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

Introduction

- 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).

E. J. Candes, J. Romberg, and T. Tao (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 52(2): 489-509.

D. L. Donoho (2006a). Compressed sensing. IEEE Transactions on Information Theory, 52(4):1289-1306.

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

Dictionary

Sparse signal to be recovered

Introduction

- 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)

(Jointly) Sparse signals (matrix) to be recovered

Dictionary

Introduction

- 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.

SHANGHAI JIAO TONG UNIVERSITY

Outline

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

Problem formulation

- Sparsity of the recovered signal in SMV model is achieved by minimizing the cardinality of the nonezero components, which is formulated as the following optimization problem

$$(\mathrm{p0}): \quad \min_{x} \|x\|_0, \quad \text{s.t.} \quad Ax = b$$

L_0 (quasi) norm, i.e., the number of nonzero components of x

Problem formulation

- 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

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

L_1 norm, convex

Problem formulation

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

$$(\mathrm{P0}): \quad \min_{X} \|X\|_{2,0}, \quad \text{s.t.} \quad AX = B$$

Matrix $L_{2,0}$ (quasi) norm, i.e., the number of nonzero rows of the matrix X

Problem formulation

- Similarly, (P0) is NP-hard and usually relaxed to the following (P1) problem

$$(\mathrm{P1}): \quad \min_{X} \|X\|_{2,1}, \quad \text{s.t.} \quad AX = B$$

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

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

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]

1. J.A. Tropp, A.C. Gilbert, and M.J. Strauss. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

2. J.A. Tropp. Algorithms for simultaneous sparse approximation. Part II: Convex relaxation. *Signal Processing*, 86(3):589–602, 2006.

3. S.F. Cotter, B.D. Rao, Kjersti Engan, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing*, 53(7):2477–2488, 2005.

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

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

Primal problem

Replace the (2,1)-norm in (P1) by its square

L. Sun, J. Liu, J. Chen, J. Ye (2009). Efficient Recovery of Jointly Sparse Vectors. In *Adv. NIPS*.

Existing methods

Dual problem

- First, derive the dual problem

$$\max_{Y} \{ -\frac{1}{2}\|A^T Y\|_{2,\infty}^2 + < Y, B > \}$$

- 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

A. Nemirovski (2005). Prox-method with rate of convergence O(1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM Journal on Optimization, 15(1):229-251.

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

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

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

The constraint is relaxed due to noise

- 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

Our algorithm (MMV-ADM)

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

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

- Its augmented Lagrangian function is

$$L(X,E,Y) = \|X\|_{2,1} + \frac{1}{2\mu}\|E\|_F^2$$

$$- < Y, AX + E - B > + \frac{\beta}{2}\|AX + E - B\|_F^2$$

The penalty parameter

The Lagrangian multiplier

The matrix inner product defined as $<A,B>=tr(A^TB)$

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,

$$\min_{E}\left\{\frac{1}{2\mu}\|E\|_F^2 - < Y^{(k)}, E + AX^{(k)} - B > + \frac{\beta}{2}\|E + AX^{(k)} - B\|_F^2\right\}$$

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

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

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

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

$G^{(k)}$ 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

$$x^* = \arg\min_{x} \{ \lambda \|x\|_1 + \frac{1}{2} \|x - c\|_2^2 \}$$

c is a constant vector

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Our algorithm (MMV-ADM)

- The closed form solution is

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

- 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

$$X^* = \mathrm{Row\_Shrink}(C, \lambda)$$

- Where the function $\mathrm{Row\_Shrink}(C, \lambda)$ is defined as

$$(x^*)^i = \begin{cases} \frac{\|c^i\|_2 - \lambda}{\|c^i\|_2} c^i, & \text{if} \quad \|c^i\|_2 > \lambda \\ 0, & \text{otherwise} \end{cases}$$

The *i*th row of X^*

The *i*th row of C

Our algorithm (MMV-ADM)

- 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

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Our algorithm (MMV-ADM)

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

Algorithm 1 MMV-ADM algorithm

Input: Sensing matrix A, multiple measurement data matrix B, parameters μ,β,τ, and γ.

Initialization: Randomly initialize $X^{(0)}$, $Y^{(0)}$.

$k = 0$;

while (*not converged*)

1. $E^{(k+1)} = \frac{\mu\beta}{1+\mu\beta}\left[\frac{1}{\beta}Y^{(k)} - (AX^{(k)} - B)\right]$.
2. $G^{(k)} = A^T(AX^{(k)} + E^{(k+1)} - B - \frac{1}{\beta}Y^{(k)})$.
3. $X^{(k+1)} = \mathrm{Row\_Shrink}(X^{(k)} - \tau G^{(k)}, \frac{\tau}{\beta})$.
4. $Y^{(k+1)} = Y^{(k)} - \gamma\beta(AX^{(k+1)} + E^{(k+1)} - B)$
5. $k = k + 1$;

end

Output: The jointly sparse signal X, the Lagrangian multiplier Y and the residue E.

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

$$B_{m\times n} = A_{m\times d}\bar{X}_{d\times n} + E_{m\times n}$$

Gaussian noise of zero mean

Experiments

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

$$\mathrm{RelErr} = \frac{\|X - \bar{X}\|_F}{\|\bar{X}\|_F}$$

- 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)

$$m = 50$$

$$d = 100$$

$$n = 80$$

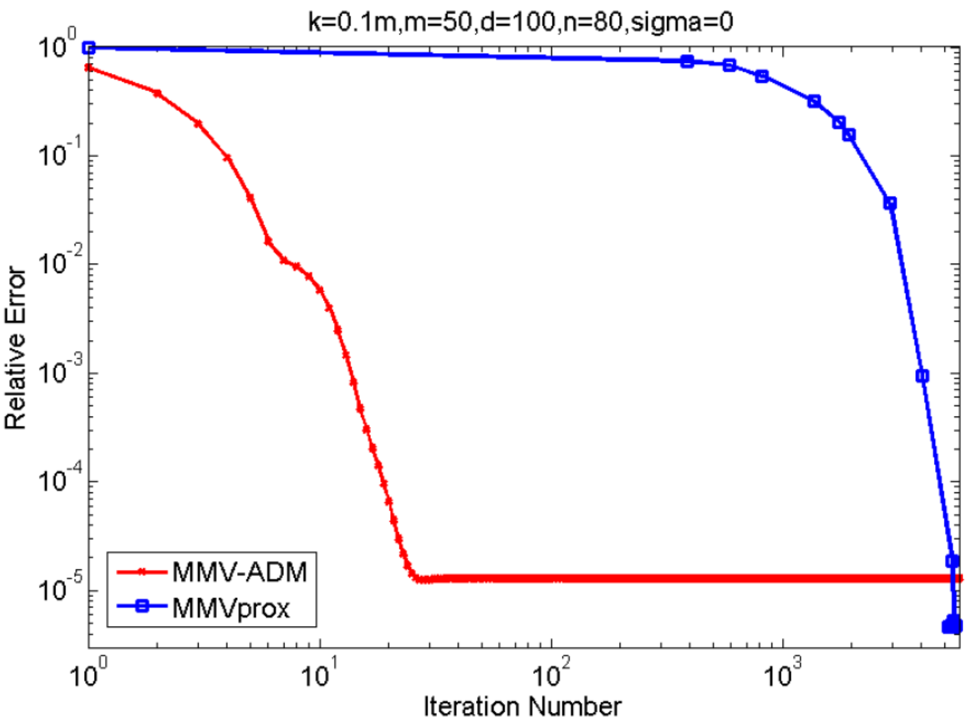


Figure 1: Relative errors versus iteration numbers for signals free of noise.

Experiments

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

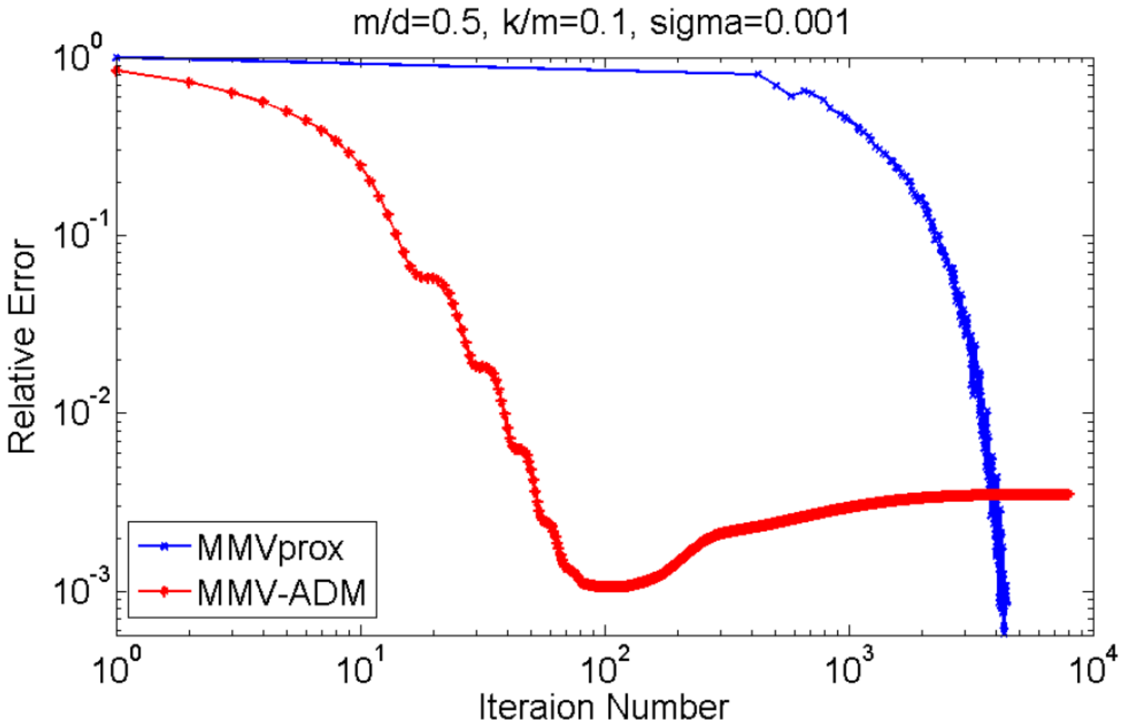


Figure 2: Relative errors versus iteration numbers for noisy signals.

Experiments

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

Table 1: Performance of our MMV-ADM algorithm

$d = 100$		**MMV-ADM**		
m/d	k/m	Iter	RelErr	AvTi
0.5	0.1	48.7	0.00556	0.0939
0.5	0.2	62.1	0.00631	0.1242
0.4	0.1	49.1	0.00602	0.0863
0.4	0.2	63.9	0.00637	0.1168
0.3	0.1	47.4	0.00768	0.0840
0.3	0.2	65.8	0.00931	0.1188

Average CPU time (in second)

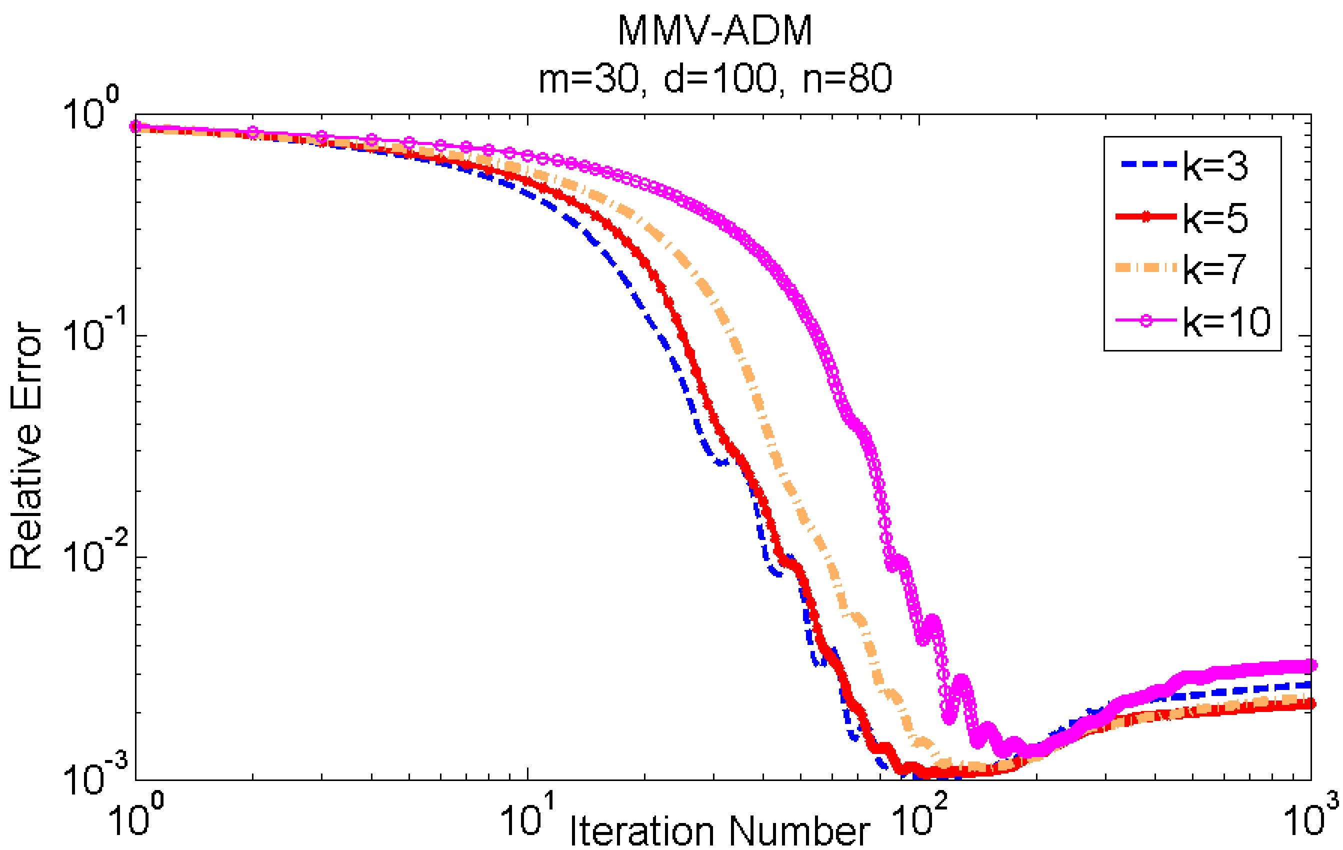
Table 2: Performance of MMVprox algorithm

$d = 100$		**MMVprox**		
m/d	k/m	Iter	RelErr	AvTi
0.5	0.1	4072	0.0015	3.7832
0.5	0.2	4147.8	0.0018	3.8291
0.4	0.1	4179.6	0.0047	3.4934
0.4	0.2	4132.1	0.0055	3.4767
0.3	0.1	5686.3	0.0212	4.2048
0.3	0.2	5445.2	0.0062	3.9961

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

Conclusion

- 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