A Fast Algorithm for Recovery of Jointly Sparse Vectors based on the Alternating Direction Methods

Hongto Lu (Presented by Heng Luo)
Department of Computer Science and Engineering
Shanghai Jiao Tong University
China
April 11, Ft. Lauderdale, FL, USA
The author (Hongtao Lu)

Cannot attend AISTATS2011 due to the Visa reason.
Ask his colleague Mr. Heng Luo to give the talk in place of him
Shanghai Jiao Tong University

One of the top research University in China
Outline

- Introduction
- Problem formulation
- Existing methods
- Our Algorithm
- Experiments
- Conclusion and future works
Compressive sensing (CS) has recently emerged as an active research area which aims to recover sparse signals from measurement data (Candes et al., 2006; Donoho, 2006a).


Introduction

Single Measurement Vector Model (SMV)

- The basic compressive sensing (CS) is a single measurement model, which aims to recover sparse signal from single measurement vector

\[ b = Ax \]

- Measurement vector
- Sparse signal to be recovered
- Dictionary
Multiple Measurement Vector Model (MMV)

- Recovery of sparse signals from multiple measurement vectors
- In this paper, we consider the recovery of jointly sparse signals in the MMV model where multiple signal measurements are represented as a matrix and the sparsity of signal occurs in common locations.

\[ B = AX \]

- Measurement vectors (matrix)
- Dictionary
- (Jointly) Sparse signals (matrix) to be recovered
MMV models arise in

- Magnetoencephalography (MEG)
- DNA microarrays
- sparse communication channels
- echo cancellation
- sparse solutions to linear inverse problems
- source localization in sensor networks
- …
Introduction

- Solving MMV model is much more difficult than solving SMV model.
- Current methods are slow and not scale up well.
- In this paper, we propose a fast algorithm for MMV model based on the alternating direction methods (ADM) which is much faster than the state-of-the-art method MMVprox.
Outline

- Introduction and motivation
- Problem formulation
- Existing methods
- Our Algorithm
- Experiments
- Conclusion and future works
Sparsity of the recovered signal in SMV model is achieved by minimizing the cardinality of the nonzero components, which is formulated as the following optimization problem:

\[(p0): \min_{x} \| x \|_0, \quad \text{s.t.} \quad Ax = b\]
The objective function of (p0) is not convex and the optimization is a combinatorial optimization, so is NP-hard. (p0) is thus relaxed to the following convex $L_1$ minimization problem

$$(p1): \min_{x} \|x\|_1, \quad \text{s.t.} \quad Ax = b$$
MMV

- In MMV model, joint sparsity is desired, which can be achieved by the following minimization problem

$$\begin{align*}
(P0): \min_X \|X\|_{2,0}, \quad \text{s.t.} \quad AX = B
\end{align*}$$

Matrix $L_{2,0}$ (quasi) norm, i.e., the number of nonzero rows of the matrix $X$
Similarly, (P0) is NP-hard and usually relaxed to the following (P1) problem:

\[ (P1): \min_X \|X\|_{2,1}, \quad \text{s.t.} \quad AX = B \]

The objective function is the matrix $L_{2,1}$ norm, i.e., the sum of row vector $L_2$ norms of $X$. 
Outline

- Introduction and motivation
- Problem formulation
- Existing methods
- Our Algorithm
- Experiments
- Conclusion and future works
Existing methods

Methods for MMV

• (2,1)-norm minimization of MMV is much more difficult than L1 minimization in SMV
• Simultaneous OMP (Greedy method) ¹
• Convex relaxation (via mathematical programming) ²
• FOCal Underdetermined System Solver (FOCUSS) ³


Existing methods

- All of these methods do not scale to problems of moderate size
- MMVprox—the state-of-the-art method by (Sun et al. 2009). MMVprox consider the following problem

\[
\min_X \frac{1}{2} \|X\|_{2,1}^2, \quad \text{s.t.} \quad AX = B
\]

Existing methods

First, derive the dual problem

\[
\max_Y \left\{ -\frac{1}{2} \| A^T Y \|_{2,\infty}^2 + \langle Y, B \rangle \right\}
\]

Then the prox-method developed by (Nemirovski, 2005) is applied to solve the dual.

The prox-method has almost dimension-independent convergence rate of O(1/t)

Much faster than former methods

Outline

- Introduction and motivation
- Problem formulation
- Existing methods
- Our Algorithm (MMV-ADM)
- Experiments
- Conclusion and future works
Our algorithm (MMV-ADM)

Motivations

- MMVprox is still slow from a perspective of SMV’s algorithms.
- Among the state-of-the-art algorithms for solving the \( L_1 \)-minimization problem in SMV, the well-known iterative shrinkage-thresholding algorithm (ISTA) has a worst-case convergence rate of \( O(1/t) \), the same as the prox-method.
- The fast iterative shrinkage-thresholding algorithms (FISTA) (Beck et al., 2009) has convergence rate of \( O(1/t^2) \).
- The alternating direction method (ADM, Yang 2009) is empirically faster than FISTA.
- So, we extend ADM to MMV model.
Our algorithm (MMV-ADM)

Our model, the following optimization problem

\[
\min_X \| X \|_{2,1}, \quad \text{s.t.} \quad \| AX - B \|_F < \delta
\]

This model is equivalent to the following unconstrained optimization problem, which is the model we focus in this paper

\[
\min_X \left\{ \| X \|_{2,1} + \frac{1}{2\mu} \| AX - B \|_F^2 \right\}
\]

- (2,1)-norm, measuring the joint sparsity of rows of X
- Frobenius norm, measuring the reconstruction error

The constraint is relaxed due to noise
Our algorithm (MMV-ADM)

- By introducing an auxiliary matrix variable $E$ to measure the residue between $AX$ and $B$, we have

\[
\min_{X, E} \left\{ \|X\|_{2,1} + \frac{1}{2\mu} \|E\|_F^2 \right\} \quad \text{s.t.} \quad AX + E = B
\]

- Its augmented Lagrangian function is

\[
L(X, E, Y) = \|X\|_{2,1} + \frac{1}{2\mu} \|E\|_F^2 + \langle Y, AX + E - B \rangle + \frac{\beta}{2} \|AX + E - B\|_F^2
\]

The penalty parameter

The Lagrangian multiplier

The matrix inner product defined as $\langle A, B \rangle = \text{tr}(A^\top B)$
Our algorithm (MMV-ADM)

- Given \((X^{(k)}, E^{(k)}, Y^{(k)})\), we derive update rules for 
  \((X^{(k+1)}, E^{(k+1)}, Y^{(k+1)})\) sequentially based on ADM.
- Firstly, given \((X^{(k)}, Y^{(k)})\), we update \(E^{(k+1)}\)
- By removing terms irrelative to \(E\) in 
  \[
  \min_{X,E} L(X, E, Y)
  \]
- We have the following objective function,

\[
\min_E \left\{ \frac{1}{2\mu} \| E \|^2_F - \langle Y^{(k)}, E + AX^{(k)} - B \rangle + \frac{\beta}{2} \| E + AX^{(k)} - B \|^2_F \right\}
\]
Our algorithm (MMV-ADM)

This is a quadratic optimization problem with respect to $E$, the minimizer is given by

$$E^{(k+1)} = \frac{\mu \beta}{1 + \mu \beta} \left[ \frac{1}{\beta} Y^{(k)} - (AX^{(k)} - B) \right]$$

1. Update rule for $E$
Our algorithm (MMV-ADM)

- Next, given \((X^{(k)}, E^{(k+1)}, Y^{(k)})\), update \(X\).
- By removing terms irrelative to \(X\), we have

\[
\min_X \left\{ \| X \|_{2,1} + \frac{\beta}{2} \| AX + E^{(k+1)} - B - \frac{1}{\beta} Y^{(k)} \|_F^2 \right\}
\]

- Approximate the second term by its Taylor expansion at \(x^{(k)}\) up to the second order, and after some algebra, we have

\[
\min_X \left\{ \| X \|_{2,1} + \beta \langle G^{(k)}, X - X^{(k)} \rangle \right. \\
\left. + \frac{\beta}{2\tau} \| X - X^{(k)} \|_F^2 \right\}
\]

\(G^{(k)}\) is gradient of the second term
Our algorithm (MMV-ADM)

- This is further equivalent to

\[
\min_X \left\{ \frac{\tau}{\beta} \|X\|_{2,1} + \frac{1}{2} \|X - (X^{(k)} - \tau G^{(k)})\|_F^2 \right\}
\]

- This is a special form of optimization enjoying closed form solution.

- Recall the famous Iterative Shrinkage Thresholding (IST) algorithm for the following optimization SMV

\[
x^* = \arg \min_x \left\{ \lambda \|x\|_1 + \frac{1}{2} \|x - c\|_2^2 \right\}
\]

\textbf{c is a constant vector}
Our algorithm (MMV-ADM)

- The closed form solution is

\[ x_i^* = \text{Shrink}(c_i, \lambda) \triangleq \begin{cases} 
  c_i - \lambda, & \text{if } c_i > \lambda \\
  c_i + \lambda, & \text{if } c_i < -\lambda \\
  0, & \text{otherwise}
\end{cases} \]

- Similarly, for the following matrix (2,1)-norm minimization problem

\[ X^* = \arg\min_X \{ \lambda \|X\|_{2,1} + \frac{1}{2} \|X - C\|_F^2 \} \]

\( C \) is a constant matrix
Our algorithm (MMV-ADM)

- The minimizer is given by

\[ X^* = \text{Row}_i \text{ Shrink}(C, \lambda) \]

- Where the function \( \text{Row}_i \text{ Shrink}(C, \lambda) \) is defined as

\[(x^*)^i = \begin{cases} 
\frac{\|c^i\|_2 - \lambda}{\|c^i\|_2} c^i, & \text{if } \|c^i\|_2 > \lambda \\
0, & \text{otherwise}
\end{cases} \]
Finally, the Lagrangian multiplier $Y$ is updated by

$$Y^{(k+1)} = Y^{(k)} - \gamma \beta (AX^{(k+1)} + E^{(k+1)} - B)$$

3. Update rule for $Y$
Our algorithm (MMV-ADM)

Integrating three update rules, our algorithm called MMV-ADM is as follows

<table>
<thead>
<tr>
<th>Algorithm 1 MMV-ADM algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> Sensing matrix $A$, multiple measurement data matrix $B$, parameters $\mu, \beta, \tau$, and $\gamma$.</td>
</tr>
<tr>
<td><strong>Initialization:</strong> Randomly initialize $X^{(0)}$, $Y^{(0)}$. $k = 0$;</td>
</tr>
<tr>
<td><strong>while</strong> (not converged)</td>
</tr>
<tr>
<td>1. $E^{(k+1)} = \frac{\mu \beta}{1 + \mu \beta} \left[ \frac{1}{\beta} Y^{(k)} - (AX^{(k)} - B) \right]$.</td>
</tr>
<tr>
<td>2. $G^{(k)} = A^T (AX^{(k)} + E^{(k+1)} - B - \frac{1}{\beta} Y^{(k)})$.</td>
</tr>
<tr>
<td>3. $X^{(k+1)} = \text{Row-Shrink} (X^{(k)} - \tau G^{(k)}, \frac{\tau}{\beta})$.</td>
</tr>
<tr>
<td>4. $Y^{(k+1)} = Y^{(k)} - \gamma \beta (AX^{(k+1)} + E^{(k+1)} - B)$.</td>
</tr>
<tr>
<td>5. $k = k + 1$;</td>
</tr>
<tr>
<td><strong>end</strong></td>
</tr>
<tr>
<td><strong>Output:</strong> The jointly sparse signal $X$, the Lagrangian multiplier $Y$ and the residue $E$.</td>
</tr>
</tbody>
</table>
Outline

- Introduction and motivation
- Problem formulation
- Existing methods
- Our Algorithm
- Experiments
- Conclusion and future works
Experiments

- We compare MMV-ADM with the state-of-the-arts method MMVprox.

Data
- Matrix A: randomly generate obeying the normal distribution \( N(0, 1) \) of mean 0 and variance 1.
- A \( d \) by \( n \) jointly sparse matrix \( \bar{X} \) with \( k \) nonzero rows, which is reserved as the ground truth for comparison with the recovered signal \( X \).
- We then generate measurement matrix by
  \[
  B_{m \times n} = A_{m \times d} \bar{X}_{d \times n} + E_{m \times n}
  \]
Experiments

- To measure the accuracy of the recovered signal, we use the relative error defined as

\[
\text{RelErr} = \frac{\|X - \bar{X}\|_F}{\|\bar{X}\|_F}
\]

- Throughout our experiments, the parameters are fixed as \( \beta = 0.2, \tau = 0.8, \gamma = 0.5 \)

- Stopping condition is

\[
\|X^{(k+1)} - X^{(k)}\|_F / \|X^{(k)}\|_F < \varepsilon
\]
Experiments

The average relative error versus the iteration numbers (noise free)

\[ m = 50 \]
\[ d = 100 \]
\[ n = 80 \]

Figure 1: Relative errors versus iteration numbers for signals free of noise.
Experiments

- The average relative error versus the iteration numbers (Gaussian noise with $\sigma = 0.001$)

Figure 2: Relative errors versus iteration numbers for noisy signals.
Experiments

Different combinations of parameters (m,k), other parameters are the same as Fig.2

Table 1: Performance of our MMV-ADM algorithm

<table>
<thead>
<tr>
<th>$d = 100$</th>
<th>MMV-ADM</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/d</td>
<td>k/m</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table 2: Performance of MMVprox algorithm

<table>
<thead>
<tr>
<th>$d = 100$</th>
<th>MMVprox</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/d</td>
<td>k/m</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Experiments

- The relative error versus the iteration numbers of the MMV-ADM for $\bar{X}$ with various number $k$ of nonzero rows.

Fig. 5.
Outline

- Introduction and motivation
- Problem formulation
- Existing methods
- Our Algorithm
- Experiments
- Conclusion and future works
In this paper, a fast algorithm for jointly sparse vector recovery in MMV model of compressive sensing is proposed. The proposed algorithm, MMV-ADM, is based on the alternating direction algorithm of the augmented Lagrangian multiplier method. The MMV-ADM alternately updates the signal, the multiplier and the residue. It is simple, easy to implement and much faster than the state-of-the-art method MMVprox.
Future Works

- The theoretical treatment of convergence of the MMV-ADM algorithm and its applications to real problems are the future research topics.

- For further discussion and comments, please contact Hongtao Lu:  htlu@sjtu.edu.cn