Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2016, Article ID 3934872, 11 pages
http://dx.doi.org/10.1155/2016/3934872

Research Article

A General Solution to Least Squares Problems with Box Constraints and Its Applications

Yueyang Teng, Shouliang Qi, Dayu Xiao, Lisheng Xu, Jianhua Li, and Yan Kang

Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang 110004, China

Correspondence should be addressed to Yan Kang; kangyan@bmie.neu.edu.cn

Received 15 November 2015; Revised 4 March 2016; Accepted 18 April 2016

Academic Editor: Masoud Hajarian

The main contribution of this paper is presenting a flexible solution to the box-constrained least squares problems. This solution is applicable to many existing problems, such as nonnegative matrix factorization, support vector machine, signal deconvolution, and computed tomography reconstruction. The key concept of the proposed algorithm is to replace the minimization of the cost function at each iteration by the minimization of a surrogate, leading to a guaranteed decrease in the cost function. In addition to the monotonicity, the proposed algorithm also owns a few good features including the self-constraint in the feasible region and the absence of a predetermined step size. This paper theoretically proves the global convergence for a special case of below-bounded constraints. Using the proposed mechanism, some valuable algorithms can be derived. The simulation results demonstrate that the proposed algorithm provides performance that is comparable to that of other commonly used methods in numerical experiment and computed tomography reconstruction.

1. Introduction

Solving the linear system $Y = AX$ is a classic inverse problem, where $X \in R^N$ and $Y \in R^M$ are vectors and $A \in R^{M\times N}$ is a matrix. This problem is applicable in many fields, including nonnegative matrix factorization (NMF), support vector machine (SVM), signal deconvolution, and medical image reconstruction (e.g., computed tomography (CT)) [1–5]. In many cases, we need to impose a box constraint on the problem. A classical approach to choosing X is to minimize the least squares (LS) error between Y and AX:

$$\min_{C_1\le X\le C_2} F(X) = \|AX - Y\|^2, \tag{1}$$

where C_1 and C_2 are given constant vectors and $\|\cdot\|$ denotes the Euclidean distance.

In fact, (1) can be easily converted into an equivalent form by a linear transformation $\xi = X - C_1$, which is convenient to study. In the following, we only focus on such LS problems and still denote variables with X for simple notation; subsequently, the readers may generalize all the results of this paper to the generic form by themselves:

$$\min_{0\le X\le C} F(X) = \|AX - Y\|^2, \tag{2}$$

where C is a constant vector.

If we denote $C = +\infty$, then it is the widely used case, that is, the nonnegativity constrained LS problem:

$$\min_{X\ge 0} F(X) = \|AX - Y\|^2. \tag{3}$$

There are many methods for solving nonnegativity constrained problems. A simple method for solving such problems is by using

$$\begin{aligned} X &= \left(A^T A\right)^{-1} A^T Y, \\ X &= \max\{X, 0\}. \end{aligned} \tag{4}$$

Of course, we must truncate the negative pixel values to satisfy the nonnegativity constraint. However, we will encounter two difficulties: perhaps $A^T A$ is not invertible or $A^T A$ is too large. A viable example for the minimization of (3) is

the gradient-based method. The steepest descent method is perhaps the simplest technique to implement, which takes the negative gradient as the descent direction:

$$X^{t+1} = X^t - \alpha(t)\nabla F\left(X^t\right),$$
$$X^{t+1} = \max\left\{X^{t+1}, 0\right\}, \tag{5}$$

where the superscript (t) denotes the tth iteration and $\alpha(t)$ is the step size.

Let $\alpha(t) = \alpha$ be a constant, where $0 < \alpha < 2/\|\nabla^2 F(X^t)\|_E$, and $\|\cdot\|_E$ denotes the maximum eigenvalue for a matrix; then, (5) becomes the projected Landweber (PL) method [6]:

$$X^{t+1} = X^t - \alpha\nabla F\left(X^t\right),$$
$$X^{t+1} = \max\left\{X^{t+1}, 0\right\}. \tag{6}$$

In addition, Lantéri et al. [7, 8] provided a general multiplicative algorithm, which was also a gradient-based method. Let $a_j(X^t)$ be a positive function with positive values for any $X_j^t > 0$; then, $-\operatorname{diag}[X_j^t a_j(X^t)]\nabla F(X^t)$ is also a descent direction. Therefore, the algorithm in a modified gradient form can be written as

$$X_j^{t+1} = X_j^t - X_j^t a_j\left(X^t\right)\nabla_j F\left(X^t\right). \tag{7}$$

Again, let $U_j(X^t)$ and $V_j(X^t)$ be two positive functions for any $X_j^t > 0$, which satisfy $-\nabla_j F(X^t) = U_j(X^t) - V_j(X^t)$. Taking $a_j(X^t) = 1/V_j(X^t)$, Lantéri et al. obtained a multiplicative update

$$X_j^{t+1} = X_j^t \frac{U_j\left(X^t\right)}{V_j\left(X^t\right)}. \tag{8}$$

The conjugate gradient (CG) method is also a popular method, and it is often implemented as an iterative algorithm, applicable to sparse systems that are too large to be handled by a direct implementation.

The above approaches are not effective for processing the upper-bounded constraints, and, generally, a truncation is imposed to the update rule, which may lead to a divergent modification.

In this paper, we consider a specific application of the surrogate-based methods [9] to a specific type of LS problem with box constraints. We derive a multiplicative update rule to iteratively and monotonically (in the sense of decreasing the cost function) solve the problem, similar to the EM algorithm [10]. The nonnegativity constraints will be satisfied automatically, and the upper-bounded constraints are performed by an upper truncation. Meanwhile, the algorithm still has the monotonicity and does not require an adjustable step size. We provide a rigorous global convergence proof for the case of only nonnegativity constraints, which can be easily generalized to the below-constrained case. We demonstrate the power of this mechanism by deriving many existing and new algorithms. We also present some computer simulation results to show the desirable behavior of the proposed algorithm with respect to the convergence rate and the stability compared with existing methods.

2. Methodology

A note about our notation: all vectors will be column vectors unless transposed to a row vector by a prime superscript T. For a matrix or vector X, $X \geq 0$ means that any component of X is equal to or greater than 0. For a matrix A, A_{i*} and A_{*i} represent the ith row and ith column of A, respectively.

As mentioned in many articles [1, 2, 9, 11], a surrogate as defined below is useful in algorithm derivations and convergence proofs.

Definition 1 (surrogate). Denote $\psi(X \mid X^t)$ as a surrogate for $\Psi(X)$ at X^t (fixed) if $\psi(X^t \mid X^t) = \Psi(X^t)$ and $\psi(X \mid X^t) \geq \Psi(X)$.

Clearly, $\Psi(X)$ is decreasing under the update $X^{t+1} = \min_X \psi(X \mid X^t)$ because

$$\Psi\left(X^{t+1}\right) \leq \psi\left(X^{t+1} \mid X^t\right) \leq \psi\left(X^t \mid X^t\right) = \Psi\left(X^t\right). \tag{9}$$

There are two obvious and important properties for a surrogate: additivity and transitivity. For the former, the sum of two surrogates is a surrogate of the sum of the two original functions. For the latter, the surrogate of a surrogate of a function is a surrogate of this function.

In the following, we proceed step by step: firstly, assume that $A \geq 0$ and $C = +\infty$; secondly, remove the restriction of $A \geq 0$; and thirdly, remove the restriction of $C = +\infty$.

2.1. $A \geq 0$ and $C = +\infty$. Let $-Y = Y^+ - Y^-$, where Y^+ and Y^- are nonnegative vectors; then, $F(X) = \|AX + Y^+ - Y^-\|^2$. Note that we decompose $-Y$ instead of Y for convenient derivation, which can be seen below. We construct a surrogate $f(X \mid X^t)$ by the convexity of $F(X)$. Denote

$$\lambda_{i*} = \frac{Y_i^+}{(AX^t + Y^+)_i},$$
$$\lambda_{ij} = \frac{A_{ij}X_j^t}{(AX^t + Y^+)_i} \tag{10}$$

that satisfy $\lambda_{i*}, \lambda_{ij} \geq 0$ and $\lambda_{i*} + \sum_{j=1}^N \lambda_{ij} = 1$. They can be the convex combination coefficients such that

$$f\left(X \mid X^t\right) = \sum_{i=1}^M \left[\lambda_{i*}\left(\frac{Y_i^+}{\lambda_{i*}} - Y_i^-\right)^2 + \sum_{j=1}^N \lambda_{ij}\left(\frac{A_{ij}X_j}{\lambda_{ij}} - Y_i^-\right)^2\right]. \tag{11}$$

It is easy to verify that $f(X^t \mid X^t) = F(X^t)$. If considering Jensen's inequality and the convex combination coefficients λ_{ij}, then $f(X \mid X^t) \geq F(X)$ is proven by the following inequality:

$$\lambda_{i*}\left(\frac{Y_i^+}{\lambda_{i*}} - Y_i^-\right)^2 + \sum_{j=1}^N \lambda_{ij}\left(\frac{A_{ij}X_j}{\lambda_{ij}} - Y_i^-\right)^2 \geq \left[(AX)_i + Y_i^+ - Y_i^-\right]^2. \tag{12}$$

Take the partial derivatives of $f(X \mid X^t)$; then, we can solve the one-dimensional equations $\partial f(X \mid X^t)/\partial X_j = 0$ to obtain a multiplicative update rule:

$$X_j^{t+1} = X_j^t \frac{\left(A^T Y^-\right)_j}{\left[A^T\left(AX^t + Y^+\right)\right]_j}. \tag{13}$$

2.2. Any A but C = +∞. Let $A = A^+ - A^-$, $w_1 + w_2 = 1$, and $w_1, w_2 > 0$, where A^+ and A^- are two nonnegative matrices; then, we can construct the surrogate $f_{\text{mid}}(X \mid X^t)$ for $F(X)$ as follows:

$$f_{\text{mid}}\left(X \mid X^t\right) = \overline{f}\left(X \mid X^t\right) + \widehat{f}\left(X \mid X^t\right), \tag{14}$$

where

$$\begin{aligned}
\overline{f}\left(X \mid X^t\right) &= w_1 \left\| \frac{1}{w_1}\left[A^+X + Y^+\right] - \left(\frac{1}{w_1} - 1\right)\left[A^+X^t + Y^+\right] - \left[A^-X^t + Y^-\right] \right\|^2, \\
\widehat{f}\left(X \mid X^t\right) &= w_2 \left\| \frac{1}{w_2}\left[A^-X + Y^-\right] - \left[A^+X^t + Y^+\right] - \left(\frac{1}{w_2} - 1\right)\left[A^-X^t + Y^-\right] \right\|^2.
\end{aligned} \tag{15}$$

It is easy to verify that $f_{\text{mid}}(X^t \mid X^t) = F(X^t)$. By the convexity of $F(X)$, it is clear that

$$\begin{aligned}
f_{\text{mid}}\left(X \mid X^t\right) &= w_1 \left\| \frac{1}{w_1}\left[A^+X + Y^+\right] - \left(\frac{1}{w_1} - 1\right)\left[A^+X^t + Y^+\right] - \left[A^-X^t + Y^-\right] \right\|^2 \\
&+ w_2 \left\| -\frac{1}{w_2}\left[A^-X + Y^-\right] + \left[A^+X^t + Y^+\right] + \left(\frac{1}{w_2} - 1\right)\left[A^-X^t + Y^-\right] \right\|^2 \geq F(X) \\
&\qquad \left(\text{note that } w_1 + w_2 = 1\right).
\end{aligned} \tag{16}$$

Following the same process as in Section 2.1, we can construct surrogates for $\overline{f}(X \mid X^t)$ and $\widehat{f}(X \mid X^t)$. Let

$$\begin{aligned}
\lambda_{ij}^+ &= \frac{A_{ij}^+ X_j^t}{\left(A^+X^t + Y^+\right)_i}, \\
\lambda_{i*}^+ &= \frac{Y_i^+}{\left(A^+X^t + Y^+\right)_i}, \\
\lambda_{i*}^- &= \frac{Y_i^-}{\left(A^-X^t + Y^-\right)_i}, \\
\lambda_{ij}^- &= \frac{A_{ij}^- X_j^t}{\left(A^-X^t + Y^-\right)_i}, \\
\overline{B} &= \left(\frac{1}{w_1} - 1\right)\left[A^+X^t + Y^+\right] + \left[A^-X^t + Y^-\right], \\
\widehat{B} &= \left[A^+X^t + Y^+\right] + \left(\frac{1}{w_2} - 1\right)\left[A^-X^t + Y^-\right];
\end{aligned} \tag{17}$$

then,

$$\begin{aligned}
\overline{\overline{f}}\left(X \mid X^t\right) &= w_1 \sum_{i=1}^{M}\sum_{j=1}^{N} \lambda_{ij}^+ \left[\frac{1}{w_1}\frac{A_{ij}^+ X_j}{\lambda_{ij}^+} - \overline{B}_i\right]^2 + \text{const}, \\
\widehat{\widehat{f}}\left(X \mid X^t\right) &= w_2 \sum_{i=1}^{M}\sum_{j=1}^{N} \lambda_{ij}^- \left[\frac{1}{w_2}\frac{A_{ij}^- X_j}{\lambda_{ij}^-} - \widehat{B}_i\right]^2 + \text{const}.
\end{aligned} \tag{18}$$

Note that λ_{i*}^+ and λ_{i*}^- are relative to the constant terms, which do not work for the optimization problems; thus, we ignore them. We can now obtain a surrogate for $F(X)$ at X^t:

$$f\left(X \mid X^t\right) = \overline{\overline{f}}\left(X \mid X^t\right) + \widehat{\widehat{f}}\left(X \mid X^t\right). \tag{19}$$

Take the partial derivatives of $\overline{\overline{f}}(X \mid X^t)$ and $\widehat{\widehat{f}}(X \mid X^t)$:

$$\begin{aligned}
\frac{\partial \overline{\overline{f}}\left(X \mid X^t\right)}{\partial X_j} &= \frac{2X_j}{w_1 X_j^t}\left[\left(A^+\right)^T\left(A^+X^t + Y^+\right)\right]_j - 2\left[\left(A^+\right)^T \overline{B}\right]_j, \\
\frac{\partial \widehat{\widehat{f}}\left(X \mid X^t\right)}{\partial X_j} &= \frac{2X_j}{w_2 X_j^t}\left[\left(A^-\right)^T\left(A^-X^t + Y^-\right)\right]_j - 2\left[\left(A^-\right)^T \widehat{B}\right]_j.
\end{aligned} \tag{20}$$

Then, we solve the one-dimensional equations $\partial f(X \mid X^t)/\partial X_j = 0$ to obtain a multiplicative update rule:

$$\begin{aligned}
X_j^{t+1} &= X_j^t \cdot \frac{w_1 w_2\left[\left(A^+\right)^T \overline{B} + \left(A^-\right)^T \widehat{B}\right]_j}{\left[w_2\left[\left(A^+\right)^T\left(A^+X^t + Y^+\right)\right]_j + w_1\left[\left(A^-\right)^T\left(A^-X^t + Y^-\right)\right]_j\right.} \\
&= X_j^t \cdot \frac{\left\{\left(w_2 A^+ + w_1 A^-\right)^T\left[w_2\left(A^+X^t + Y^+\right) + w_1\left(A^-X^t + Y^-\right)\right]\right\}_j}{\left[w_2\left[\left(A^+\right)^T\left(A^+X^t + Y^+\right)\right]_j + w_1\left[\left(A^-\right)^T\left(A^-X^t + Y^-\right)\right]_j\right.}.
\end{aligned} \tag{21}$$

2.3. Any A and C. As shown, (21) will ensure the monotonic decrease of the cost function and satisfy the nonnegativity constraints. In the following, it will be modified by a truncation at the end of each iteration to satisfy the box constraints:

$$X = \min\{X, C\}. \tag{22}$$

We will show that the truncation still ensures the monotonic decrease of the cost function.

It is easy to see the separability of the variables of (19); that is, $f(X, X^t) = \sum_{j=1}^{N} f_j(X_j, X^t)$, and every separated function $f_j(X_j, X^t)$ has a quadratic form:

$$f_j\left(X_j, X^t\right) = w_1 \sum_{i=1}^{M} \lambda_{ij}^{+} \left[\frac{1}{w_1} \frac{A_{ij}^{+} X_j}{\lambda_{ij}^{+}} - \overline{B}_i\right]^2 + w_2 \sum_{i=1}^{M} \lambda_{ij}^{-} \left[\frac{1}{w_2} \frac{A_{ij}^{-} X_j}{\lambda_{ij}^{-}} - \widehat{B}_i\right]^2 + \text{const}. \tag{23}$$

We further simplify it into a general form:

$$g\left(X_j\right) = a\left(X_j - b\right)^2 + \text{const}, \tag{24}$$

where $a > 0$, $b > 0$, and $X_j^t \in (0 \quad C]$. It is clear that $X_j = b$ minimizes the quadratic function. If $b \le C$, we change nothing. If $b > C$, then it is easy to obtain that $g(X_j^t) \ge g(C)$ because $g'(X_j) = 2a(X_j - b) < 0$ on the open interval $(X_j^t \quad b)$ such that $g(X_j)$ strictly monotonously decreases on that open interval. Thus, the proof is established.

In fact, we know that $X = \min\{X, +\infty\}$; therefore, we may provide a generic solution to (2), which is the main result of this paper.

Algorithm 2. Start from an initial point $X^0 > 0$; then,

(1) update X by (21);

(2) truncate X by (22).

It is easy to generalize the constraints from $0 \le X \le C$ to $C_1 \le X \le C_2$ by the transformation $\xi = X - C_1$, which is left for readers to complete by themselves.

3. Convergence

The convergence proof with box constraints is very difficult; thus, we only focus on the below-bounded constraints, which can be viewed as a special case of box constraints with $C = +\infty$. Additionally, we consider the equivalence between the general below-constrained form and the nonnegative constrained one by a linear transformation $\xi = X - C_1$; then, we will theoretically prove the global convergence of the latter case for simplification, which is easy to extend to that of the former. In theory, the Kuhn-Tucker (KT) point will be a global solution if the cost function is convex. It is easy to see the convexity of $F(X)$. By Theorem 2.19 in [12], the KT conditions of (3) ($X \ge 0$) are as follows:

$$\frac{\partial F(X)}{\partial X_j} = 0 \quad \text{if } X_j > 0, \tag{25}$$

$$\frac{\partial F(X)}{\partial X_j} \ge 0 \quad \text{if } X_j = 0. \tag{26}$$

We entertain several important and reasonable assumptions.

Assumption 3. For the iteration sequence $\{X^t\}$, we assume that

(1) the algorithm starts from a positive image;

(2) $(A^T A)_{jj} > 0$ for all j;

(3) F is a strictly convex function.

The first assumption forces the iterations to be positive, but the limit may be zero. The second condition is reasonable because $(A^T A)_{jj} = 0$ suggests $A_{ij} = 0$ for any i. Thus, the equation $Y = AX$ is irrelevant to X_j, and, then, X_j in X is removable. The third one is indeed restrictive; however, it is important for the following derivation.

We will prove the global convergence along the lines of [13–17]. First, we provide several useful lemmas.

Lemma 4. *The set of accumulation points of a bounded sequence $\{Z^t\}$ with $\{\|Z^{t+1} - Z^t\|\} \to 0$ is connected and compact.*

Proof. This is Theorem 28.1 from Ostrowski [18]. The reader is kindly referred to this paper for the proof. □

Lemma 5. *The iteration sequence $\{X^t\}$ is bounded.*

Proof. Because F is a strictly convex function, $A^T A$ is a positive definite symmetric matrix, and, thus, it can be factorized into $A^T A = U^T U$ by Cholesky's method, where U is an invertible matrix. We assume that X^* is the sole minimum, and, then, we expand F on X^* using a Taylor series:

$$\begin{aligned} F(X) - F(X^*) &= (X - X^*)^T \left(A^T A\right)(X - X^*) \\ &= (X - X^*)^T U^T U (X - X^*) \ge 0. \end{aligned} \tag{27}$$

Let $\xi = U(X - X^*)$; then, for any constant L, it is easy to know the boundedness of $S = \{\xi : \xi^T \xi < L\}$ such that $\{X : X = U^{-1}\xi + X^*,\ \xi \in S\}$ is bounded, which is equivalent to $\{X : F(X) - F(X^*) < L\}$ being bounded. Let $L_0 = F(X^0) - F(X^*)$; then, $\{X^t\} \subset \{X : F(X) - F(X^*) < L_0\}$ because $\{F(X^t)\}$ monotonically decreases. Therefore, $\{X^t\}$ is bounded. □

Lemma 6. *The sequence $\{\|X^t - X^{t+1}\|\} \to 0$.*

Proof. Take into account that

$$\begin{aligned} \nabla_{jj}^2 f\left(X \mid X^t\right) &= \frac{2}{w_1 X_j^t} \left[(A^+)^T \left(A^+ X^t + Y^+\right)\right]_j \\ &\quad + \frac{2}{w_2 X_j^t} \left[(A^-)^T \left(A^- X^t + Y^-\right)\right]_j \\ &\ge \frac{2}{w_1 X_j^t} \left[(A^+)^T A^+ X^t\right]_j \\ &\quad + \frac{2}{w_2 X_j^t} \left[(A^-)^T A^- X^t\right]_j \end{aligned}$$

$$\geq \frac{2}{w_1}\sum_{i=1}^{M}\left(A_{ij}^{+}\right)^2 + \frac{2}{w_2}\sum_{i=1}^{M}\left(A_{ij}^{-}\right)^2 = \gamma_j > 0 \quad \left(\text{see the second one of Assumption 3}\right). \tag{28}$$

Let $\gamma = \min_j\{\gamma_j\}$. Because $\nabla f(X^{t+1} \mid X^t) = 0$,

$$\begin{aligned} F\left(X^t\right) - F\left(X^{t+1}\right) &\geq f\left(X^t \mid X^t\right) - f\left(X^{t+1} \mid X^t\right) \\ &= \frac{1}{2}\left(X^t - X^{t+1}\right)^T \nabla^2 f\left(X^{t+1} \mid X^t\right)\left(X^t - X^{t+1}\right) \\ &\geq \frac{\gamma}{2}\left\|X^t - X^{t+1}\right\|^2. \end{aligned} \tag{29}$$

Because $\{F(X^t)\}$ monotonically decreases and is bounded from below, $\{F(X^t) - F(X^{t+1})\} \to 0$; therefore, $\{\|X^t - X^{t+1}\|\} \to 0$. □

Lemma 7. *If a subsequence $\{X^{t_s}\} \to X^*$, then $\{X^{t_s+1}\} \to X^*$ as well.*

Proof. By contradiction, if $\{X^{t_s+1}\}$ diverges, then it must have a convergent subsequence $\{X^{t_{ss}+1}\} \to X^{**} \neq X^*$ because of the boundness by Lemma 5. Let $\epsilon_0 = \|X^* - X^{**}\| > 0$, and if we consider the two convergent subsequences $\{X^{t_{ss}}\}$ and $\{X^{t_{ss}+1}\}$, then there must be a positive integer $\mathcal{T}$ to make $\|X^{t_{ss}} - X^*\| < \epsilon_0/4$ and $\|X^{t_{ss}+1} - X^{**}\| < \epsilon_0/4$ when $t_{ss} > \mathcal{T}$. By the triangle inequality, we can obtain the contradictive result to $\{\|X^t - X^{t+1}\|\} \to 0$ as

$$\begin{aligned} &\left\|X^{t_{ss}} - X^{t_{ss}+1}\right\| + \left\|X^{t_{ss}} - X^*\right\| + \left\|X^{t_{ss}+1} - X^{**}\right\| \\ &\quad \geq \left\|X^* - X^{**}\right\| \Longrightarrow \left\|X^{t_{ss}} - X^{t_{ss}+1}\right\| > \frac{\epsilon_0}{2}. \end{aligned} \tag{30}$$

□

Lemma 8. *At each iteration, one knows that*

$$\begin{aligned} X_j^{t+1} &= X_j^t - \alpha_j\left[\nabla F\left(X^t\right)\right]_j, \\ \alpha_j &= \frac{w_1 w_2 X_j^t}{2\left[\, w_2\left[(A^+)^T\left(A^+X^t + Y^+\right)\right]_j + w_1\left[(A^-)^T\left(A^-X^t + Y^-\right)\right]_j\right.}. \end{aligned} \tag{31}$$

Proof. We consider that

$$\begin{aligned} \nabla F\left(X^t\right) &= 2\left(A^+ - A^-\right)^T\left[\left(A^+X^t + Y^+\right) - \left(A^-X^t + Y^-\right)\right] = 2\left(A^+\right)^T\left(A^+X^t + Y^+\right) \\ &\quad + 2\left(A^-\right)^T\left(A^-X^t + Y^-\right) - 2\left(A^+\right)^T\left(A^-X^t + Y^-\right) \\ &\quad - 2\left(A^-\right)^T\left(A^+X^t + Y^+\right) = \frac{2}{w_1}\left[\left(A^+\right)^T \cdot \left(A^+X^t + Y^+\right)\right] + \frac{2}{w_2}\left[\left(A^-\right)^T\left(A^-X^t + Y^-\right)\right] \\ &\quad - 2\left(A^+\right)^T\left[\left(\frac{1}{w_1} - 1\right)\left(A^+X^t + Y^+\right) + \left(A^-X^t + Y^-\right)\right] - 2\left(A^-\right)^T\left[\left(A^+X^t + Y^+\right) + \left(\frac{1}{w_2} - 1\right)\left(A^-X^t + Y^-\right)\right] \\ &= \frac{2}{w_1 w_2}\left\{w_2\left[\left(A^+\right)^T\left(A^+X^t + Y^+\right)\right] + w_1\left[\left(A^-\right)^T\left(A^-X^t + Y^-\right)\right]\right\} \\ &\quad - \frac{2}{w_1 w_2}\left\{\left(w_2A^+ + w_1A^-\right)^T \cdot \left[w_2\left(A^+X^t + Y^+\right) + w_1\left(A^-X^t + Y^-\right)\right]\right\} \end{aligned} \tag{32}$$

$$= \nabla f\left(X^t \mid X^t\right) \quad \left(\text{compare (32) with (20)}\right). \tag{33}$$

It is easy to verify the correctness of (31) if replacing $\nabla F(X^t)$ by (32). □

The global convergence will be proven from the three theorems below.

Theorem 9. *Let $\{X^{t_s}\} \to X^*$ be any convergent subsequence; then, X^* meets the first KT condition (25).*

Proof. When $X_j^* > 0$, it is easy to obtain that $[\nabla F(X^*)]_j = [\nabla f(X^* \mid X^*)]_j$ by (33). In addition, $\partial F(X \mid X^{t_s})/\partial X_j|_{X^{t_s+1}} = 0$ and $\{X^{t_s}\} \to X^*$ such that $\{X^{t_s+1}\} \to X^*$ (Lemma 7); thus,

$$\left.\frac{\partial F\left(X \mid X^*\right)}{\partial X_j}\right|_{X^*} = \lim_{t_s \to +\infty}\left.\frac{\partial F\left(X \mid X_s^t\right)}{\partial X_j}\right|_{X^{t_s+1}} = 0. \tag{34}$$

□

Theorem 10. *The entire sequence $\{X^t\}$ converges.*

Proof. According to Lemmas 4, 5, and 6, the set of accumulation points of $\{X^t\}$ is connected and compact. If we can prove that the number of accumulation points is finite, then the desired result follows because a finite set can be connected only if it consists of a single point [16].

To prove the existence of a finite number of accumulation points, we consider any accumulation point X^*. Given an integer set $\Omega = \{1, 2, \ldots, \mathcal{T}\}$, where $\mathcal{T}$ is the total number of components of X, $\Omega^* = \{j : X_j^* = 0\}$ is a subset of Ω. Let F_{Ω^*} be the restrictions of F to the set $\{X : X_j = 0,\ j \in \Omega^*\}$, which is a strictly convex function of the reduced variables. It follows that F_{Ω^*} has a unique minimum (Theorem 9: $\partial F(X^*)/\partial X_j = 0$ if $X_j^* > 0$). This means that an accumulation point must correspond to a subset of Ω. The number of subsets of Ω is finite, and, thus, the number of accumulation points is also finite. □

In Theorem 9, we prove that every accumulation point meets the first KT condition, by which the full sequence convergence is provided in Theorem 10. Naturally, the limit of $\{X^t\}$ satisfies the first KT condition. In the following, we will show that the second KT condition is satisfied.

Theorem 11. *The limit X^* of $\{X^t\}$ satisfies the second KT condition (26).*

Proof. When $X_j^* = 0$, by contradiction, we assume that there is an $X_j^* = 0$ that satisfies $[\nabla F(X^*)]_j < 0$. Because $\{X^t\} \to X^*$, there exists an $\epsilon < 0$ and a positive integer $\mathcal{T}$ such that $[\nabla F(X^t)]_j < \epsilon$ for $t > \mathcal{T}$; then,

$$X_j^{t+1} - X_j^t = -\alpha_j \left[\nabla F\left(X^t\right)\right]_j > -\alpha_j \epsilon > 0 \quad \text{(by Lemma 8)}. \tag{35}$$

Thus, we can obtain that $X_j^{t+1} > X_j^t$, which is a contradiction to $\{X_j^t\} \to 0$. □

4. Solutions to Some Existing and New Examples

4.1. NMF. NMF has been widely used in pattern recognition, machine learning, and data mining. It decomposes the nonnegative matrix Y by the product of two other nonnegative matrices A and X. Lee and Seung [1] have proposed using the square of the Euclidean distance to measure the similarity between Y and AX:

$$\min_{A\geq 0, X\geq 0} F(A, X) = \|Y - AX\|^2. \tag{36}$$

Minimizing them under the constraints $A \geq 0$ and $X \geq 0$, Lee and Seung [1] alternated between solving an optimization problem in the variables A and then solving another one in the variables X. As can be seen, A and X have the same roles.

To take X, for instance, with $X^t > 0$ and $A > 0$ given, denote $\lambda_{ijk} = A_{ik}^t X_{kj}^t / (AX^t)_{ij}$ such that the surrogate function can be constructed:

$$f\left(X \mid X^t\right) = \sum_{ijk} \left[\lambda_{ijk} \left(\frac{A_{ik} X_{kj}}{\lambda_{ijk}} - Y_{ij}\right)^2\right]. \tag{37}$$

The update rule can be obtained by solving $\partial f(X, X^t)/\partial X_{kj} = 0$. Reversing the roles of A and X, we can similarly construct the surrogate for A and acquire the update rule:

$$\begin{aligned} X_{kj}^{t+1} &= X_{kj}^t \frac{\left(A^T Y\right)_{kj}}{\left(A^T A X^t\right)_{kj}}, \\ A_{ik}^{t+1} &= A_{ik}^t \frac{\left(Y X^T\right)_{ik}}{\left(A^t X X^T\right)_{ik}}. \end{aligned} \tag{38}$$

4.2. Variation of Linearized SVM. SVM attempts to separate points belonging to two given sets in real Euclidean space R^n by a surface. Several practical applications of SVMs use nonlinear kernels, such as the polynomial and radial basis function kernels. However, in applications such as text classification, linear SVMs are still used because it has been observed that many text classification problems are linearly separable [19]. Most literature on large-scale SVM training have targeted the linear SVM problem, citing this fact. A dual form of linear SVM is usually used because of the simple structure, which can be formulated as [20]

$$\begin{aligned} \min \quad & \frac{1}{2} X^T D A^T A D X - 1^T X \\ \text{s.t.} \quad & 0 \leq X \leq C, \\ & D^T X = 0, \end{aligned} \tag{39}$$

where A is the input data, D is the class label, C is the penalty parameter, and X represents the Lagrangian multiplier that needs to be optimized. Hsieh et al. [21] handled the constraint $D^T X = 0$ by removing it, which corresponded to removing the "intercept" from the classifier. For a simplified expression, we denote $B = AD$, and, then, we obtain the following simple formula:

$$\begin{aligned} \min \quad & \frac{1}{2} X^T B^T B X - 1^T X \\ \text{s.t.} \quad & 0 \leq X \leq C. \end{aligned} \tag{40}$$

We can solve this optimization problem using the proposed method. Let $w_1 = w_2 = 1/2$ and $B = B^+ - B^-$, where B^+ and B^- are two nonnegative matrices; then,

$$f_{\text{mid}}\left(X \mid X^t\right) = \overline{f}\left(X \mid X^t\right) + \widehat{f}\left(X \mid X^t\right) - 1^T X, \tag{41}$$

where

$$\begin{aligned} \overline{f}\left(X \mid X^t\right) &= \frac{1}{4} \left\|2B^+ X - B^+ X^t - B^- X^t\right\|^2, \\ \widehat{f}\left(X \mid X^t\right) &= \frac{1}{4} \left\|2B^- X - B^+ X^t - B^- X^t\right\|^2. \end{aligned} \tag{42}$$

We, respectively, construct surrogates for $\overline{f}(X \mid X^t)$ and $\widehat{f}(X \mid X^t)$. Let

$$\begin{aligned} \lambda_{ij}^+ &= \frac{B_{ij}^+ X_j^t}{(B^+ X^t)_i}, \\ \lambda_{ij}^- &= \frac{B_{ij}^- X_j^t}{(B^- X^t)_i}; \end{aligned} \tag{43}$$

then,

$$\begin{aligned}\overline{\overline{f}}\left(X \mid X^t\right) &= \frac{1}{4}\sum_{i=1}^{M}\sum_{j=1}^{N}\lambda_{ij}^{+}\left[2\frac{B_{ij}^{+}X_j}{\lambda_{ij}^{+}} - \left(B^{+}X^t + B^{-}X^t\right)_i\right]^2, \\ \widehat{\widehat{f}}\left(X \mid X^t\right) &= \frac{1}{4}\sum_{i=1}^{M}\sum_{j=1}^{N}\lambda_{ij}^{-}\left[2\frac{B_{ij}^{-}X_j}{\lambda_{ij}^{-}} - \left(B^{+}X^t + B^{-}X^t\right)_i\right]^2.\end{aligned} \tag{44}$$

We now obtain a surrogate for $F(X)$ at X^t:

$$f\left(X \mid X^t\right) = \overline{\overline{f}}\left(X \mid X^t\right) + \widehat{\widehat{f}}\left(X \mid X^t\right). \tag{45}$$

We solve the one-dimensional equation $\partial f(X \mid X^t)/\partial X_j = 0$ and truncate it to obtain an update:

$$\begin{aligned}X_j^{t+1} &= X_j^t \frac{1 + \left[\left(B^{+} + B^{-}\right)^T\left(B^{+} + B^{-}\right)X^t\right]_j}{2\left\{\left[\left(B^{+}\right)^T B^{+} + \left(B^{-}\right)^T B^{-}\right]X^t\right\}_j}, \\ X_j^{t+1} &= \min\left\{X_j^{t+1}, C\right\}.\end{aligned} \tag{46}$$

4.3. Nonnegative Image Deblurring. In many optical devices, the process of image blurring can be considered to be the result of convolution by a point spread function (PSF) [22, 23]. It is assumed that the degraded image Y is in the form $Y = AX$, where X is the true image and A is the PSF matrix consisting of PSFs at every pixel. As noted in [24, 25], a simple way to approach the deconvolution problem is to find the least squares (LS) estimation between Y and AX. It is well known that the problem of restoring the original image from the noisy and degraded version is an ill-posed inverse problem: small perturbations in the data may result in an unacceptable result [26]. The Tikhonov regularization method [27, 28] is a popular method that generally leads to a unique solution, which is formulated as

$$\min_{X\geq 0} F(X) = \|AX - Y\|^2 + \beta\|RX\|^2. \tag{47}$$

Note that we must approve negative values existing in the matrix R.

In [17], we assume that A and Y are a nonnegative matrix and vector and that there is no restriction on R; then, we, respectively, construct surrogate functions for $\|AX-Y\|^2$ and $\|RX\|^2$ using the method proposed in this paper such that we obtain a multiplicative update rule as follows:

$$X_j^{t+1} = X_j^t \frac{\left(A^TY\right)_j + \beta\left[\left(R^{+} + R^{-}\right)^T\left(R^{+} + R^{-}\right)X^t\right]_j}{\left(A^TAX^t\right)_j + 2\beta\left[\left(\left(R^{+}\right)^T R^{+} + \left(R^{-}\right)^T R^{-}\right)X^t\right]_j}. \tag{48}$$

If $\beta = 0$, it is De Pierro's ISRT (Image Space Reconstruction Technique) [4]:

$$X_j^{t+1} = X_j^t \frac{\left(A^TY\right)_j}{\left(A^TAX^t\right)_j}. \tag{49}$$

4.4. CT Reconstruction with L_1 Regularization. In CT reconstruction, let A be the system probability matrix, Y projections, and R a predetermined matrix. Then, a classical approach is to choose an X (X-ray attenuation image) such that the LS error [5] between Y and AX is minimized:

$$\min_{X\geq 0} F(X) = \|AX - Y\|^2 + \beta\|RX\|_1, \tag{50}$$

where $\beta \geq 0$ serves as a penalty parameter. Many R matrices have been used, such as the first- or second-order derivative matrix or wavelet basis matrix. Note that both A and Y are a nonnegative matrix and vector; however, we must accept negative values existing in the matrix R.

The alternating direction method of multipliers (ADMM) [29, 30] has been developed for solving optimization problems. This method decomposes the original problem into three subproblems and then sequentially solves them at each iteration.

Introducing the additional variable $V = RX$, a fixed penalty parameter ρ, and the Lagrangian multiplier μ, a unified framework can then be given to solve the L_1-norm regularized LS reconstruction problem [29].

Algorithm 12 (ADMM general framework). (1) $V_i^{t+1} = (|(RX^t + \mu^t)_i| - \beta/\rho)_+ \operatorname{sgn}[(RX^t + \mu^t)_i]$, $i = 1, \ldots, M$.

(2) $X^{t+1} = \arg\min_{X\geq 0} L(X, V^{t+1}, \mu^t) = \|AX - Y\|^2 + (\rho/2)\|RX - V^{t+1} + \mu^t\|_2^2$.

(3) $\mu^{t+1} = \mu^t + RX^{t+1} - V^{t+1}$.

The X-update is the most difficult problem that minimizes $L(X, V^{t+1}, \mu^t)$ with nonnegativity constraints. We can, respectively, construct surrogate functions for $\|AX-Y\|^2$ and $\|RX - V^{t+1} + \mu^t\|_2^2$; then minimize the sum of the surrogates to update X; however, there is a high overlap ratio with the above, so we leave them to the readers.

5. Experimental Results

We compare the performance of the proposed surrogate method, with those of the PL method (6) and the CG method on simulated data and CT projection data. For the PL method, we use $\alpha = 1/(\|A\|_1\|A\|_\infty)$ [31], where $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the 1- and ∞-norm of a matrix. The code for the CG method comes from [32], which is slightly modified to meet our criteria.

The experiments are performed on a HP Compaq PC with a 3.00 GHz Core i5 CPU and 4 GB memory. The algorithms are implemented in MATLAB 7.0. All of the algorithms are initiated by the same uniform image for a fair comparison.

Table 1: Performance comparison of the three algorithms.

	Iteration	1	50	150	200	250	300
MSE	Proposed	0.2860	0.2226	0.1961	0.1822	0.1737	0.1682
	CG	0.3441	0.2240	0.3183	0.3183	0.3183	0.3183
	PL	0.3441	0.2398	0.2136	0.2033	0.1978	0.1943
χ	Proposed	0.0031	0.0016	0.0009	0.0006	0.0004	0.0003
	CG	1.0000	0.0001	0	0	0	0
	PL	1.0000	0.0117	0.0037	0.0018	0.0010	0.0007

Table 2: Total running time of the three algorithms with 300 iterations and 50 random start points.

Algorithm	Proposed	CG	PL
Time (second)	2.22	2.13	1.95

The mean square error (MSE) is used to measure the similarity to the true solution, as given below:

$$\mathrm{MSE}(t) = \frac{1}{N}\left\|X^t - X^{\mathrm{True}}\right\|^2. \tag{51}$$

The following criterion is applied to stop the iterative process:

$$\chi(t) = \frac{\left\|X^{t+1} - X^t\right\|}{\|X^t\|} < \epsilon, \tag{52}$$

where ϵ is a difference tolerance.

5.1. Numerical Simulation. Here, we randomly generate $X \in R^{500}$ and $A \in R^{100\times 500}$, where the former are uniformly distributed on $[0 \;\; 1]$ and the latter are on $[-1 \;\; 1]$. Then we obtain Y by $Y = AX$. We will minimize $F(X) = \|Y - AX\|^2$ under the box constraints $0 \le X \le 1$. In the experiment, we generate X, Y, and A once, but we randomly generate start points (uniformly distributed on $[0 \;\; 1]$) 50 times to obtain a reliable averaged result.

Figure 1 presents a comparison of MSE and χ versus iteration number. As shown, CG rapidly finishes iteration to a static solution, so that $\chi(t) = 0$ after that, which can not be shown with a log-scaled y-axis. However, such a phenomenon does not mean a good result. Instead, it is the worst result among the three algorithms by observing the MSE curve. We can conclude that CG is not suitable to the box-constrained LS problem. PL and the proposed method always decrease the curves of MSE and χ. However, the proposed method shows a superior performance than the PL method.

Table 1 shows the quantitative comparison of the three algorithms; then we can draw the same conclusion as above. This table further proves that the proposed method is rapid and stable, and CG is not suitable to the box-constrained LS problem.

Table 2 shows the computational time of the algorithms. As can be seen, they have similar computational complexity; however, the proposed algorithm requires only a little more time because of the nonnegative decomposition of Y and A. In fact, from the above theoretical analysis, we can observe that the proposed algorithm is also some gradient-based method; thus, it has a similar computational complexity with the CG and PL methods.

5.2. CT Reconstruction. CT reconstruction is a medical imaging technique for creating a meaningful diagnostic image. A computer can take the input from the CT machine, run it through the formula, and return a set of images for a physician to examine. Here, we only consider the simplest mathematical model $\min_{X\ge 0}\|Y - AX\|^2$ to pursue the structural image. A simulated thorax phantom with 128×128 grids and 0.5 mm pixels, as shown in Figure 2, is used in the following experiments. There are many advantages to using simulated phantoms, including prior knowledge of the pixel values and the ability to control noise. The total attenuation value is approximately 5×10^7. For this case, the proposed algorithm is called ISRT, which is widely used in the field of CT reconstruction [4].

The system matrix is obtained using the "angle of view" method [10]. From the system matrix, we forward project the phantom on the sinogram with 128 radial bins (bin size of 0.5 mm) and 128 angular views evenly spaced over π. The noisy projections are obtained by the Poisson-random formula $y_i = \mathrm{Poisson}(y_i^*)$, where y_i^* is the noise-free projection.

Figure 3 shows the reconstructions with 50 iterations for every algorithm. As shown, the CG's image suffers from serious noise artifacts. The PL algorithm provides a reconstructed image that is slightly blurred compared with that of ISRT. The proposed ISRT provides the sharpest edge and clearest interior region.

Figure 4 quantitatively compares the histograms of the reconstructed images. The CG's result still shows a clear turbulence around the true value. PL and ISRT provide two comparable curves, and neither is significantly better than the other.

We also present a comparison of the MSE curves and χ with increasing iteration number in Figure 5. As shown, CG exhibits an unsteady reconstruction property in which more iterations may lead to a worse result. Both PL and ISRT overcome this shortcoming so that more iterations always provide considerably better results. In addition, ISRT will provide a more rapid convergence than PL.

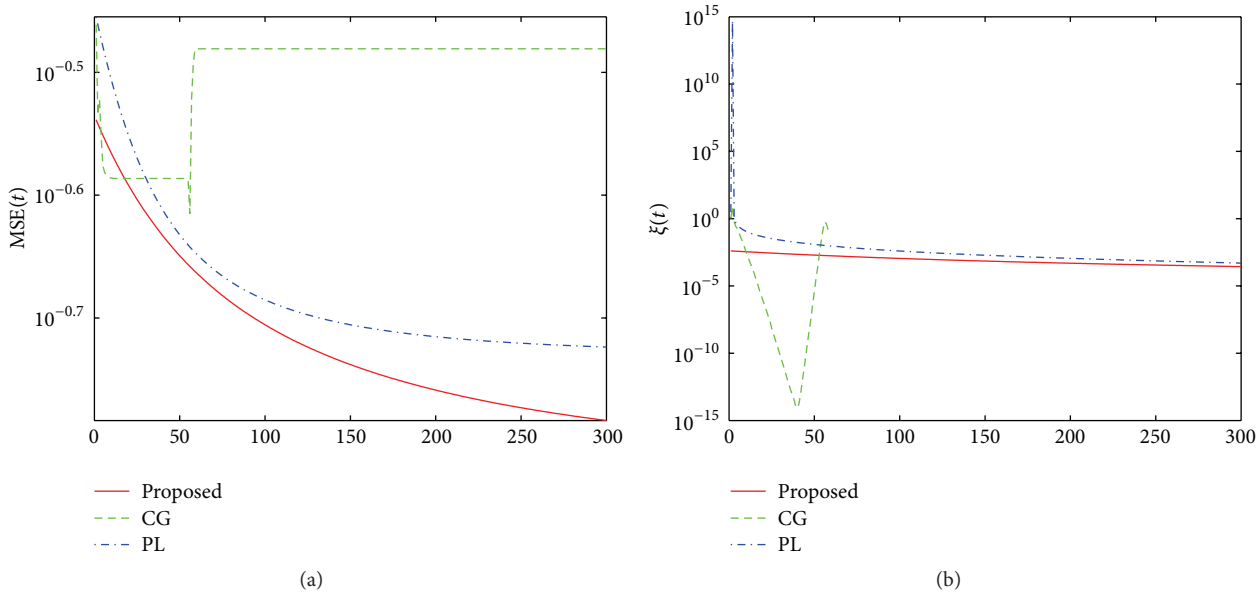

Figure 1: Simulated data: MSE and χ versus iteration number: (a) MSE and (b) χ.

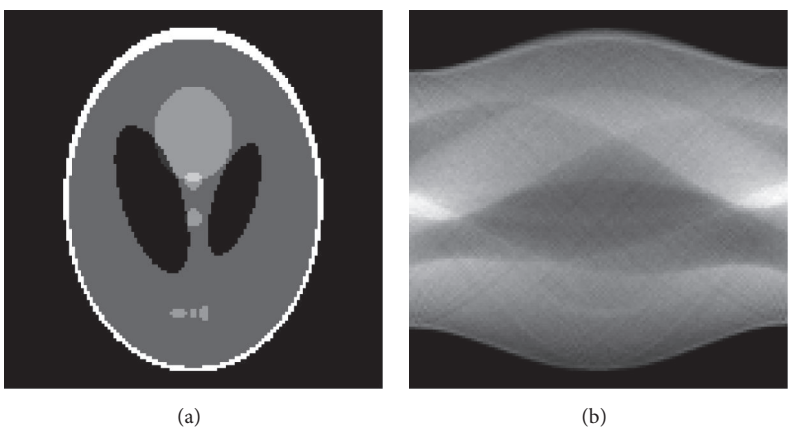

Figure 2: Shepp-Logan phantom with 128×128 grids: (a) true phantom and (b) noisy projections.

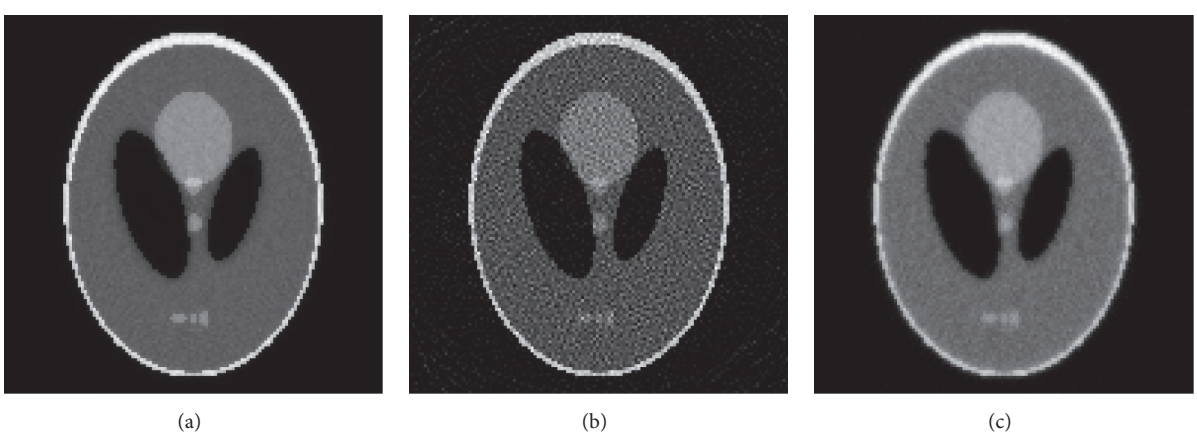

Figure 3: Reconstructed images with 50 iterations by (a) the proposed ISRT, (b) CG, and (c) PL.

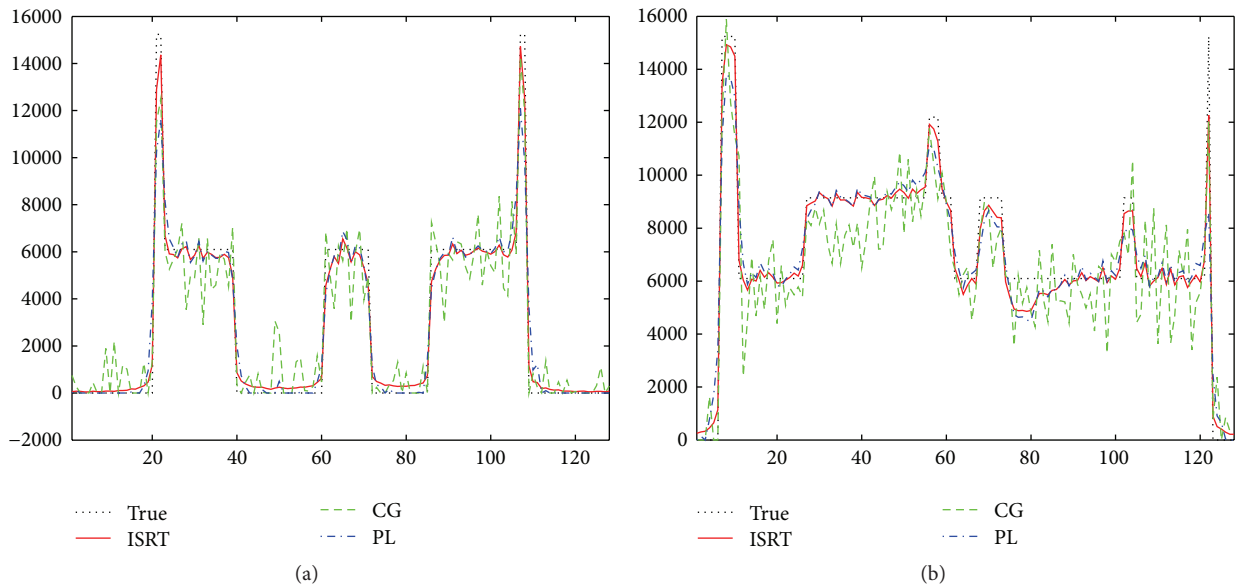

Figure 4: Intensity of the true and reconstructed images: (a) horizontal central line and (b) vertical central line.

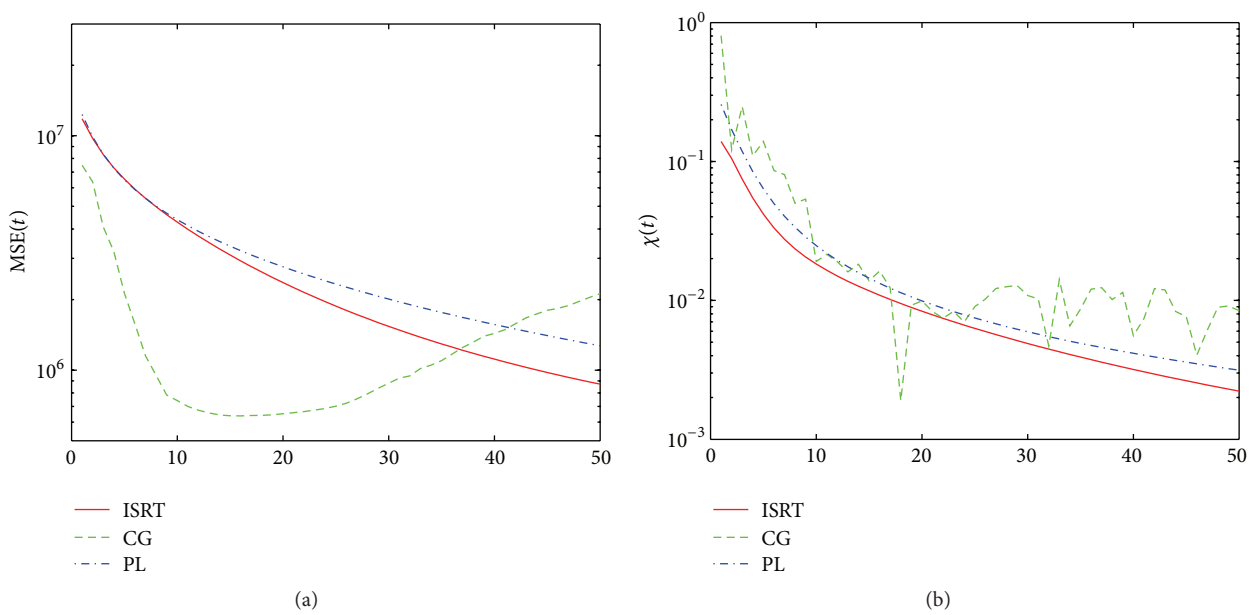

Figure 5: MSE and χ versus iteration number: (a) MSE and (b) χ.

6. Conclusion

We present a special application of a surrogate-based method for solving box-constrained LS problems. A new update rule is developed to iteratively update variables, and this rule exhibits desirable properties, including monotonic decrease of the cost function, self-constraining in the feasible region, and no need to impose a step size. This algorithm covers many existing algorithms, as well as new ones, as the special examples. Targeting the special case of only below-bounded constraints, we provide a rigorous theoretical global convergence proof. We use the simulated data to evaluate the performance of the algorithm, demonstrating that the proposed algorithm provides a stabler and faster convergence than the PL and CG approaches.

In this paper, we only provide a global convergence proof for the below-bounded case, not the box-constrained case; thus, proving the convergence for the latter case will be the focus of future work.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundations of China (61302013 and 61372014), Science and Technology Plan of Liaoning Province of China (2014305001), and Fundamental Research Funds for the Central Universities of China (N130219001 and N141008001).

References

[1] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proceedings of the Advances in Neural Information Processing Systems 13 (NIPS '00)*, pp. 556–562, Breckenridge, Colo, USA, 2000.

[2] Y. Lin and D. D. Lee, "Bayesian regularization and nonnegative deconvolution for room impulse response estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 839–847, 2006.

[3] O. L. Mangasarian and D. R. Musicant, "Lagrangian support vector machines," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 161–177, 2001.

[4] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 12, no. 2, pp. 328–333, 1993.

[5] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 13, no. 2, pp. 290–300, 1994.

[6] M. Piana and M. Bertero, "Projected Landweber method and preconditioning," *Inverse Problems*, vol. 13, no. 2, pp. 441–463, 1997.

[7] H. Lantéri, M. Roche, O. Cuevas, and C. Aime, "A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints," *Signal Processing*, vol. 81, no. 5, pp. 945–974, 2001.

[8] H. Lanteri, M. Roche, and C. Aime, "Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms," *Inverse Problems*, vol. 18, no. 5, pp. 1397–1419, 2002.

[9] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 1–20, 2000.

[10] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.

[11] A. Cichocki, H. Lee, Y.-D. Kim, and S. Choi, "Non-negative matrix factorization with α-divergence," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1433–1440, 2008.

[12] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*, Prentice-Hall, New York, NY, USA, 1969.

[13] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 14, no. 1, pp. 132–137, 1995.

[14] A. R. De Pierro, "On the convergence of an EM-type algorithm for penlaized likelihood estimation in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 14, pp. 762–765, 1995.

[15] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.

[16] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *Journal of Computer Assisted Tomography*, vol. 8, no. 2, pp. 306–316, 1984.

[17] Y. Teng, Y. Zhang, H. Li, and Y. Kang, "A convergent nonnegative deconvolution algorithm with Tikhonov regularization," *Inverse Problems*, vol. 31, no. 3, Article ID 035002, 2015.

[18] A. M. Ostrowski, *Solution of Equations in Euclidean and Banach Spaces*, Academic Press, New York, NY, USA, 1973.

[19] B. Scholkopf, C. Burges, and A. Smola, "Making large-scale SVM learning practical," in *Advances in Kernel Methods*, pp. 169–184, MIT Press, Cambridge, Mass, USA, 1998.

[20] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods Support Vector Learning*, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds., pp. 185–208, MIT Press, Cambridge, Mass, USA, 1999.

[21] C.-J. Hsieh, K.-W. Chang, C.-J. Lin, S. S. Keerthi, and S. Sundararajan, "A dual coordinate descent method for large-scale linear SVM," in *Proceedings of the 25th International Conference on Machine Learning (ICML '08)*, pp. 408–415, Helsinki, Finland, July 2008.

[22] H. Andrews and B. Hunt, *Digital Image Restoration*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1977.

[23] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, IOP Publishing, London, UK, 1998.

[24] S. Walrand, F. Jamar, and S. Pauwels, "Improved solution for ill-posed linear systems using a constrained optimization ruled by a penalty: evaluation in nuclear medicine tomography," *Inverse Problems*, vol. 25, no. 11, Article ID 115004, 2009.

[25] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[26] H. Liu, L. Yan, Y. Chang, H. Fang, and T. Zhang, "Spectral deconvolution and feature extraction with robust adaptive Tikhonov regularization," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 2, pp. 315–327, 2013.

[27] A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*, Kluwer Academic Publishers, Boston, Mass, USA, 1995.

[28] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, Winston & Sons, Washington, DC, USA, 1977.

[29] S. Boyd, N. Parikh, E. Chu, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[30] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Püschel, "D-ADMM: a communication-efficient distributed algorithm for separable optimization," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2718–2723, 2013.

[31] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopins University Press, Baltimore, Md, USA, 1996.

[32] http://web.stanford.edu/group/SOL/index.html.