 1$, which means each group contains only one variable.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(3) The sparse logistic regression model in [45, 34, 35],

$$\min\left\{\frac{1}{n}\sum_{i=1}^{n}\log\Big(1+\exp\big(-r_i(D_i^T x+x_0)\big)\Big)+\gamma\|x\|_1,\ \mid x\in\Re^d, x_0\in\Re\right\}, \tag{6.3}$$

where $x \in \Re^d$ is the coefficient vector, $x_0 \in \Re$ is the intercept scalar, $D_i \in \Re^d$ $(i = 1, 2, \ldots, n)$ are training data points, $r_i \in \{-1, 1\}$ $(i = 1, 2, \ldots, n)$ are corresponding labels, n is the number of data points, d is the dimension of data, and $\gamma > 0$ is a regularization parameter. After obtaining the coefficient x and x_0 in (6.3), we can predict a binomial categorical label r_{pred} with the given input $D_{\text{in}} \in \Re^d$ by

$$r_{\text{pred}} = \text{sign}\left(\frac{1}{1+e^{-(x_0+D_{\text{in}}^T x)}}\right).$$

Now we illustrate how to implement the ADMM (1.3), the PRSM (1.4), and the strictly contractive PRSM (1.5) for solving these statistical learning models. First, for the LASSO model (6.1), by introducing an auxiliary variable $y \in \Re^d$, the model (6.1) can be reformulated as

$$\min\left\{\frac{1}{2}\|Dx-r\|_2^2+\gamma\|y\|_1 \mid x-y=0, x\in\Re^d,\ y\in\Re^d\right\}, \tag{6.4}$$

which is a special case of (1.1) where $f(x) = \frac{1}{2}\|Dx - r\|_2^2$, $g(y) = \gamma\|y\|_1$, $n_1 = n_2 = d$, $A = I_d$, $B = -I_d$, $b = 0$, $\mathcal{X} = \Re^d$, and $\mathcal{Y} = \Re^d$. Therefore, the iterative scheme of the ADMM (1.3) for (6.4) is

$$\left\{\begin{array}{l} x^{k+1} = (D^TD+\beta I)^{-1}(D^Tr+\beta y^k+\lambda^k),\\ y^{k+1} = S_{\gamma/\beta}(x^{k+1}-\lambda^k/\beta),\\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}}-\beta(x^{k+1}-y^{k+1}).\end{array}\right. \tag{6.5}$$

Moreover, the iterative schemes of the PRSM (1.4) and the strictly contractive PRSM (1.5) for (6.4) read, respectively, as

$$\left\{\begin{array}{l} x^{k+1} = (D^TD+\beta I)^{-1}(D^Tr+\beta y^k+\lambda^k),\\ \lambda^{k+\frac{1}{2}} = \lambda^k-\beta(x^{k+1}-y^k),\\ y^{k+1} = S_{\gamma/\beta}(x^{k+1}-\lambda^{k+\frac{1}{2}}/\beta),\\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}}-\beta(x^{k+1}-y^{k+1})\end{array}\right. \tag{6.6}$$

and

$$\left\{\begin{array}{l} x^{k+1} = (D^TD+\beta I)^{-1}(D^Tr+\beta y^k+\lambda^k),\\ \lambda^{k+\frac{1}{2}} = \lambda^k-\alpha\beta(x^{k+1}-y^k),\\ y^{k+1} = S_{\gamma/\beta}(x^{k+1}-\lambda^{k+\frac{1}{2}}/\beta),\\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}}-\alpha\beta(x^{k+1}-y^{k+1}).\end{array}\right. \tag{6.7}$$

In (6.5)–(6.7), $S_\kappa(a)$ is the soft-thresholding operator defined as

$$(S_\kappa(a))_i = (1-\kappa/|a_i|)_+\cdot a_i, \quad i=1,\ldots,d,$$

for $\kappa > 0$ and $a \in \Re^d$; see, e.g., [14].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Similarly, by introducing $y_i = x_i$ for $i = 1, 2, \ldots, N$, we can reformulate the group LASSO model (6.2) as

(6.8)
$$\min\left\{\frac{1}{2}\|Dx - r\|_2^2 + \gamma\sum_{i=1}^{N}\|y_i\|_2 \mid x_i - y_i = 0,\ i = 1, 2, \ldots, N;\ x_i \in \Re^{d_i},\ y_i \in \Re^{d_i}\right\},$$

for which the ADMM (1.3), the PRSM (1.4), and the strictly contractive PRSM (1.5) are implementable. The resulting iterative schemes are, respectively, as follows:

$$\left\{\begin{array}{l} x^{k+1} = (D^TD + \beta I)^{-1}(D^Tr + \beta y^k + \lambda^k), \\ y_i^{k+1} = S_{\gamma/\beta}(x_i^{k+1} - \beta\lambda_i^k), \quad i = 1, \ldots, N, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(x^{k+1} - y^{k+1}), \end{array}\right. \tag{6.9}$$

$$\left\{\begin{array}{l} x^{k+1} = (D^TD + \beta I)^{-1}(D^Tr + \beta y^k + \lambda^k), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(x^{k+1} - y^k), \\ y_i^{k+1} = S_{\gamma/\beta}(x_i^{k+1} - \beta\lambda_i^{k+\frac{1}{2}}), \quad i = 1, \ldots, N, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(x^{k+1} - y^{k+1}), \end{array}\right. \tag{6.10}$$

and

$$\left\{\begin{array}{l} x^{k+1} = (D^TD + \beta I)^{-1}(D^Tr + \beta y^k + \lambda^k), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1} - y^k), \\ y_i^{k+1} = S_{\gamma/\beta}(x_i^{k+1} - \beta\lambda_i^{k+\frac{1}{2}}), \quad i = 1, \ldots, N, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(x^{k+1} - y^{k+1}). \end{array}\right. \tag{6.11}$$

Finally, the sparse logistic regression model (6.3) can be reformulated as

(6.12)
$$\min\left\{\frac{1}{n}\sum_{i=1}^{n}\log(1+\exp(-r_i(D_i^Tx+x_0))+\gamma\|y\|_1) \mid x-y=0,\ x_0 \in \Re, x \in \Re^d, y \in \Re^d\right\},$$

where $y \in \Re^d$ is an auxiliary variable. Therefore, the ADMM (1.3), the PRSM (1.4), and the strictly contractive PRSM (1.5) are all applicable to (6.12). For succinctness, we list only the iterative scheme of the strictly contractive PRSM (1.5) in detail:

(6.13)
$$\left\{\begin{array}{l} x^{k+1} = \arg\min_x \frac{1}{n}\sum_{i=1}^{n}\log(1+\exp(-r_i(D_i^Tx+x_0))) + \frac{\beta}{2}\|x - y^k - \lambda^k/\beta\|_2^2, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1} - y^k), \\ y_i^{k+1} = S_{\gamma/\beta}(x_i^{k+1} - \beta\lambda_i^{k+\frac{1}{2}}), \quad i = 1, \ldots, N, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(x^{k+1} - y^{k+1}). \end{array}\right.$$

Note that in (6.13) the x-subproblem has no closed-form solutions. We implement Newton's method (coded by MATLAB) to solve this subproblem with a tolerance of 10^{-5} and a maximum iteration number of 10.

6.1.2. Implementation details. Now we specify the setting for the sparse learning models to be tested. For the LASSO model (6.1), we set $n = 2000$ and $d = 4000$; each entry of D is drawn from $\mathcal{N}(0, 1)$, and then all columns of D are normalized; we generate a random sparse vector in $\Re^{4000}$ with 100 nonzero entries from $\mathcal{N}(0, 1)$ as x; the noise vector $\epsilon \sim \mathcal{N}(0, 10^{-3}I)$ and the vector $r = Dx + \epsilon$; the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

regularization parameter is set as $\gamma = 0.1\|D^T r\|_\infty$. For the group LASSO model (6.2), we set $n = 1500$; D is also drawn from $\mathcal{N}(0,1)$ and normalized; the noise vector $\epsilon \sim \mathcal{N}(0, 10^{-3}I)$ and the vector $r = Dx + \epsilon$; we generate $N = 200$ blocks with size n_i uniformly distributed between 1 and 50, $d = \sum_{i=1}^N n_i$. Among all the blocks, 5% have entries drawn from the standard Gaussian distribution, and the other blocks have entries all equal to zero; the regularization parameter γ is set as $\gamma = 0.1 \max\{\|D_1^T r\|_2, \ldots, \|D_N^T r\|_2\}$, where $\{D_1, D_2, \ldots, D_N\}$ with $D_i \in \Re^{n\times d_i}$ is a disjointed partition of D's columns in correspondence with the partition of $\{x_1, x_2, \ldots, x_N\}$. For the sparse logistic regression model (6.3), we set $n = 50$ and $d = 500$; each vector $D_i \in \Re^{500}$ has 10 nonzero entries from $\mathcal{N}(0,1)$; the vectors r_i are generated by $r_i = \text{sign}(D_i^T x + x_0 + \epsilon_i)$, where ϵ_i is the noise drawn from $\mathcal{N}(0, 0.1)$; the vector $x \in \Re^d$ contains 100 nonzero entries drawn from $\mathcal{N}(0,1)$; the intercept x_0 is also from $\mathcal{N}(0,1)$, and γ is set according to [34]: $\gamma = 0.1\gamma_{\max}$, where $\gamma_{\max}$ is the maximum regularization parameter above which the solution x has all zero entries.

Then, we provide some details to implement the methods to be tested. Our code was constructed based on the code available at http://www.stanford.edu/∼boyd/papers/admm/. Hence, the ADMM (1.3) can be implemented directly by this code package, while the implementation of the PRSM (1.4) and the strictly contractive PRSM (1.5) require only a slight modification on the Lagrange multiplier to this package. It is worth mentioning that in (6.5)−(6.11), we need to compute $(D^T D + \beta I)^{-1}$ and $D^T r$, which is quite time consuming if N and d are large. However, since these two terms are invariant in each iteration, we need only compute it once before all iterations. We define the stopping criterion as

$$\max\left\{\beta\|y^k - y^{k+1}\|_2, \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|_2\right\} \le \sqrt{d}\epsilon, \tag{6.14}$$

where $\beta\|y^k - y^{k+1}\|_2$ and $\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|_2$ measure the primal and dual residuals, respectively, and $\epsilon > 0$ is a tolerance; see, e.g., [6, 30, 56]. Note that because of Lemma 3.2 it is also reasonable to use this stopping criterion for the PRSM (1.4) and the strictly contractive PRSM (1.5). For LASSO (6.1) and group LASSO (6.2), we set $\epsilon = 10^{-4}$; and for the sparse logistic regression model (6.3), we set $\epsilon = 10^{-3}$ since its x-subproblem is solved approximately at each iteration. For the penalty parameter β, we set it as 1 for all methods; and we set $\alpha = 0.9$ for the strictly contractive PRSM (1.4) (the reason will be explained later).

6.1.3. Results. In Table 1, we report the computing time in seconds ("time (s)") and the number of iterations when the ADMM (1.3), the PRSM (1.4), and the strictly contractive PRSM (1.5) are applied to solve the above-mentioned statistical learning models. According to this table, we see that the original PRSM (1.4) is convergent only for the group LASSO model (6.2). For this case, the original PRSM (1.4) is really faster than the ADMM (1.3), and it is almost as efficient as the strictly contractive PRSM (1.4). The assertions in [20, 23] are thus verified again. For the other two models, however, the convergence of PRSM is not witnessed. But the proposed strictly contractive PRSM (1.4) still performs well—faster than ADMM.

To further observe the convergence of the ADMM and the strictly contractive PRSM, in Figures 1–3 we visualize the evolution of convergence when these two methods are applied to solve these three sparse learning models. The evolution of the objective function value, the reduction of primal and dual residuals, and $\|v^{(k)} - v^*\|_H$ with respect to the iterations are plotted. Plots in Figure 1 show that the strictly contractive PRSM and ADMM reach the primal tolerance at almost the same time,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

but the former reaches the dual tolerance faster than the latter. Plots in Figure 2 indicate that the strictly contractive PRSM reaches both the primal and dual tolerances faster than ADMM. Last, plots in Figure 3 reveal that ADMM reaches the primal tolerance faster (but possibly jumps back to above the tolerance again), while the strictly contractive PRSM reaches the dual tolerance much faster than ADMM. The lower right plots in Figure 1–3 illustrate that the sequence $\|v^{(k)} - v^*\|_H$ decreases monotonically, which further back up our theoretical results in (3.19).

TABLE 1
Quantitative comparison among the strictly contractive PRSM, the original PRSM, and the ADMM for statistical learning models.

	LASSO		Group LASSO		Sparse logistic regression	
Algorithm	time(s)	# iterations	time(s)	# iterations	time(s)	# iterations
Strictly contractive PRSM	**1.27**	**53**	**0.53**	**21**	**2.14**	**41**
ADMM	2.13	89	0.96	36	3.75	72
PRSM	-	-	0.53	23	-	-

"-" means that the stopping criterion $\max\{\beta\|y^k - y^{k+1}\|_2, \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|_2\} \le \sqrt{d}\epsilon$ is not satisfied after $10,000$ iterations.

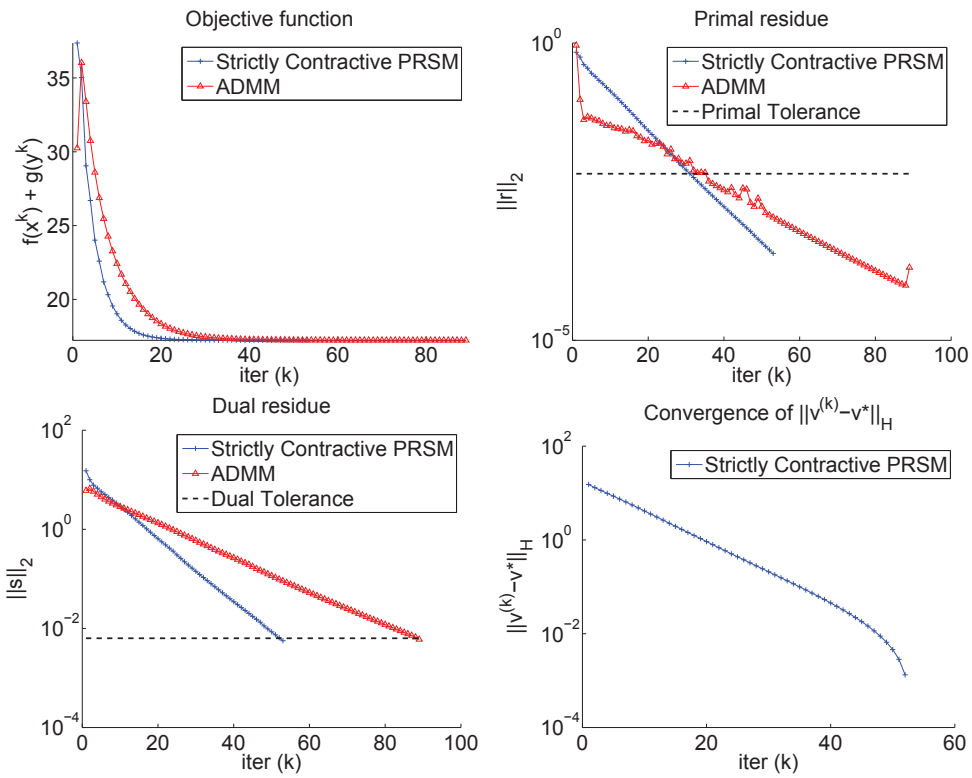


FIG. 1. *LASSO model* (6.1): *Evolution of the objective function value, primal residual, and dual residual, and* $\|v^{(k)} - v^*\|_H$ *for ADMM and the strictly contractive PRSM.*

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

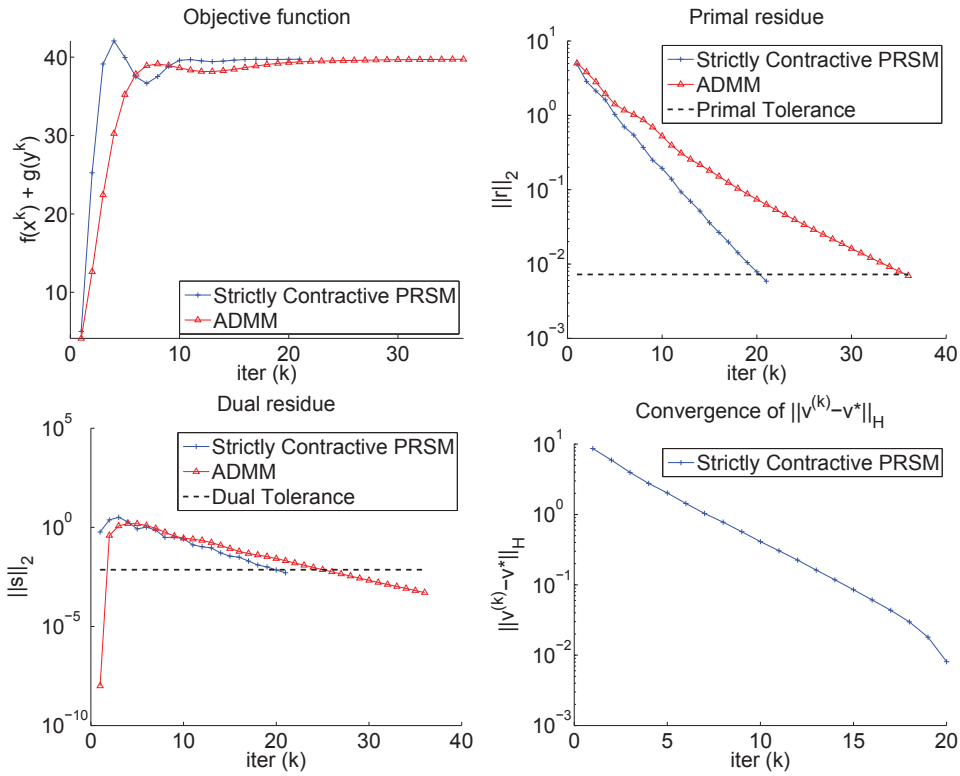


FIG. 2. *Group LASSO model* (6.2): *Evolution of the objective function, primal residual, and dual residual, and* $\|v^{(k)} - v^*\|_H$ *for ADMM and the strictly contractive PRSM.*

6.1.4. Sensitivity to α. As we have analyzed, attaching an underdetermined relaxation factor $\alpha \in (0,1)$ to the original PRSM (1.4) can make the resulting iterative sequence strictly contractive with respect to the solution set of (1.1). Thus, it becomes possible to ensure the convergence and establish a worst-case $O(1/t)$ convergence rate in a nonergodic sense for the strictly contractive PRSM (1.5). Despite its significant theoretical role, we would emphasize that this underdetermined relaxation factor can be chosen easily to implement the strictly contractive PRSM (1.5). This is an important convenience for the implementation of the strictly contractive PRSM (1.5).

In this subsection, we take the LASSO model (6.1) to test the sensitivity of α for the strictly contractive PRSM (1.5). We fix $\beta = 1$ and choose different values of α in the interval $[0.05, 0.99]$. (More specifically, we choose $\alpha = \{0.05, 0.10, 0.15, \ldots, 0.85, 0.90, 0.91, 0.92, \ldots, 0.98, 0.99\}$.) The computing time and number of iterations required by the strictly contractive PRSM (1.5) are recorded for each choice of α. Then, we plot them in Figure 4. For comparison purposes, we also plot for ADMM with $\beta = 1$. According to the curves in Figure 4, we see that the underdetermined relaxation factor α works for a wide range of values; thus it can be chosen easily in implementation. In particular, based on our experiments, some aggressive values close to 1 (e.g., $[0.8, 0.9]$) are preferred.

As is well known in the literature, the numerical performance of some augmented-Lagrangian-based methods including the ADMM (1.3) is highly dependent on the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

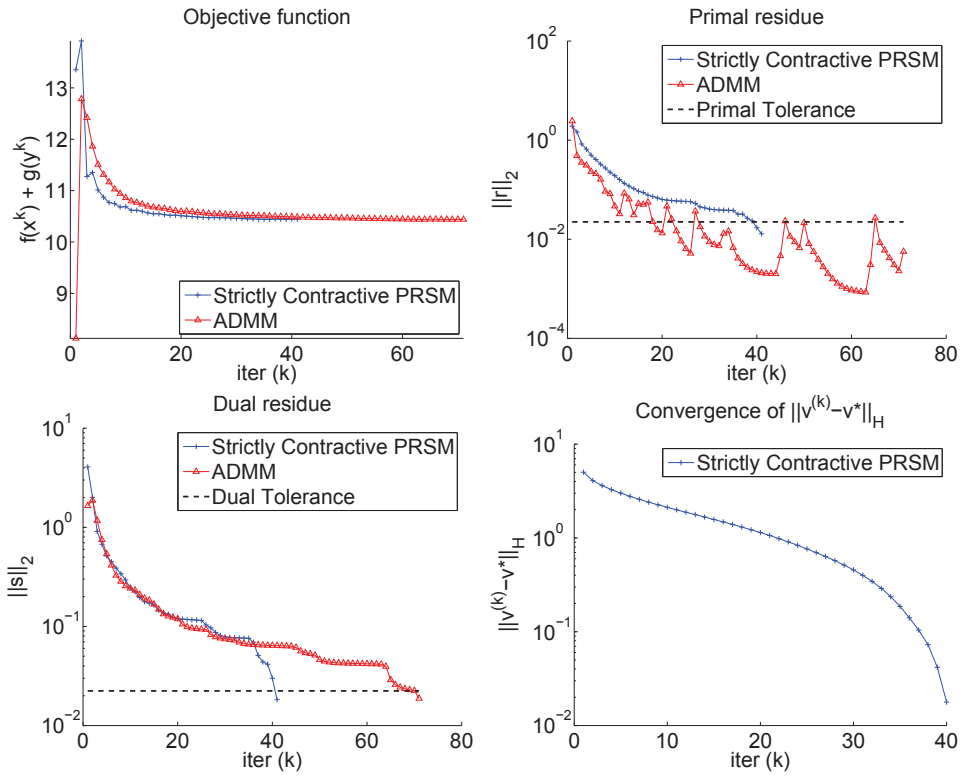
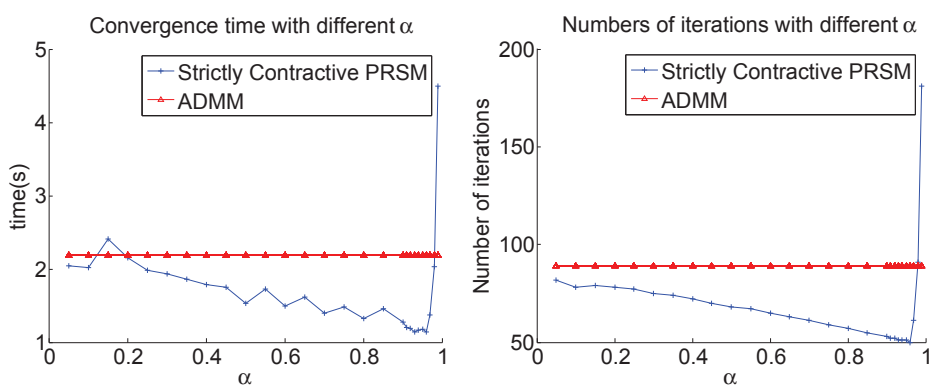


FIG. 3. *Sparse logistic regression model* (6.3): *Evolution of the objective function, primal residual, and dual residual, and* $\|v^{(k)} - v^*\|_H$ *for ADMM and the strictly contractive PRSM.*



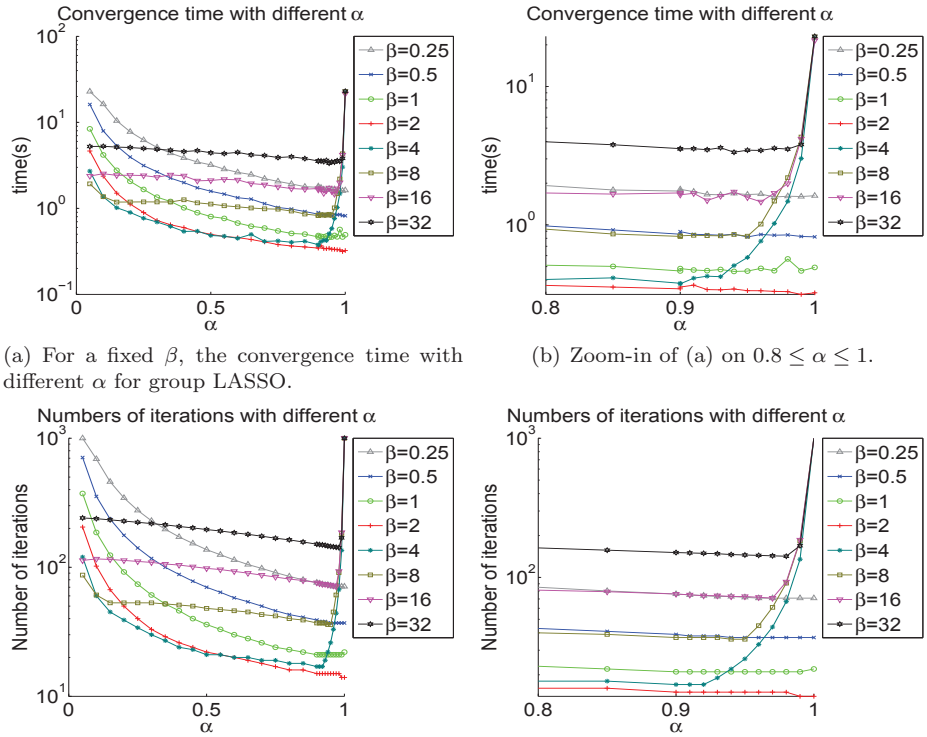
(a) α's effect on the execution time. (b) α's effect on the number of iterations.

FIG. 4. *Sensitivity test on the underdetermined relaxation factor* α *when* $\beta = 1$.

penalty parameter β. Theoretically, some strategies of adjusting this parameter automatically have been proposed; see, e.g., [29]. But for some concrete applications (especially some large-scale problems or models with matrix variables), realizing this kind of self-adaptive strategy might result in too much computation. Thus, a more popular way to choose this parameter is to tune manually and then fix it as a tuned

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php



(a) For a fixed β, the convergence time with different α for group LASSO.

(b) Zoom-in of (a) on $0.8 \leq \alpha \leq 1$.

(c) For a fixed β, the number of iterations with different α for group LASSO.

(d) Zoom-in of (c) on $0.8 \leq \alpha \leq 1$.

FIG. 5. *Acceleration of α for the group LASSO model* (6.2).

value throughout. To the best of our knowledge, it is not clear so far how to determine an optimal value for β; it is highly possible that it is problem-dependent. This difficultly occurs also for the original PRSM (1.4). In this subsection, we test some fixed values of β and empirically verify that when implementing the strictly contractive PRSM (1.5), it is easy to choose $\alpha \in (0, 1)$ to accelerate the convergence.

We take the group LASSO model (6.2) to demonstrate the effectiveness of α for a fixed β. In our experiments, we test a set of value $\beta = 0.25, 0.5, 1, 2, 4, 8, 16, 32$. For each β, we choose different values of $\alpha = \{0.05, 0.10, 0.15, \ldots, 0.85, 0.90, 0.91, 0.92, \ldots, 0.98, 0.99\}$ and plot the computing time in seconds and number of iterations with respect to different choices of α in Figure 5. According to the plots in Figure 5, it seems that the original PRSM is very sensitive to the value of β. In fact, for some β such as $\beta = 4, 8, 16, 32$, the original PRSM (1.4) (i.e., $\alpha = 1$) fails to satisfy the stopping criterion within $10,000$ iterations. This further emphasizes the importance of choosing β when implementing the original PRSM (1.4). At the same time, we see that for each of the tested β, $\alpha \in (0, 0.9)$ tends to accelerate the convergence of PRSM.

6.2. Image reconstruction models. In this subsection, we test some digital image reconstruction models. Our aim is to further verify the efficiency of the proposed strictly contractive PRSM (1.5) by comparing it numerically with four well-known algorithms in the imaging literature: SALSA [1], TwIST [4], SpaRSA [52], FISTA [2], and YALL1/TVAL3 [36, 37, 38, 54, 57].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

6.2.1. Models and iterative schemes. We first briefly review the background of digital image reconstruction problems; for more details we refer the reader to [28, 47]. A fundamental task in many areas such as medical and astronomical imaging, film restoration, image or video coding, and synthetic aperture radar (see, e.g., [49, 50, 33]), the image reconstruction problem is to reconstruct the original image $p \in \Re^n$ from its degraded image $p^0 \in \Re^n$. Note that we vectorize an $N \times M$-pixel image P into an n-dimensional vector p in lexicographical order with $n = NM$. The relationship between p and p^0 is given by

$$p^0 = Dp + \epsilon, \tag{6.15}$$

where $\epsilon \in \Re^n$ is a noise corrupting the original image p, and $D \in \Re^{n\times n}$ is the matrix representation of a distortion operator such as a blurring (convolution), vignetting, inpainting, or zooming operator.

According to [13], we can classify image reconstruction models into two categories: the synthesis approach and the analysis approach. The synthesis approach defines $x \in \Re^d$ as the vector of wavelet coefficients of the original image p under a wavelet dictionary. Let $W \in \Re^{n\times d}$ be the matrix of a wavelet dictionary, e.g., a group of orthogonal bases; we then have $p = Wx$. Since (6.15) is usually ill-posed, certain regularization techniques are required. Note that the image p processes a sparse representation under the wavelet dictionary W; that is, x is sparse with many zero entries. Therefore, it is natural to use $\|x\|_1$, the l_1 norm of x, to regularize the data-fidelity term. We thus have

$$\min\left\{\frac{1}{2}\|DWx - p^0\|_2^2 + \gamma\|x\|_1 \mid x \in \Re^d\right\} \tag{6.16}$$

as the synthesis approach of an image reconstruction model. Note that we consider only the case of additive nose. Thus the l_2 norm is used for the data-fidelity term in (6.16). Other cases such as the impulsive or uniform noise can also be considered. On the other hand, the analysis approach considers reconstructing the image directly and not under a wavelet domain. Let the image be represented by a vector $x \in \Re^n$. Under the consideration of additive noise in (6.15), the data-fidelity term is $\frac{1}{2}\|Dx - p^0\|_2^2$. For the regularization term, a very popular choice is the total variation (TV) regularization proposed in the seminal work [48], which is well known to be capable of preserving the edges of images. We thus have

$$\min\left\{\frac{1}{2}\|Dx - p^0\|_2^2 + \gamma\mathrm{TV}(x) \mid x \in \Re^n\right\} \tag{6.17}$$

as the analysis approach of an image reconstruction model. In (6.17), $\mathrm{TV}(x)$ denotes the nonsmooth isotropic TV norm [48]:

$$\mathrm{TV}(x) = \sum_{i=1}^{N-1}\sum_{j=1}^{M-1}\sqrt{\left(x_{i,j}-x_{i+1,j}\right)^2 + \left(x_{i,j}-x_{i,j+1}\right)^2} + \sum_{i=1}^{N-1}|x_{i,M}-x_{i+1,M}| + \sum_{j=1}^{M-1}|X_{N,j}-x_{N,j+1}|,$$

where $x \in \Re^n$ is the vectorized original $N \times M$ two-dimensional image in lexicographic order with $n = NM$ and $x_{i,j}$ denotes pixel value at the position (i, j).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Now we show how to implement the proposed strictly contractive PRSM (1.5) to solve the models (6.16) and (6.17). For (6.16), by introducing an auxiliary variable $y \in \Re^d$, it can be reformulated as

$$\min\left\{\frac{1}{2}\|DWx - p^0\|_2^2 + \gamma\|y\|_1 \mid x - y = 0,\ x \in \Re^d, y \in \Re^d\right\}. \tag{6.18}$$

Similarly, the model (6.17) can be reformulated as

$$\min\left\{\frac{1}{2}\|Dx - p^0\|_2^2 + \gamma\mathrm{TV}(y) \mid x - y = 0,\ x \in \Re^n,\ y \in \Re^n\right\}. \tag{6.19}$$

Therefore, when the strictly contractive PRSM (1.5) is applied to solve the models (6.18) and (6.19), the iterative schemes read, respectively, as

$$\left\{\begin{array}{l} x^{k+1} = ((DW)^T(DW) + \beta I)^{-1}((DW)^T p^0 + \beta y^k + \lambda^k), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1} - y^k), \\ y^{k+1} = S_{\gamma/\beta}(x^{k+1} - \lambda^{k+\frac{1}{2}}/\beta), \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(x^{k+1} - y^{k+1}) \end{array}\right. \tag{6.20}$$

and

$$\left\{\begin{array}{l} x^{k+1} = (D^T D + \beta I)^{-1}(D^T p^0 + \beta y^k + \lambda^k), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1} - y^k), \\ y^{k+1} = \arg\min\left\{\gamma\mathrm{TV}(y) + \frac{\beta}{2}\|x^{k+1} - \lambda^{k+\frac{1}{2}}/\beta + y\|^2 \;\middle|\; y \in \mathcal{Y}\right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(x^{k+1} - y^{k+1}). \end{array}\right. \tag{6.21}$$

Let us explain how to solve the subproblems in (6.20) and (6.21). For example, the x-subproblem in (6.20) might be computationally expensive due to the high dimensionality of x. For example, for the analysis approach, $x \in \Re^{262144}$ for a 512×512-pixel image. According to [1], the matrices D and W are of special structures (e.g., $W^T W = I$); fast solvers such as the fast Fourier transform (FFT) are thus applicable. For the y-subproblem in (6.21) whose closed-form solution does not exist, we adopt the algorithm proposed in [8] to solve it.

We will test three scenarios for the models (6.16) and (6.17):

(1) The synthesis-based (6.16) image deblurring model where D is the matrix representation of the blurring operator [27]. Here we use the 9×9 uniform convolution kernel with every element being $1/81$ as the blurring operator.
(2) The analysis-based (6.17) image inpainting model where D is the matrix representation of the missing pixel operator [27]. Specifically, D is a highly sparse matrix with only ones and zeros in the diagonal. The zeros in the diagonal correspond to the missing pixels.
(3) The analysis-based (6.17) MRI image reconstruction model where D is matrix representation of the 22-radial-line mask in the frequency domain, which is visualized in Figure 11.

6.2.2. Implementation details. Among the methods to be compared, SALSA in [1] and YALL1/TVAL3 in [38, 57] are ADMM-based algorithms (YALL1 is for l_1 norm regularized problems and TVAL3 is for TV norm regularized problems; thus we implement YALL1 to (6.16) and TVAL3 to (6.17)). The proposed strictly contractive PRSM (1.5) thus can be easily coded based on the source code of SALSA, which is publicly available. We followed the user guide [38, 57] to code YALL1/TVAL3 with

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

tuned parameters. Codes for all other methods are downloaded from the respective authors' web pages. We terminate the iteration of the strictly contractive PRSM when

$$|\text{Objective}_{k+1} - \text{Objective}_k| \leq 10^{-4},$$

where "Objective$_k$" represents the objective function value at the kth iterate for the model under consideration. To compare, for all other methods, their iterations are terminated when their objective functions are less than or equal to the objective function value obtained by the strictly contractive PRSM (1.5). In other words, we compare these methods subject to the criterion that they achieve the same objective function value.

We measure the noise of an image by the signal-to-noise ratio in units of dB,

$$\text{SNR} := 10 \log_{10}(\|p\|_2^2 / \|p^0 - p\|_2^2),$$

where p and p^0 are the clean image and the distorted image, respectively. The error of reconstruction is measured by the mean squared error (MSE)

$$\text{MSE} = \|p - \hat{p}\|_2^2 / n,$$

where $\hat{p}$ is the reconstructed image. We also define the improved signal-to-noise ratio as

$$\text{ISNR} := 10 \log_{10}(\|p^0 - p\|_2^2 / \|p - \hat{p}\|_2^2)$$

as a uniform measurement of the quality of reconstructed images. For different methods, we compare the speed in terms of computing time and number of iterations to achieve the same quality of reconstruction which is measured by MSE or ISNR.

For the synthesis image deblurring application, we set $\gamma = 0.0075$ and test the 256×256-pixel image of Lena. We choose the four-level redundant Haar wavelet frame as W [11]. To generate the convolution operator D, we choose a 9×9 uniform blur kernel in which every element equals $1/81$ (with zero padding in the boundary). For the corrupted image, its SNR value is 40dB. The Gaussian noise vector ϵ is thus generated by $\mathcal{N}(0, 0.449)$. We list the clean and blurred images in Figure 10. To implement the strictly contractive PRSM (1.5), we set $\beta = 0.0075$ and $\alpha = 0.8$.

For the analysis image inpainting application, we set $\gamma = 0.15$ and test the 256×256-pixel image of Lena. The inpainting operator D contains 40% missing pixels which are chosen randomly. The masked image is shown in Figure 10. For the corrupted image, its SNR value is also 40dB, which means the noise is generated from $\mathcal{N}(0, 0.529)$. To implement the strictly contractive PRSM (1.5), we set $\beta = 0.05$ and $\alpha = 0.9$. To solve the y-subproblem in (6.21) by the method in [8], we allow a maximum of 20 for the inner iteration.

For the analysis MRI application, we set $\gamma = 0.0001$ and test the 128×128-pixel image of the Shepp–Logan phantom. The clean image is masked by 22 radial lines on its discrete Fourier transform, and only the frequency components covered by the radial lines are observed. We contaminate the frequency components using the circular complex Gaussian noise with $\sigma_\epsilon^2 = 0.5 \times 10^{-3}$; i.e., the real and imaginary parts of the noise are independent Gaussian with the standard deviation σ_ϵ. We list the clean and masked images of the Shepp–Logan phantom in Figure 11. The SNR value is 5.42. To implement the strictly contractive PRSM (1.5), we set $\beta = 0.01$ and $\alpha = 0.9$. To solve the y-subproblem in (6.21) by the method in [8], we allow a maximum of 40 for the inner iteration.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

6.2.3. Results. We first list the comparison of different methods for the imaging construction models in Tables 2–4.[4] Then, we visualize the evolution of the objective function when these methods are applied to solve these imaging models in Figures 6–8. The evolution of the MSE is also plotted in Figure 9. In Figure 12, we display the clean, corrupted, and reconstructed images by the strictly contractive PRSM (1.5) for the tested scenarios. These tables and figures clearly show the efficiency of the proposed strictly contractive PRSM (1.5).

TABLE 2
Quantitative comparison on image deblurring model.

	Image deblurring			
Algorithm	time(s)	#iterations	MSE	ISNR(dB)
Strictly contractive PRSM	**2.60**	**15**	**97.8**	**6.60**
SALSA	6.86	41	97.7	6.60
TwIST	7.87	60	97.5	6.61
FISTA	8.18	97	103	6.63
SpaRSA	9.40	86	107	6.21
YALL1	9.38	86	107	6.21

TABLE 3
Quantitative comparison on image inpainting model.

	Image inpainting			
Algorithm	time(s)	#iterations	MSE	ISNR(dB)
Strictly contractive PRSM	**3.86**	**28**	**88.9**	**17.5**
SALSA	6.82	46	93.2	17.3
TwIST	14.4	83	89.7	17.4
FISTA	6.76	91	90.4	17.4
TVAL3	6.04	117	90.2	17.4

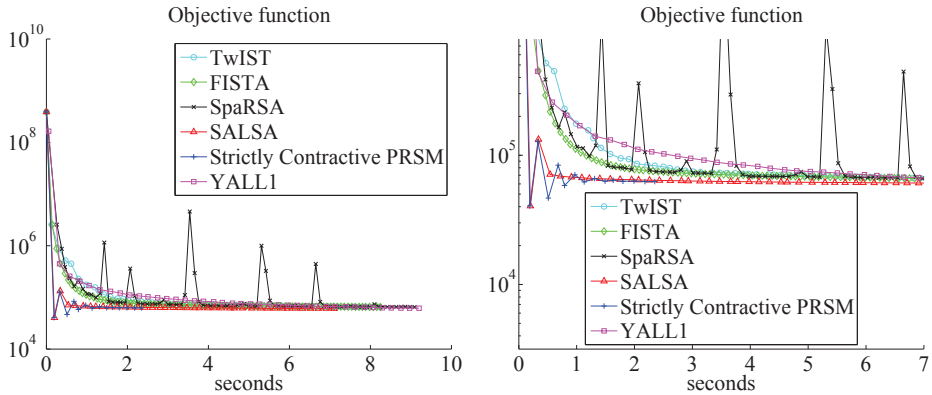
TABLE 4
Quantitative comparison on MRI model.

Algorithm	time(s)	#iterations	MSE	ISNR(dB)
Strictly contractive PRSM	**19.81**	**207**	**1.03652e-06**	**42.2073**
SALSA	30.46	311	1.03435e-06	42.2164
TwIST	53.69	451	1.10616e-06	41.9673
FISTA	25.23	460	1.04220e-06	42.1836
SpaRSA	97.74	1001	1.09543e-06	41.9250
TVAL3	45.36	601	1.03762e-06	42.2027

7. Conclusions. As a classical operator splitting method in the literature, the Peaceman–Rachford splitting method (PRSM) may fail to be convergent for solving a convex optimization problem with linear constraints and a separable objective function. This paper shows that this failure can be illustrated by showing that its iterative sequence is not strictly contractive with respect to the solution set of the model under consideration. This understanding from a contraction perspective inspires us to tackle the deficiency of PRSM by embedding an underdetermined factor

[4]In our experiments, the open source code of SpaRSA does not work for the image inpainting model under testing. Table 3 thus does not include any comparison with SpaRSA.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.



(a) Image deblurring.

(b) Image deblurring. (Zoom in on the first 7 seconds.)

FIG. 6. *Visualization of the objective functions on image deblurring model.*

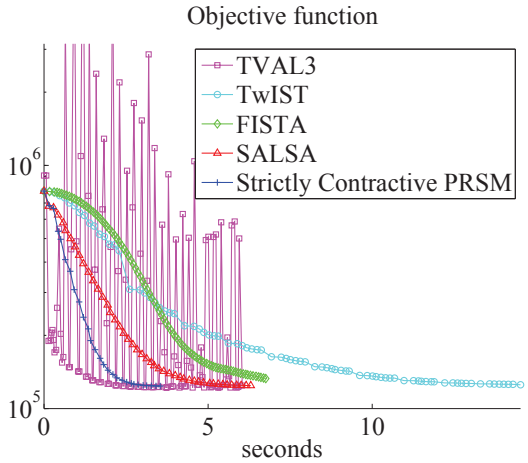
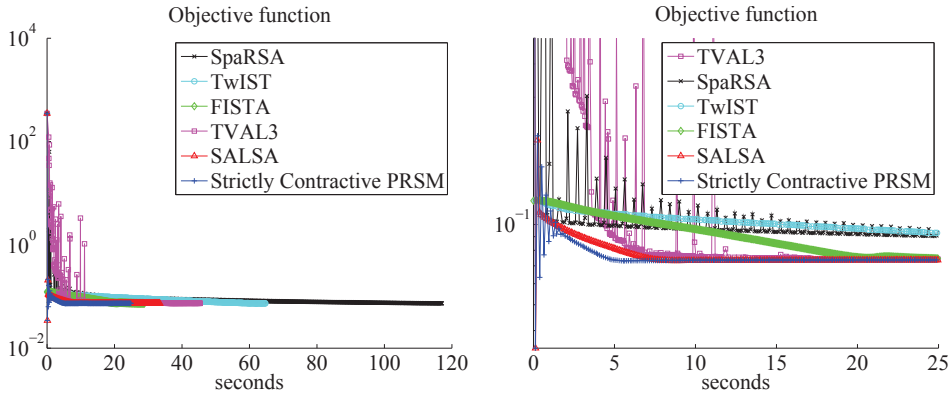


FIG. 7. *Visualization of the objective functions on image inpainting model.*



(a) The convergence of the objective functions.

(b) The convergence of the objective functions. (Zoom in on the first 25 seconds.)

FIG. 8. *Visualization of the objective functions on MRI image reconstruction.*

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

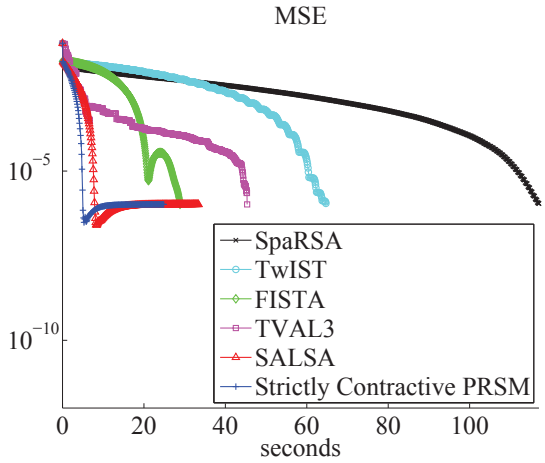


FIG. 9. *Visualization of MSE on MRI image reconstruction.*

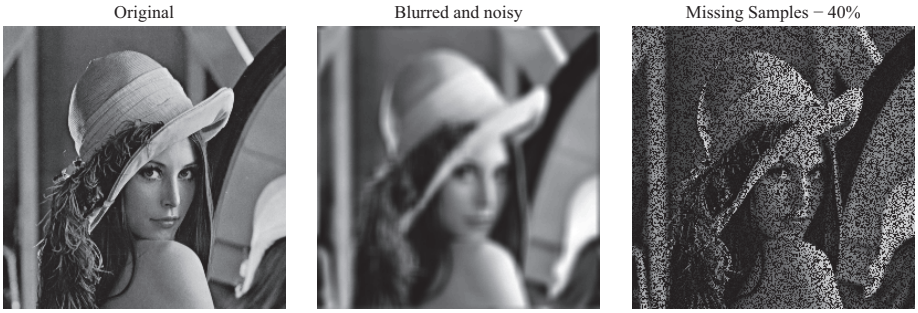


FIG. 10. *The original, blurred, and masked (with* 40% *missing pixels) images of Lena.*

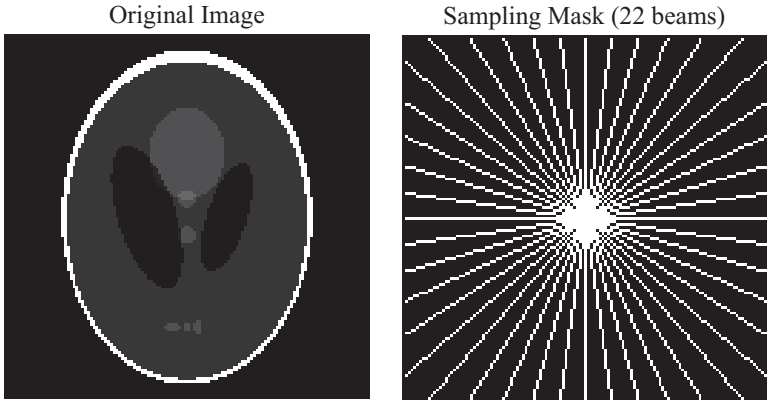


FIG. 11. *The original and masked images of the Shepp–Logan phantom.*

into the iterative scheme of PRSM and to propose a strictly contractive PRSM. The strictly contractive PRSM is as easy to implement as that of the alternating direction method of multipliers (ADMM), and it is numerically faster. We verify these advantages by some applications in statistical learning and image processing. We also

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

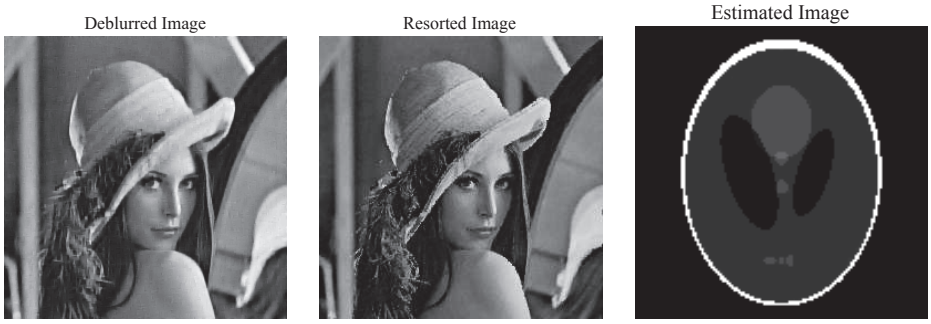


FIG. 12. *The deblurred and reconstructed images of Lena, and the reconstructed Shepp–Logan phantom image by strictly contractive PRSM. The SNRs for the reconstructed images are* 46.6*dB,* 57.5*dB, and* 47.6*dB, respectively.*

study the convergence rate of the proposed strictly contractive PRSM, establishing the worst-case $O(1/t)$ convergence rate in both the ergodic and nonergodic senses.

REFERENCES

[1] M.V. Afonso, J.M. Bioucas-Dias, and M.A.T. Figueiredo, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Imaging Process., 9 (2010), pp. 2245–2256.

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[3] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[4] J.M. Bioucas-Dias and M.A.T. Figueiredo, *A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration*, IEEE Trans. Imaging Process., 2 (2007), pp. 2992–3004.

[5] E. Blum and W. Oettli, *Mathematische Optimierung. Grundlagen und Verfahren*, Ökonometrie und Unternehmensforschung, Springer-Verlag, Berlin, Heidelberg, New York, 1975.

[6] 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 Machine Learning, 3 (2010), pp. 1–122.

[7] M. Caner, *Lasso-type GMM estimator*, Econometric Theory, 25 (2009), pp. 270–290.

[8] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis., 1 (2004), pp. 89–97.

[9] T.F. Chan and R. Glowinski, *Finite Element Approximation and Iterative Solution of a Class of Mildly Non-linear Elliptic Equations*, Technical report, Stanford University, Stanford, CA, 1978.

[10] S. Chatterjee, K. Steinhaeuser, A. Banerjee, S. Chatterjee, and A. Ganguly, *Sparse group Lasso: Consistency and climate applications*, in Proceedings of the SIAM International Conference on Data Mining, SIAM, Philadelphia, 2012, pp. 47–58.

[11] C.K. Chui, *An introduction to wavelets*, Academic Press Professional, New York, 1992.

[12] E. Corman and X.M. Yuan, *A Generalized Proximal Point Algorithm and Its Convergence Rate*, manuscript, 2012.

[13] G. Demoment, *Image reconstruction and restoration: Overview of common estimation structures and problems*, IEEE Trans. Acoust. Speech Signal Process., 37 (1989), pp. 2024–2036.

[14] D.L. Donoho and Y. Tsaig, *Fast solution of l_1-norm minimization problems when the solution may be sparse*, IEEE Trans. Inform. Theory, 54 (2008), pp. 4789–4812.

[15] J. Douglas and H.H. Rachford, *On the numerical solution of the heat conduction problem in* 2 *and* 3 *space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 420–439.

[16] J. Eckstein, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph.D. Dissertation, MIT, Cambridge, MA, 1989.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

[17] J. Eckstein, *Augmented Lagrangian and Alternating Direction Methods for Convex Optimization: A Tutorial and Some Illustrative Computational Results*, manuscript, 2012.

[18] J. Eckstein and D.P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[19] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems Vol.* I, Springer Ser. Oper. Res., Springer-Verlag, New York, 2003.

[20] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrange Methods: Applications to the Solution of Boundary-Valued Problems, M. Fortin and R. Glowinski, eds., North–Holland, Amsterdam, 1983, pp. 299–331.

[21] 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.

[22] R. Glowinski, *On Alternating Direction Methods of Multipliers: A Historical Perspective*, manuscript, 2012.

[23] R. Glowinski, T. Kärkkäinen, and K. Majava, *On the convergence of operator-splitting methods*, in Proceedings of CIMNE 2003: Numerical Methods for Scientific Computing, Variational Problems and Applications, Y. Kuznetsov, P. Neittanmaki, and O. Pironneau, eds., Barcelona, Spain, 2003, pp. 67–79.

[24] R. Glowinski and A. Marrocco, *Approximation par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires*, RAIRO Anal. Numér., 9 (1975), pp. 41–76.

[25] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator Splitting Methods in Nonlinear Mechanics*, SIAM Stud. Appl. Math. 9, SIAM, Philadelphia, 1989.

[26] E.G. Gol'shtein and N.V. Tret'yakov, *Modified Lagrangian in convex programming and their generalizations*, Math. Program. Stud., 10 (1979), pp. 86–97.

[27] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Addison–Wesley Longman, Harlow, UK, 2002.

[28] P.C. Hansen, J.G. Nagy, and D.P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, Fund. Algorithms 3, SIAM, Philadelphia, 2006.

[29] B.S. He, L.Z. Liao, D.R. Han, and H. Yang, *A new inexact alternating directions method for monontone variational inequalities*, Math. Program., 92 (2002), pp. 103–118.

[30] B.S. He and H. Yang, *Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities*, Oper. Res. Lett., 23 (1998), pp. 151–161.

[31] B.S. He and X.M. Yuan, *On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.

[32] B.S. He and X.M. Yuan, *On nonergodic convergence rate of Douglas-Rachford alternating direction method of multipliers*, Numer. Math., submitted, 2012.

[33] G.T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Springer, New York, 2009.

[34] K. Koh, S.-J. Kim, and S. Boyd, *An interior-point method for large-scale l_1-regularized logistic regression*, J. Mach. Learn. Res., 8 (2007), pp. 1519–1555.

[35] B. Krishnapuram, L. Carin, M.A. Figueiredo, and A.J. Hartemink, *Sparse multinomial logistic regression: Fast algorithms and generalization bounds*, IEEE Trans. Pattern Anal. Machine Intell., 27 (2005), pp. 957–968.

[36] C. Li, *Compressive Sensing for 3D Data Processing Tasks: Applications, Models and Algorithms*, Ph.D. thesis, Rice University, Houston, TX, 2011.

[37] C. Li, W. Yin, H. Jiang, and Y. Zhang, *An efficient augmented Lagrangian method with applications to total variation minimization*, Comput. Optim. Appl., 56 (2013), pp. 507–530.

[38] C. Li, W. Yin, and Y. Zhang, *User's Guide for TVAL3: TV Minimization by Augmented Lagrangian and Alternating Direction Algorithms*, Rice University CAAM Technical Report, Rice University, Houston, TX, 2009.

[39] P.L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.

[40] H. Liu, M. Palatucci, and J. Zhang, *Blockwise coordinate descent procedures for the multitask lasso, with applications to neural semantic basis discovery*, in Proceeding of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 2009, pp. 649–656.

[41] S. Ma, X. Song, and J. Huang, *Supervised group Lasso with applications to microarray data analysis*, BMC Bioinformatics, 8 (2007), pp. 60–72.

[42] A.S. Nemirovsky and D.B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley-Interscience Series in Discrete Mathematics, John Wiley and Sons, New York, 1983.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[43] Y.E. Nesterov, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$*, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547 (in Russian).

[44] Y. Nesterov, *Gradient Methods for Minimizing Composite Objective Function*, Core Discussion Paper 2007/96, Center for Operations Research and Econometrics, Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2007.

[45] M.Y. Park and T. Hastie, *L_1-regularization path algorithm for generalized linear models*, J. R. Stat. Soc. Ser. B Methodol., 69 (2007), pp. 659–677.

[46] D.W. Peaceman and H.H. Rachford, Jr., *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Indust. Appl. Math., 3 (1955), pp. 28–41.

[47] W.K. Pratt, *Digital Image Processing: PIKS Inside*, 3rd ed., John Wiley and Sons, New York, 2001.

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

[49] M. Soumekh, *Synthetic Aperture Radar Signal Processing*, Wiley, New York, 1999.

[50] B.P. Sutton, D.C. Noll, and J.A. Fessler, *Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities*, IEEE Trans. Med. Imaging, 17 (2003), pp. 178–188.

[51] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 32 (1996), pp. 267–288.

[52] S.J. Wright, R.D. Nowak, and M.A.T. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 43 (2009), pp. 2479–2493.

[53] T.T. Wu, Y.F. Chen, T. Hastie, E. Sobel, and K. Lange, *Genome-wide association analysis by lasso penalized logistic regression*, Bioinformatics, 19 (2009), pp. 714–720.

[54] J. Yang and Y. Zhang, *Alternating direction algorithms for ℓ_1-problems in compressive sensing*, SIAM J. Sci. Comput., 33 (2011), pp. 250–278.

[55] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, J. R. Stat. Soc. Ser. B Stat. Methodol., 68 (2006), pp. 49–67.

[56] X.M. Yuan, *Alternating direction methods for covariance selection models*, J. Sci. Comput., 51 (2012), pp. 261–273.

[57] Y. Zhang, *User's Guide for YALL1: Your Algorithms for L1 Optimization*, Rice University CAAM Technical Report TR09-17, Rice University, Houston, TX, 2009.

Downloaded 07/28/14 to 158.182.168.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.