FOCUSS ALGORITHM FOR RANK AWARE ROW-SPARSE MMV RECOVERY

A. Majumdar, R. K. Ward and T. Aboulnasr

Department of Electrical and Computer Engineering, University of British Columbia
angshulm@ece.ubc.ca, rababw@ece.ubc.ca and taboulnasr@apsc.ubc.ca

ABSTRACT

In this paper we propose a non-convex method for rank aware row-sparse Multiple Measurement Vector (MMV) recovery. Recent studies in row-sparse MMV recovery observed that better results can be achieved when the recovery algorithm takes into account the fact that the MMV matrix to be recovered is low-rank. The proposed non-convex problem requires minimizing the sum of $l_{2,p}$-norm and Schatten-q norm subject to data constraints. We derive an algorithm to solve the said problem based on the FOCally Under-determined System Solver (FOCUSS) approach. We compare our proposed method with state-of-the-art methods in rank aware and rank blind row-sparse MMV recovery. Our method always yields the best results in terms of probability of recovery.

Index Terms— sparse recovery, low-rank matrix recovery, non-convex algorithm

1. INTRODUCTION

In this paper, we propose an algorithm for row-sparse Multiple Measurement Vector (MMV) recovery taking into account the fact that the MMV matrix to be recovered is low-rank. Such problems arise in a variety of biomedical engineering problems like Magnetic Resonance Imaging [1] and EEG signal reconstruction [2]. Until recently most theoretical work on MMV recovery have only exploited the row-sparsity of the MMV matrix. But in recent papers [3, 4] it was shown that better recovery can be achieved when the low-rank property is also considered.

Generally in Compressed Sensing (CS), we are interested in solving an under-determined system of linear equations where the solution (x) is known to be sparse.

$$y_{r\times 1} = A_{r\times n} x_{n\times 1},\ r < n \tag{1}$$

Here, y is the observation vector, A is the measurement matrix and x is the unknown vector to be solved.

In the row-sparse MMV recovery, the problem is the following,

$$Y_{r\times N} = A_{r\times n} X_{n\times N} \tag{2}$$

where $Y_{r\times N} = \left[y_1 | \ldots | y_N\right]$ and $X_{n\times M} = \left[x_1 | \ldots | x_N\right]$.

In (2) all the x_i's vectors have the same support, i.e., they have non-zero elements at the same locations. Therefore, the MMV matrix X is row-sparse. But at the same time, since only a few of the rows in X are non-zeroes, the matrix is low-rank as well.

Previous studies in solving (2) have only used the row-sparsity information of X [5-9]. Their solution was based on the following optimization,

$$\min_X \|X\|_{m,p}^p \text{ subject to } Y = AX \tag{3}$$

where $\|X\|_{m,p}^p = \sum_{j=1}^{n} \|X^{j\rightarrow}\|_m^p$ ($X^{j\rightarrow}$ is the j^{th} row of X).

In [6], the values $m=2$ and $p\leq 1$ were proposed. Since values of p<1, make the problem non-convex, $m=2$ and $p=1$ are used in [5]. The choice of such values for the norms can be understood intuitively. The l_2-norm over every row ($X^{j\rightarrow}$) enforces non-zero values on all elements of the row vector whereas the summation over the l_2-norm ($\sum_{j=1}^{n} \|X^{j\rightarrow}\|_2$) of the rows enforces row-sparsity, i.e. the selection of few rows.

There are several algorithms to solve (3) [5-9]. These algorithms only rely on the group-sparsity of X and do not consider the fact that the solution X is low-rank as well. Recent studies [3, 4] showed that, better results can be obtained when the low-rank property is used along with the row-sparsity constraints while solving (2). In [3], a greedy algorithm called Order Recursive Matching Pursuit (ORMP) was proposed to solve (2); in [4] an optimization based algorithm was suggested based on the following optimization,

$$\min_X \|X\|_{2,1} + \eta \|X\|_* \text{ subject to } Y = AX \tag{4}$$

where $\|X\|_*$ denotes the convex nuclear norm of the matrix X and is defined as the sum of its singular values. Instead of minimizing the rank of a matrix, the convex nuclear norm is used in (4), because minimizing the rank of matrix is an NP hard problem [10, 11]. This is similar to CS where the convex l_1-norm is used instead of the NP hard l_0-norm (since the l_1-norm is the closest convex surrogate to the l_0-norm).

Studies in sparse recovery [12, 13] and low-rank matrix recovery [14-16] have shown that, when non-convex surrogates of the NP hard l_0-norm and rank of matrix are employed instead of their convex counterparts (i.e. l_p-norm instead of l_1-norm for sparsity and Schatten-q norm instead of nuclear norm for rank-deficiency, 0<p,q≤1), better reconstruction results can be achieved.

In this work, we address the problem of recovering row-sparse MMV matrix in a rank-aware fashion. Our approach however is motivated by findings in non-convex sparse recovery and non-convex low-rank matrix recovery. Therefore instead of (4), we propose to solve the following optimization problem,

$$\min_X \|X\|_{2,p}^p + \eta \|X\|_{S_q}^q \text{ subject to } Y = AX \tag{8}$$

where $\|X\|_{S_q}$ denotes the Schatten-q norm of the matrix.

To solve (8) we will employ the FOCally Under-determined System Solver (FOCUSS) approach. Previous this approach has been successfully used for row-sparse MMV recovery [6] and general low-rank matrix recovery [15]. In the next section, we derive the algorithm for rank-aware row-sparse MMV recovery. Section 3, will present the results for experimental evaluation. The conclusions of this work are discussed in section 4.

2. DERIVATION OF ALGORITHM

The problem is to solve (8). As mentioned before, we will use the FOCUSS approach to solve it. This approach has been successfully used to solve problems of row-sparse MMV recovery [6] and general low-rank matrix recovery [15] but has not been attempted on problems of rank aware row-sparse MMV recovery (8). The rank aware row-sparse MMV problem is however studied in [3, 4]. An algorithm is proposed in [4] to solve this problem using a convex formulation (4). This work proposes to solve (8) with a non-convex formulation.

To solve (8), the unconstrained Lagrangian expression is considered where we replace the Schatten-q norm by the equivalent Ky-Fan norm,

$$L(X,\lambda) = \|X\|_{2,p}^p + \eta Tr(X^T X)^{q/2} + \lambda^T (Y - AX) \tag{9}$$

where λ is the vector of Lagrangian multipliers.

The Karush-Kuhn-Tucker (KKT) conditions for an extremum to exist are the following normal equations,

$$\nabla_X L(X,\lambda) = p\left|\left\|X^{j\rightarrow}\right\|_2\right|^{p/2-1} \cdot X + \frac{1}{2}\eta q (XX^T)^{\frac{q}{2}-1} X + A^T\lambda = 0 \tag{10a}$$

$$\nabla_\lambda L(X,\lambda) = AX - y = 0 \tag{10b}$$

In (10a) $\left|\left\|X^{j\rightarrow}\right\|_2\right|^{p/2-1} \cdot X$ implies that the j^{th} row of X ($X^{j\rightarrow}$) is multiplied by $\left\|X^{j\rightarrow}\right\|_2^{p/2-1}$. In a more compact manner (10a) can be expressed as,

$$DX + A^T\lambda = 0 \tag{11}$$

where $D = pDiag(\left\|X^{j\rightarrow}\right\|_2^{p/2-1}) + \frac{1}{2}\eta q (XX^T)^{q/2-1}$; the *Diag* operation means that the values $\left\|X^{j\rightarrow}\right\|_2^{p/2-1}$ are placed as elements of a diagonal matrix.

Solving, for X in (11) we get,

$$X = -D^{-1}A^T\lambda \tag{12}$$

D is a block diagonal matrix with positive semi-definite blocks along the diagonal. Since D is positive semi-definite, the solution is not numerically stable. Such a problem was encountered while using FOCUSS for sparse signal recovery in Compressed Sensing [13]. To reach a stable solution, D must be positive definite. Following previous studies in l_p-norm minimization [13] and Schatten-p norm minimization [16], we ensure D is positive definite by adding a small term along the diagonal, i.e. we replace,

$$D \rightarrow D + \varepsilon I$$

Here ε is a small constant that ensures D is positive definite. This also guarantees that its inverse is positive definite.

Solving λ from (10b) and (11) we get,

$$\lambda = -(AD^{-1}A^T)^{-1}Y \tag{13}$$

Substituting the value of λ back in (16), we get,

$$X = D^{-1}A^T(AD^{-1}A^T)^{-1}Y \tag{14}$$

In order to efficiently compute X in each iteration, we re-write (14) as,

$$X = R\tilde{X}, \text{ where } \tilde{X} = (AR)^T((AR)(AR)^T)^{-1}Y \tag{15}$$

Here R is the Cholesky decomposition of D^{-1}. The decomposition exists since D^{-1} is a positive definite matrix. The reason, we expressed (14) in the form (15) is because $\tilde{X}$ can be solved very efficiently using the LSQR algorithm [17]. The problem (8) is solved iteratively. In each iteration k, the matrix D_k is computed based on the value of X from the previous value (X_{k-1}); the value of X_k is updated by solving the least squares problem. The algorithm is concisely expressed as follows:

Intitialize: $X_0 = A^T(AA^T)^{-1}Y$ which is a least squares solution; define ε
Repeat until stopping criterion is met:
Compute:
$D_k = pDiag(\left\|X_{k-1}^{j\rightarrow}\right\|_2^{p/2-1}) + \eta q (X_{k-1}X_{k-1}^T)^{q/2-1} + \varepsilon I$ and
$R_k R_k^T = D_k^{-1}$.
Update: $\tilde{X}_k = (AR_k)^T((AR_k)(AR_k)^T)^{-1}y$ and $X_k = R\tilde{X}_k$.
Decrease: $\varepsilon = \varepsilon/10$ iff $\|X_k - X_{k-1}\|_2 \le tol$

There are two stopping criteria. The first one is a limit on the maximum number of iterations. The second is the change in the value of the objective function in subsequent iterations; if the change is nominal, the iterations stop; it assumes that the solution has reached a local minimum with a tolerance (*tol*). The update step of the algorithm is solved by LSQR. The LSQR method runs for 20 iterations. The value of ε is initialized as 1. The tolerance level for deciding the decrease of ε is fixed at 10^{-3}.

2.1. Iterative Re-weighted Least Squares

In the past the Iterative Reweighted Least Squares (IRLS) approach was used for recovering sparse signals from under-sampled measurements [11]. This approach is closely related to the FOCUSS. Similarly a method for low-rank matrix recovery was developed using the IRLS method [14]. In this sub-section, we will discuss how a solution for (8) can be alternatively derived based the IRLS approach.

The focus of this paper is on solving the combined $l_{2,p}$-norm and Schatten-q norm minimization problem (8).

$$\min_X \|X\|_{2,p}^p + \eta\|X\|_{S_q}^q \text{ subject to } Y = AX$$

Above, we have derived a FOCUSS based algorithm solve it. In this sub-section, we will propose an alternate solution based on IRLS. It can be shown that $\|X\|_{2,p}^p = \|W_1 x\|_2^2$ and $\|X\|_{S_q}^q = \|W_2 x\|_2^2$ where, $W_1 = Diag(\|X^{j\rightarrow}\|_2^{1-p/2}) + \varepsilon I$ and $W_2 = (XX^T)^{1-q/2} + \varepsilon I$. Using these substitutions, (8) is expressed as,

$$\min_X \|W_1 X\|_2^2 + \eta\|W_2 X\|_2^2 \text{ subject to } Y = AX \qquad (16)$$

Alternatively,

$$\min_X \left\| \begin{pmatrix} W_1 \\ \sqrt{\eta} W_2 \end{pmatrix} x \right\|_2^2 \text{ subject to } Y = AX \qquad (17)$$

Here $\begin{pmatrix} W_1 \\ \sqrt{\eta} W_2 \end{pmatrix}$ implies that the matrices are vertically concatenated.

This has a closed form solution. However, it can be solved more efficiently using Conjugate Gradient. The values of W_1 and W_2 are updated at each iteration based on the values of X from the previous iteration. The IRLS algorithm can be concisely represented as following:

Initialize: $X_0 = \min_X \|Y - AX\|_2^2$

At iteration k (repeat until convergence)

$$W_1 = Diag(\|X_k^{j\rightarrow}\|_2^{1-p/2}) + \varepsilon I \text{ and } W_2 = (X_k X_k^T)^{1-q/2} + \varepsilon I$$

$$X_k = \min_X \left\| \begin{pmatrix} W_1 \\ \sqrt{\eta} W_2 \end{pmatrix} x \right\|_2^2 \text{ subject to } Y = AX$$

$\varepsilon = \varepsilon / 10$ iff $\|X_k - X_{k-1}\|_2 \leq tol$

The IRLS based algorithm is similar to our previous FOCUSS based algorithm. The least squares problem is solved by LSQR. Also the stopping criteria and the thresholds are the same.

3. EXPERIMENTAL RESULTS

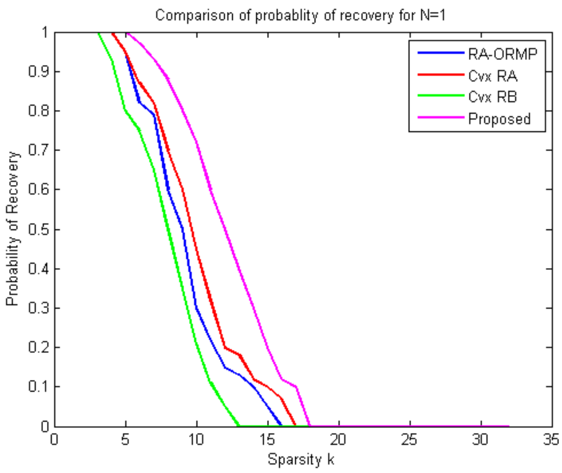
The experimental methodology followed here is based on [1, 3, 6]. The columns of matrix A are drawn from a Normal distribution and are normalized to have unit norms. The size of A is 32 X 256. Four sets of experiments are carried out; the number of measurements vectors in X (N) are 1, 4, 8 and 16 for the four sets. The number of non-zero rows in X is varied between 1 and 32 in steps of 1. The non-zero entries in X are drawn from a Normal distribution.

We compare our proposed non-convex rank aware (non-cvx RA) method with three existing methods:

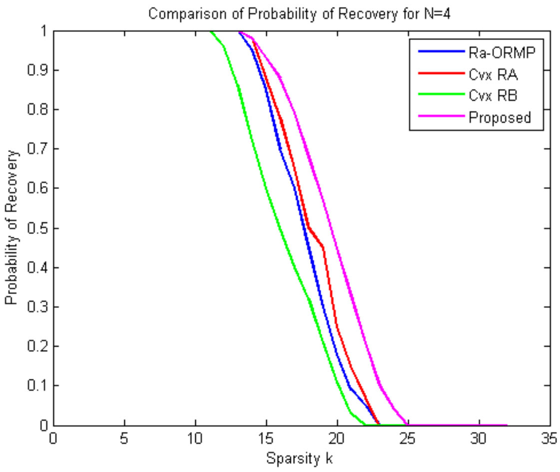
1. Rank Aware Order Recursive Matching Pursuit (RA-ORMP) [3].
2. Convex formulation for rank-aware row-sparse MMV recovery (4) (Cvx RA) [4].
3. Convex formulation for (rank-blind) row-sparse MMV recovery (3) (Cvx RB) [8].

Our method requires three values to be specified – the values p and q in the $l_{2,p}$-norm and Schatten-q norms and the value of η that controls the relative importance of the two norms. We found that the values $p=0.8$ and $q=0.8$ yield the best results. However, the variation in the results for values of p and q varying between 0.6 and 1 is only slight. The value of η is fixed at 0.1; this value was found to give good results always.

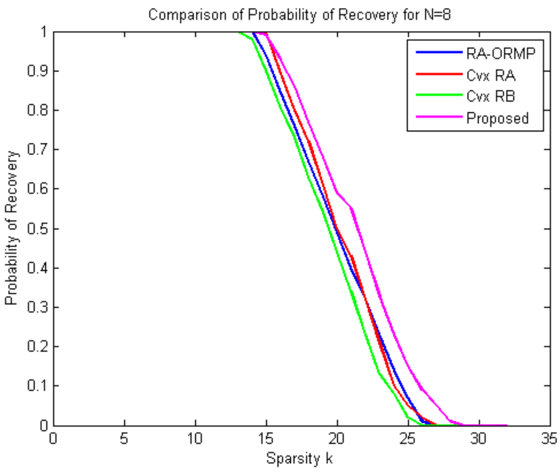
For the following graphs, a recovery is considered successful if the normalized mean squared error between the recovered signal and the ground-truth is less than a certain threshold (10^{-3} in this case). The probability of recovery is defined as the number of successful recoveries divided by the total number of trials for each configuration (1000 in this case).



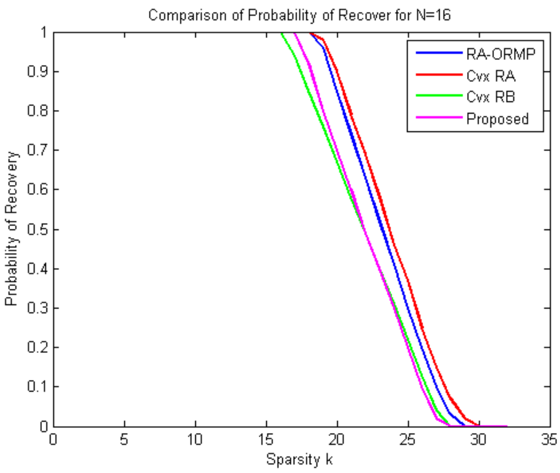
(a)



(b)



(c)



(d)

Fig. 1. Comparison of Probability of Recovery. (a) For one measurement vector, (b) for 4 measurement vectors, (c) for 8 measurement vectors and (d) for 16 measurement vectors.

Figure 1, gives the comparative probability of recoveries from various algorithm for different number of measurement vectors. The probability of recovery from RA-ORMP is almost the same as reported in [3]. The following conclusions can be drawn:

1. Our proposed non-convex rank aware recovery method always yields the best probability of recovery.
2. For all methods, the probability of recovery increases as the number of measurement vectors increases. This observation has been made in previous studies as well [3, 5, 8].

The improvement in probability of recovery is evident from Fig. 1. In Fig. 1a (N=1), at a sparsity of 10, the probability of recovery from our method is 0.78 whereas the probability of recovery from the convex formulation is 0.57 and from RA-ORMP is 0.32. Similar observations can be made from the other graphs as well (Fig. 1b-d).

4. CONCLUSION

In row-sparse MMV recovery, the unknown vectors to be reconstructed have a common support; therefore when they are stacked as columns of the MMV matrix, the resulting MMV matrix is row-sparse. Until recently, this problem was solved by taking into account the row-sparsity of the solution. However such an MMV matrix is low-rank as well. Recent studies [3, 4] showed that when the information about the MMV matrix's low rank property is exploited along with row-sparsity, the recovery results improve. In [3] a greedy algorithm called Rank aware Order Recursive Matching Pursuit (RA-ORMP) is proposed to solve the row-sparse MMV recovery problem in a rank aware fashion. In [4] a convex optimization based method is proposed for solving this problem. This method minimizes the sum of the $l_{2,1}$-norm (promoting row-sparsity) and the Nuclear norm (promoting rank deficiency) subject to the data constraints is.

Following studies in non-convex methods for sparse and low-rank recovery we here propose a non-convex optimization problem that minimizes the sum of thr $l_{2,p}$-norm and the Schatten-q norm subject to data constraints. We derive an algorithm to solve it based on the FOCUSS approach. We compare our method against state-of-the-art methods in rank aware and rank blind row-sparse MMV recovery. Experimental validation shows that our proposed method always yields better results.

REFERENCES

[1] A. Majumdar and R. K. Ward, "Calibration-less Multi-Coil MR Image Reconstruction", Magnetic Resonance Imaging, Vol. 30 (7), pp. 1032–1045, 2012.

[2] Z. Zhang, T. P. Jung, S. Makeig and B. D. Rao, "Compressed Sensing of EEG for Wireless Telemonitoring with Low

Energy Consumption and Inexpensive Hardware", arXiv:1206.3493v1.
[3] M. E. Davies and Y. C. Eldar, "Rank Awareness in Joint Sparse Recovery", IEEE Transactions on Information Theory, Vol.58 (2), pp.1135-1146, 2012.
[4] M. Golbabaee and P. Vandergheynst, "Guaranteed recovery of a low-rank and joint-sparse matrix from incomplete and noisy measurements", SPARS 2011.
[5] E. van den Berg and M. P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements", IEEE Trans. Info. Theory, Vol. 56 (5), pp. 2516-2527, 2010.
[6] S. F. Cotter, B. D. Rao, K. Engang, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors". IEEE Trans. Sig. Proc., Vol. 53 (7), pp. 2477-2488, 2005.
[7] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors", IEEE Trans. Sig. Proc., Vol. 56(10), pp. 4692-4702, 2008.
[8] A. Majumdar and R. K. Ward, "Synthesis and Analysis Prior Algorithms for Joint-Sparse Recovery", ICASSP, pp. 3421-3424, 2012.
[9] J. A. Tropp, "Algorithms for simultaneous sparse approximation: Part II: Convex relaxation", Signal Processing, Vol. 86, pp. 589,-602, 2006.
[10] B. Recht, W. Xu, and B. Hassibi, "Null Space Conditions and Thresholds for Rank Minimization", Mathematical Programming. Ser B. Vol. 127, Pages 175-211. 2011.
[11] E. J. Candès and T. Tao. "The power of convex relaxation: Near-optimal matrix completion", IEEE Trans. Inform. Theory 56(5), 2053-2080.
[12] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing", Inverse Problems, vol. 24, no. 035020, pp. 1--14, 2008.
[13] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing", ICASSP 2008.
[14] A. Majumdar and R. K. Ward, "Some Empirical Advances in Matrix Completion", Signal Processing, Vol. 91 (5), pp. 1334-1338, 2011.
[15] A. Majumdar, R. K. Ward and T. Aboulnasr, "A FOCUSS Based Method for Low Rank Matrix Recovery", IEEE International Conference on Image Processing, 2012 (accepted).
[16] K. Mohan, M. Fazel, Iterative Reweighted Least Squares for Matrix Rank Minimization, Proc. Allerton Conference on Communications, Control, and Computing, Sep 2010.
[17] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares", ACM Transactions on Mathematical Software, Vol. 8(1), pp. 43-71, 1982.