A Splitting-based Iterative Method for Sparse Reconstruction

Liquan Kang, Ying Chen, Zefeng Yu, Heng Wu, Zijun Zheng, Shanzhou Niu*

School of Mathematics and Computer Sciences, Gannan Normal University, Ganzhou, 341000, China.

(Received: 11 December 2015; Accepted: 20 February 2016)

Abstract  In this paper, we study a $\ell_1$-norm regularized minimization method for sparse solution recovery in compressed sensing and X-ray CT image reconstruction. In the proposed method, an alternating minimization algorithm is employed to solve the involved $\ell_1$-norm regularized minimization problem. Under some suitable conditions, the proposed algorithm is shown to be globally convergent. Numerical results indicate that the presented method is effective and promising.

Keywords  Compressed sensing, Sparse reconstruction, $\ell_1$-norm regularized minimization, Variable splitting, Convergence

AMS 2010 subject classifications 68U10, 90C30

DOI: 10.19139/soic.v4i1.204

1. Introduction

In this paper, we mainly focus on the following $\ell_1$-norm regularized minimization problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2_2 + \lambda \|x\|_1,$$

(1)

where $A \in \mathbb{R}^{m \times n} (m \ll n)$ is a linear operator, and $b \in \mathbb{R}^m$ is an observation, parameter $\lambda > 0$ is used to trade off $\ell_2$ term and $\ell_1$ term for minimization. The above problem (1) is well-known as the unconstrained basis pursuit denoising problem.

Up to now, numerous approaches have been proposed and analyzed for solving the aforementioned problem (1) [7, 12, 23, 25]. Figueiredo et al presented a GPSR [12] method, in which the $\ell_1$ problem (1) was reformulated as a box-constrained quadratic program and was minimization by the Barzilai-Borwein gradient projection method [1] with an efficient nonmonotone line search. Iterative shrinkage-thresholding algorithm (ISTA) is one of the most widely studied methods for solving problem (1), which was first proposed for wavelet-based image deconvolution [13, 18], and then was applied to many other fields [10, 11, 21, 22]. Using operator splitting, Hale, Yin and Zhang derived the iterative shrinkage-thresholding fixed point continuation algorithm (FPC) [16, 17]. Another similar method is the sparse reconstruction algorithm SpaRSA, which was proposed by Wright, Nowak and Figueiredo in [24]. Recently, Beck and Teboulle proposed a FISTA algorithm with $O(1/k^2)$ convergence rate [4]. In [25], Yin, Osher, Goldfarb and Darbon proposed a Bregman iterative regularization method to solving the problem (1). Furthermore, a linearized Bregman method was proposed and analyzed subsequently in [7, 8, 26]. Using the spectral gradient projection method with an efficient Euclidean projection on $\ell_1$-norm ball, Friedlander and Van den Berg [27] proposed a SPGL1 method for problem (1). Becker, Bobin and Candès [3] proposed a NESTA algorithm based on the Nesterov’s smoothing technique to solve the problem (1). Moreover, an alternating direction
algorithms for \( \ell_1 \)-norm regularized minimization problem were proposed, which was derived from either the primal or the dual form of \( \ell_1 \)-norm problem [28].

In this paper, based on the variable splitting scheme [14, 20], we present the following \( \ell_1 \)-norm regularized minimization model:

\[
\min_{x, y \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|u\|_1,
\]  

s.t. \( \|x - u\|_2^2 \leq \varepsilon. \) 

It is clear that the constrained problem (2) will reduce to (1) when \( \varepsilon = 0 \). However, in general, the constrained problem (2) is a strict inequality constrained problem for the effect of noise. The parameter \( \varepsilon > 0 \) is the standard deviation of additive noise. The above constrained problem can be transformed into an unconstrained problem as follows:

\[
\min_{x, u \in \mathbb{R}^n} J(x, u) = \min_{x, u \in \mathbb{R}^n} \frac{1}{2\lambda_1} \|Ax - b\|_2^2 + \frac{1}{2\lambda_2} \|u - x\|_2^2 + \|u\|_1,
\]

where \( \lambda_1 \) and \( \lambda_2 \) are positive regularization parameters. In fact, this variable splitting method has been proposed in image restoration [15], where the authors proposed a fast TV minimization method for image restoration.

For the problem (3), as we can see that, if we fixed \( u \), it is a smooth quadratic optimization problem and if we fixed \( x \), the optimization problem on \( u \) has a closed form that it can be explicitly calculated by Iterative Shrinkage/Thresholding Algorithm.

The rest of this paper is organized as follows. In Section 2, the alternating minimization algorithm is developed to solve the proposed \( \ell_1 \)-norm minimization problem (3) and its global convergence of the iterative algorithm is established under some suitable conditions. In Section 3, numerical examples are presented to demonstrate the effectiveness of the proposed model. Finally, we have a conclusion section.

2. Optimization algorithm

In this section, we present a variable splitting method (VSM) to solve the problem (3). Starting from an initial guess \( u_0 \), this method yields a sequence

\[ u_0 \rightarrow x_1 \rightarrow u_1 \rightarrow x_2 \rightarrow u_2 \rightarrow \cdots \rightarrow x_k \rightarrow u_k \cdots \]

such that

\[
\begin{align*}
S_A(u_{k-1}) &:= x_k = \arg\min_x f(x) = \arg\min_x \frac{1}{2\lambda_1} \|Ax - b\|_2^2 + \frac{1}{2\lambda_2} \|u_{k-1} - x\|_2^2, \\
S_{\ell_1}(x_k) &:= u_k = \arg\min_u g(u) = \arg\min_u \frac{1}{2\lambda_2} \|u - x_k\|_2^2 + \|u\|_1.
\end{align*}
\]

Then, we can obtain that

\[ u_k = S_{\ell_1}(S_A(u_{k-1})). \]

Let \( T(\cdot) = S_{\ell_1}(S_A(\cdot)) \), then

\[ u_k = T(u_{k-1}). \]

2.1. Minimization with respect to \( x \)

The subproblem (4a) is a quadratic convex optimization problem, which is equivalent to solving a linear system:

\[
(A^T A + \frac{\lambda_1}{\lambda_2} I)x = A^T b + \frac{\lambda_1}{\lambda_2} u,
\]

where \((A^T A + \frac{\lambda_1}{\lambda_2} I)\) is a symmetric positive definite matrix.

In this paper, the conjugate gradient method was utilized to solve (4a). The iterative scheme is as follows:

\[ x_{k+1} = x_k + \alpha_k d_k, \]

where \( \alpha_k = \arg\min \alpha f(x_k + \alpha d_k) \), \( d_k \) is the conjugate gradient direction,

\[ d_k = \begin{cases} -\nabla f(x_k) & \text{if } k = 0, \\ -\nabla f(x_k) + \beta_k d_{k-1} & \text{otherwise}. \end{cases} \]

where \( \beta_k = \frac{\nabla f(x_k)^T y_{k-1}}{d_{k-1}^T y_{k-1}} \), \( y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1}) \).

**Remark:** The optimal step size can be determined by

\[ \alpha_k = \frac{\lambda_2 (A d_k)^T b + \lambda_1 d_k^T u_k - \lambda_2 (A d_k)^T A x_k - \lambda_1 d_k^T x_k}{\lambda_2 (A d_k)^T A d_k + \lambda_1 d_k^T d_k}. \]

### 2.2. Minimization with respect to \( u \)

Since \( u \) is separable in the subproblem (4b), each of its components can be independently obtained by the shrinkage operator in [9], i.e.,

\[ u_k^i = \max\{ x_k^i - \lambda_2, 0 \} \text{sgn}(x_k^i), \quad i = 1, 2, \ldots, n \quad (7) \]

where \( u_k^i \) is the \( i \)th component of the \( u_k \).

### 2.3. Convergence analysis

This section is devoted to analysis the convergence of the poposd algorithm for (3). First, let us give some definitions.

**Definition 1**

An operator \( P \) is called nonexpansive if, for any \( x_1, x_2 \in \mathbb{R}^N \), we have

\[ \| P(x_1) - P(x_2) \| \leq \| x_1 - x_2 \|. \]

If there exists some nonexpansive operator \( A \) and \( \kappa \in (0, 1) \) such that \( P = (1 - \kappa) I + \kappa A \), then \( P \) is called \( \kappa \)-averaged nonexpansive.

In the following, we give the Lemma 2.4 in [6].

**Lemma 1**

Let \( \varphi \) be convex and semicontinuous and \( \beta > 0 \). Suppose \( \hat{u} \) is defined as follows:

\[ \hat{u} = \arg\min_u \| u - x \|^2 + \beta \varphi(u). \]

Define \( S \) such that \( \hat{u} = S(x) \) for each \( x \). Then \( S \) is \( \frac{1}{\beta} \)-averaged nonexpansive.

It is known that \( \| u \|_1 \) is convex and semicontinuous function. By Lemma 1, we can derive that \( S_{\ell_1} \) is \( \frac{1}{\beta} \)-averaged nonexpansive.

**Lemma 2**

The operator \( T \) defined in (5) is nonexpansive.

**Proof:** Since \( A^T A + \frac{\lambda_1}{\lambda_2} I \) is symmetric positive definite and its eigenvalues are larger than \( \frac{\lambda_1}{\lambda_2} \), then for any \( u, v \)

\[ \| \frac{\lambda_2}{\lambda_1} (A^T A + \frac{\lambda_1}{\lambda_2} I)^{-1} (u - v) \|_2 \leq \| u - v \|_2. \]
Therefore,
\[
\|T(u) - T(v)\|_2 = \|S_{\ell_1}(S_A(u)) - S_{\ell_1}(S_A(v))\| \\
\leq \|S_A(u) - S_A(v)\|_2 \\
= \|(A^TA + \frac{\lambda_1}{\lambda_2}I)^{-1}(A^Tb + \frac{\lambda_1}{\lambda_2}u) - (A^TA + \frac{\lambda_1}{\lambda_2}I)^{-1}(A^Tb + \frac{\lambda_1}{\lambda_2}v)\|_2 \\
= \|\frac{\lambda_2}{\lambda_1}(A^TA + \frac{\lambda_1}{\lambda_2}I)^{-1}(u - v)\| \\
\leq \|u - v\|_2. \tag{8}
\]

The result follows. □

Definition 2
Let $C$ be a convex closed set in Banach space $X$, a mapping $F : C \to C$ is asymptotically regular if, for any $x \in C$,
\[
\lim_{n \to \infty} \|F^{n+1}(x) - F^n(x)\|_2 = 0.
\]

Lemma 3
Let \{u_k\} be generated by (5), then we have
\[
\lim_{k \to \infty} \|u_{k+1} - u_k\|_2 = 0.
\]

Proof: We consider the Taylor series expansion of $J(x, u_k)$ in the first variable, and we have
\[
J(x_k, u_k) = J(x_{k+1}, u_k) + (x_k - x_{k+1})^T \frac{\partial J}{\partial x}(x_{k+1}, u_k) \\
+ \frac{1}{2}(x_k - x_{k+1})^T \frac{\partial^2 J}{\partial x^2}(x_k - x_{k+1}). \tag{9}
\]
Since $x_{k+1} = \text{argmin}_x J(x, u_k)$, it holds that
\[
\frac{\partial J}{\partial x}(x_{k+1}, u_k) = 0,
\]
and
\[
\frac{\partial^2 J}{\partial x^2}(x_{k+1}, u_k) = A^TA + \frac{\lambda_1}{\lambda_2}I.
\]
As $(A^TA + \frac{\lambda_1}{\lambda_2}I)$ is symmetric positive definite and its eigenvalues are larger than $\frac{\lambda_1}{\lambda_2}$, then we can obtain that
\[
J(x_k, u_k) - J(x_{k+1}, u_k) = \frac{1}{2}(x_k - x_{k+1})^T (A^TA + \frac{\lambda_1}{\lambda_2}I)(x_k - x_{k+1}) \\
\geq \frac{\lambda_1}{\lambda_2} \|x_k - x_{k+1}\|^2. \tag{10}
\]
Since $u_{k+1} = \text{argmin}_x J(x_{k+1}, u)$, then $J(x_{k+1}, u_{k+1}) \leq J(x_{k+1}, u_k)$, therefore,
\[
J(x_k, u_k) - J(x_{k+1}, u_{k+1}) \geq J(x_k, u_k) - J(x_{k+1}, u_k) \\
\geq \frac{\lambda_1}{\lambda_2} \|x_k - x_{k+1}\|^2. \tag{11}
\]
Since $S_{\ell_1}$ is nonexpansive, by Lemma 1, we have that
\[
\|x_k - x_{k+1}\|^2 \geq \|S_{\ell_1}(x_k) - S_{\ell_1}(x_{k+1})\|_2 = \|u_k - u_{k+1}\|^2.
\]
Hence, we obtain that
\[ \mathcal{J}(x_k, u_k) - \mathcal{J}(x_{k+1}, u_{k+1}) \geq \frac{\lambda_1}{\lambda_2} \| u_k - u_{k+1} \|_2^2. \]
It follows that \( \sum_{k=0}^{\infty} \| u_k - u_{k+1} \|_2^2 \) is bounded, and follows that \( \lim_{k \to \infty} \| u_{k+1} - u_k \|_2 = 0 \).

Taking the Lemma 3 into account, we can prove that the operator \( \mathcal{T} \) is asymptotically regular.

**Lemma 4**

For any initial guess \( u_0 \), suppose \( u_k \) is generated by (5); then \( \mathcal{T} \) is asymptotically regular, i.e.,
\[
\lim_{k \to \infty} \| T^{k+1}(u_0) - T^k(u_0) \|_2 = \lim_{k \to \infty} \| T^k(u_1) - T^{k-1}(u) \|_2 = \lim_{k \to \infty} \| T(u_k) - T(u_{k-1}) \|_2 = \lim_{k \to \infty} \| u_{k+1} - u_k \|_2 = 0. \]

To proof the set of the minimum of \( \mathcal{J}(x, u) \) is nonempty, we shall show the coerciveness of \( \mathcal{J} \).

**Definition 3**

A function \( \phi : \mathbb{R}^N \to \mathbb{R} \) is proper over a set \( X \subset \mathbb{R}^N \) if \( \phi(x) < \infty \) at least one \( x \in X \) and \( \phi(x) > -\infty \) for all \( x \in X \). A function \( \phi : \mathbb{R}^N \to \mathbb{R} \) is coercive over a set \( X \subset \mathbb{R}^N \) if for every sequence \( \{ x_k \} \subset X \) such that \( \| x_k \| \to \infty \), we have \( \lim_{k \to \infty} \phi(x_k) \to \infty \).

**Lemma 5**

The function \( \mathcal{J}(x, u) \) defined in (3) is coercive.

**Proof:**
\[
\mathcal{J}(x, u) = \frac{1}{2\lambda_1} \| Ax - b \|_2^2 + \frac{1}{2\lambda_2} \| u - x \|_2^2 + \| u \|_1 \\
\geq \frac{1}{2\lambda_1} \| Ax - b \|_2^2 + \frac{1}{2\lambda_2} \| u - x \|_2^2 + \| u \|_2 \\
= \frac{1}{2} \left( \begin{pmatrix} \sqrt{\lambda_1} A & 0 \\ \sqrt{\lambda_2} I \\ 0 \end{pmatrix} \right) \begin{pmatrix} x \\ u \end{pmatrix} - \begin{pmatrix} 0 \\ \sqrt{\lambda_2} I \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} \begin{pmatrix} \lambda_1 \lambda_2 \end{pmatrix} + \| b \|_2^2 + \left( \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} \right) \begin{pmatrix} \lambda_1 \lambda_2 \end{pmatrix}.
\]
Using the matrix factorization, we note that
\[
\begin{pmatrix} \sqrt{\lambda_1} A & 0 \\ \sqrt{\lambda_2} I \\ 0 \end{pmatrix} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \lambda_1 \lambda_2 \end{pmatrix}.
\]
It is obvious that the above matrix is full rank as \( \text{Null}(A) \cap \text{Null}(I) = \emptyset \). In fact, \( \text{Null}(A) \cap \text{Null}(I) = \{ 0 \} \), and \( 0 \) is not taken into consideration in practice. Therefore, when \( \| \begin{pmatrix} x \\ u \end{pmatrix} \|_2 \to \infty \), we have \( \mathcal{J}(x, u) \to \infty \). The result follows.

**Lemma 6**

The the set of fixed points of \( \mathcal{T} \) is nonempty.

**Proof:** Since the objective function \( \mathcal{J} \) is coercive, the set of minimizers of \( \mathcal{J} \) is nonempty. Assume that \( (x^*, u^*) \) is a minimizer of \( \mathcal{J}(x, u) \), i.e.,
\[
\frac{\partial \mathcal{J}}{\partial x}(x^*, u^*) = 0, \quad \frac{\partial \mathcal{J}}{\partial x}(x^*, u^*) = 0.
\]
And
\[
\begin{align*}
x^* &= S_A(u^*) = \arg\min J(\cdot, u^*), \\
u^* &= S_{\ell_1}(x^*) = \arg\min J(x^*, \cdot).
\end{align*}
\]
Hence we get \(u^* = S_{\ell_1}(S_A(u^*)) = T(u^*),\) and \(u^*\) is a fixed point of \(T.\) The result follows.

We remark that the objective function \(J\) in (3) is strictly convex as \(A\) is a matrix of full column rank, Therefore, a fixed point of \(T\) is also a global minimizer of \(J.\) According to the Opial theorem [19], the sequence \(\{u_k\}\) converges to a fixed point of \(J,\) i.e., a minimizer of \(J.\)

Theorem 1
For any initial guess \(u^0 \in \mathbb{R}^n\), suppose \(\{u_k\}\) is generated by (5), and \(u_k\) converges to a stationary point of \(J.\)

3. Numerical experiments

In this section, we test the performance of the variable splitting method for the \(\ell_1\)-norm regularized minimization problem with application to computed tomography (CT) image reconstruction and compressed sensing.

All the programs were coded in MATLAB R2007a and run on a personal computer with an Intel Core 2 Duo CPU at 2.5 GHz and 2GB of memory.

3.1. Compressed sensing problem

In this subsection, the SVM algorithm was applied to solve the compressed sensing with comparison to FISTA method. In the numerical test, the original sparse signal \(x_s\) and the matrix \(A\) are randomly generated, and the observation \(b\) is generated according to \(b = Ax_s + \sigma \cdot \text{randn}(m,1)\), where \(\sigma\) is the standard deviation of additive Gaussian noise. Let \(k\) denote the number of nonzero components in \(x_s.\) The stopping criterion was adopted as follows:
\[
\frac{\|x_{k+1} - x_k\|}{\|x_{k+1}\|} < 5 \times 10^{-3},
\]
where \(\{x_k\}\) is the sequence generated by the test algorithms. And we measure the quality of recovered signal by its mean square error (MSE), i.e.,
\[
\text{MSE} = \frac{\|\widehat{x} - x^*\|}{n},
\]
where \(\widehat{x}\) is the signal restored by certain algorithm, \(x^*\) is the original signal, \(n\) is the length of the signal.

The value of the objective function (log10(fun)) and MSE are given in Table 1 and Table 2, respectively. Figure 1 and Figure 3 show the restored signal by FISTA algorithm and SVM algorithm, respectively. Figure 2 and Figure 4 show the objective function value and MSE vs the iteration numbers, respectively. The results indicate that the VSM algorithm outperforms FISTA algorithm.

<table>
<thead>
<tr>
<th>n=2048</th>
<th>FISTA</th>
<th>VSM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>iter</td>
<td>log10(fun)</td>
</tr>
<tr>
<td>0.3</td>
<td>232.000</td>
<td>2.515</td>
</tr>
<tr>
<td>0.3</td>
<td>195.000</td>
<td>2.963</td>
</tr>
<tr>
<td>0.2</td>
<td>260.000</td>
<td>2.225</td>
</tr>
<tr>
<td>0.2</td>
<td>203.000</td>
<td>2.703</td>
</tr>
<tr>
<td>0.1</td>
<td>278.000</td>
<td>1.785</td>
</tr>
<tr>
<td>0.1</td>
<td>207.000</td>
<td>2.197</td>
</tr>
</tbody>
</table>
Figure 1. The original signal (top), the measurement (second), and the reconstructed signals by FISTA (third) and VSM (bottom).

n=2048, m=410, k=82, sigma=0.01

Figure 2. Function value vs. iteration (left) and MSE vs. iteration (right) for FISTA and VSM.

Table 2. Comparison results in CS (sigma=0.05).

<table>
<thead>
<tr>
<th>n=2048</th>
<th>FISTA</th>
<th>VSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>i (m/n)</td>
<td>j (k/m)</td>
<td>iter</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>232.000</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>201.000</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>262.000</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>209.000</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>266.000</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>206.000</td>
</tr>
</tbody>
</table>
3.2. Sparse-view X-ray CT image reconstruction

To validate and evaluate the performance of the SVM method, in this subsection, a modified digital Shepp-Logan phantom (Figure 5) was used in the experiments. In this subsection, the total variation regularization was used in (3) for sparse-view X-ray CT image reconstruction, the (4b) is solved by Chambolle’s algorithm [5]. For the CT projection simulation, we chose a geometry that was representative of a monoenergetic fan-beam CT scanner setup. This geometry was modeled with 672 bins on a 1-D detector for 2-D image reconstruction, and projection data with 25 projection numbers of views was simulated at equal angular increment on 360 around the phantom. The imaging parameters of the CT scanner were as follows: (1) the distance from the detector arrays to the x-ray source was 1040 mm; (2) the distance from the rotation center to the x-ray source was 570 mm; (3) and the space...
of each detector bin was 1.407 mm. Each phantom is composed by $512 \times 512$ square pixels with each pixel size of $1.25 \text{ mm} \times 1.25 \text{ mm}$.

The results from the 25-view projection data in the noise-free and noisy cases are shown in Fig. 6. Serious noise-induced streak artifacts can be observed in the images reconstructed by the FBP method with ramp filter. The SVM algorithm can yield accurate reconstructions from spare-view projection data.
4. Conclusion

In this paper, we presented a splitting-based iterative method for sparse solution recovery in compressed sensing and X-ray CT image reconstruction. The numerical experiment results show that the proposed method is effective and promising.

5. Acknowledgement

This research was supported by the National Natural Science Foundation of China (No.61262026), Natural Science Foundation of Jiangxi Province (20132BAB201026), Science and technology program of Jiangxi Education Committee (LDJH12088) and the graduate student innovation project of jiangxi Province (YC2014-S411).

REFERENCES