TUNING-FREE JOINT SPARSE RECOVERY VIA OPTIMIZATION TRANSFER

Evgenia Chunikhina, Gregory Gutshall, Raviv Raich, and Thinh Nguyen

Department of EECS, Oregon State University, Corvallis, Oregon 97331

{chunikhe,gutshalg,raich,thinhq}eecs.oregonstate.edu

ABSTRACT

Multiple measurement vector (MMV) problem addresses the recovery of a set of sparse vectors that have common sparsity pattern. In this paper, we consider a variant of the MMV problem where the common sparsity pattern is obfuscated by an additive noise. Specifically, we study the conditions for perfect reconstruction of the original sparsity pattern. Based on these, we develop a tuning-free algorithm for recovering jointly sparse solutions via the transfer optimization approach. We provide a preliminary numerical evaluation to illustrate our approach.

Index Terms— Sparse representation, joint sparsity, multiple-measurement vector (MMV), optimization transfer

1. INTRODUCTION

The problem of computing sparse solutions for linear inverse problems has received notable attention in the last years, especially in the signal processing community. Many different methods have been proposed to solve this problem [1] and [2]. In order to enhance the recoverability, additional information about the underlying solution structure, such as group sparsity, can be taken into account.

In our work, we consider a special case of the group sparsity structure, which is a problem of reconstruction of jointly sparse solutions, also known as the multiple measurement vector (MMV) problem. Jointly sparse solutions share the same nonzero support and appear in many applications, such as distributed compressive sensing, source localization, and magnetic resonance imaging. Our interest was motivated by the problem of deconvolving the threshold ionization energies of a measurement recorded from a Resonant Electron Capture-Time of Flight (REC-ToF) mass spectrometer. Theoretically, a fragment ion has distinct sparse ionization threshold energies. However, the measured ionization curve is convoluted by the electron energy distribution of the ionization source, which may be characterized through measurement.

Most notable among the methods that have been developed to solve the MMV problem are the forward sequential search-based method [3], ReMBo (which, reduces a MMV problem to a set of singular measurement vector problems) [4], a method based upon the alternating direction principle [5], along with many others. The variation of problems that allow the presence of noise was recently studied in [2], [6].

Our work is focused on finding the jointly sparse solution to the MMV problem whose the original sparsity pattern has been obfuscated by an additive noise. We study the conditions for perfect reconstruction of the original sparsity pattern. Based on these, we propose a tuning-free algorithm for recovering jointly sparse solutions via the optimization transfer approach. We provide a preliminary numerical evaluation of our approach.

2. PROBLEM FORMULATION

In this paper we define matrices by uppercase letters and vectors by lowercase letters. For a matrix X, x^i indicates its *i*-th row and x_i indicates its *j*-th column.

Let $||X||_{R^0}$ denote the number of rows of matrix X that have non-zero elements, i.e., $||X||_{R^0} = card\{i|||X^Te_i||_2 \neq 0\}$, where e_i the canonical vector satisfying $e_i(j) = 1$ for j = i and 0 otherwise. The multiple measurement vector (MMV) problem can be formulated as

$$\begin{array}{ll} \underset{X}{\text{minimize}} & \|X\|_{R^{0}} \\ \text{subject to} & \sum_{l=1}^{n} \|A_{l}x_{l} - y_{l}\|_{2}^{2} \leqslant \epsilon, \end{array}$$

$$(1)$$

where the mixing matrices $A_l \in \mathbb{R}^{m \times k}$ may be different $\forall l = \overline{1, n}$, and the solution vectors $x_l \in \mathbb{R}^{k \times 1}$ and measurement vectors $y_l \in \mathbb{R}^{m \times 1}$ are such that $y_l = A_l x_{0l} + \xi_l$, with vector $\xi_l = [\xi_{l1}, \xi_{l2}, \dots, \xi_{ln}]^T$ representing the additive noise $\forall l = \overline{1, n}$ and $X_0 \in \mathbb{R}^{k \times n}$ is the original s_0 row-sparse matrix we are interested in recovering.

Here, $n \ge 1$ is the number of measurement vectors. We assume $n \ll m$. Matrices $A_l \in \mathbb{R}^{m \times k}$ are known and obtained from the physics of the problem. Without loss of generality we assume that $rank(A_l) = m, m \ll k, \forall l = \overline{1, n}$.

The main aim is to obtain the solution matrix X such that each solution vector x_l is sparse and all solution vectors have a common sparsity pattern, i.e., the indices of the nonzero elements of x_l must be the same $\forall l = \overline{1, n}$.

This work was partially supported by the Air Force Office of Scientific Research grant number FA9550-09-1-0471 for R.R. and E.C. and by the Department of Analytical Chemistry and the Mass Spectrometry faculty grant NIEHS (P30 ES00210) for G.G.

2.1. Sparse recovery properties

The problem in (1) can be used to obtain row-sparse solution for X. However, the solution (1) may exhibit a trade-off between data fit and sparsity. A sparse solution may result in a poor data fit while a solution which provides a good data fit may have many non-zero rows. This trade-off can be controlled via ϵ . We are interested in a tuning-free method, i.e., a method which fixes ϵ to a given value which guarantee exact recovery of the row-sparsity. The following results provide a theoretical justification for selecting ϵ to guarantee such exact recovery.

Proposition 2.1. Recall the matrix to be recovered $X_0 \in \mathbb{R}^{k \times n}$ and define the solution set $S_{\epsilon} = \{X | \sum_{i=1}^{n} ||A_i x_i - y_i||_2^2 \le \epsilon\}$. Assume that $X_0 \in S_{\epsilon}$. Then, X^* the solution to (1) satisfies $\|X^*\|_{R_0} \le \|X_0\|_{R_0}$.

Proof. Since X^* minimizes $||X||_{R_0}$ over the set S_{ϵ} , we have $||X^*||_{R_0} \leq ||X||_{R_0}$ for all $X \in S_{\epsilon}$. Specifically since $X_0 \in S$, we have that $||X^*||_{R_0} \leq ||X_0||_{R_0}$.

This proposition suggests the optimization in (1) produces a solution with row-sparsity which is less than or equal to the row-sparsity of the true X_0 if S_{ϵ} is sufficiently large to contain the original X_0 .

Proposition 2.2. Let matrix X_0 satisfy $||X_0||_{R_0} = s_0$. Define

$$\gamma(X) = \min_{i=\overline{1,n}} \{ \|e_i^T X\|_2 | \|e_i^T X\|_2 \neq 0 \}.$$

Then, for any matrix \tilde{X} satisfying $\|\tilde{X} - X_0\|_F < \gamma(X_0)$, we have $\|\tilde{X}\|_{R_0} \ge \|X_0\|_{R_0}$.

Proof. (By contradiction) Assume that there exist \tilde{X} such that $\|\tilde{X}\|_{R_0} < \|X_0\|_{R_0}$, then $\|\tilde{X}-X_0\|_F \ge \min_{X',\|X'\|_{R_0} < \|X_0\|_{R_0}} \|X'-X_0\|_F \ge \min_{i,\|e_i^T X_0\|_2 \neq 0} \|e_i^T X_0\|_2 = \gamma(X_0)$. By contradiction to $\|\tilde{X}-X_0\|_F < \gamma(X_0)$, the assumption $\|\tilde{X}\|_{R_0} < \|X_0\|_{R_0}$ is invalid and hence $\|\tilde{X}\|_{R_0} \ge \|X_0\|_{R_0}$. \Box

This proposition suggests that there exists no matrix \tilde{X} of lower row-sparsity then that of X_0 in the $\gamma(X_0)$ -Frobenius ball neighborhood of X_0 . Intuitively, small changes to matrix X_0 which has sufficiently large row norms cannot set those rows to zero and hence cannot lower the row-sparsity of X_0 .

If the solution set S_{ϵ} is a subset of the $\gamma(X_0)$ -Frobenius ball neighborhood of X_0 then X^* must satisfy $||X^*||_{R_0} \ge$ $||X||_{R_0}$ (in addition to $||X^*||_{R_0} \le ||X||_{R_0}$), thereby guaranteeing $||X^*||_{R_0} = ||X||_{R_0}$. Next, we present a proposition which sets the conditions for $S_{\epsilon} \subset$ the $\gamma(X_0)$ -Frobenius ball neighborhood of X_0 .

Proposition 2.3. Assume $X_0 \in S_{\epsilon}$. Let matrices A_i be such that matrix A_i satisfy $2s_0$ - restricted isometry property with $\delta_{2s_0} < 1$. Specifically $||A_ix_i||_2 \ge (1 - \delta_{2s_0})||x_i||_2$. If

$$\gamma(X_0) > \frac{2\sqrt{\epsilon}}{1 - \delta_{2s_0}}$$

then $S_{\epsilon} \subset \gamma(X_0)$ -Frobenius ball neighborhood of X_0 and $\|X^*\|_{R_0} = \|X_0\|_{R_0}$.

Proof. Consider $X \in S_{\epsilon}$, then

$$\begin{aligned} \|X - X_0\|_F &= \sqrt{\sum_i \|x_i - x_{0i}\|_2^2} \\ &\leq \frac{\sqrt{\sum_i \|A_i(x_i - x_{0i})\|_2^2}}{1 - \delta_{2s_0}} \\ &\leq \frac{\sqrt{\sum_i (\|A_i x_i - y_i\|_2 + \|A_i x_{0i} - y_i\|_2)^2}}{1 - \delta_{2s_0}} \\ &\leq \frac{\sqrt{4\epsilon}}{1 - \delta_{2s_0}} < \gamma(X_0) \end{aligned}$$

Since any $X \in S_{\epsilon}$ is in the $\gamma(X_0)$ -neighborhood of X_0 , then $S_{\epsilon} \subset \gamma(X_0)$ -Frobenius ball neighborhood of X_0 . Moreover, the solution $X^* \in S_{\epsilon}$ and hence $\|X^*\|_{R_0} \leq \|X_0\|_{R_0}$ by proposition 2.1. Since $S_{\epsilon} \subset \gamma(X_0)$ -Frobenius ball neighborhood of X_0 , $X^* \in \gamma(X_0)$ -Frobenius ball neighborhood of X_0 and therefore $\|X^*\|_{R_0} \geq \|X_0\|_{R_0}$. We conclude that $\|X^*\|_{R_0} = \|X_0\|_{R_0}$.

This proposition suggests that if the true X_0 is in the solution set S_{ϵ} and each one of its s_0 non-zero rows has sufficient large l_2 -norm, then the solution of (1) will have the same rowsparsity as that of X_0 thereby recovering the sparsity of X_0 , while providing an $\frac{2\sqrt{\epsilon}}{1-\delta_{2s_0}}$ bound on the the Frobenius norm of the error $X^* - X_0$.

One of the key issues of exact rank recovery involves finding an ϵ which guarantees that $X_0 \in S_{\epsilon}$. In the statistical setting, this typically can be obtained by understanding the noise characteristics. The following proposition suggests how to determine the value of ϵ .

2.2. Determining ϵ based on the probabilistic noise model

In general, a statistical analysis of the term $\sum_i ||A_i x_{0i} - y_i||_2^2$ can be performed to obtain in probability guarantees on the inequality $\sum_i ||A_i x_{0i} - y_i||_2^2 \le \epsilon$.

Proposition 2.4. Let matrix $X_0 \in \mathbb{R}^{k \times n}$ be given. Let $y_i = A_i x_{0i} + \xi_i$, $\forall i = \overline{1, n}$, where $\xi_i \sim Normal(0, I_{m \times m})$. Let $p \in (0, 1)$ be given and set ϵ according to

$$\epsilon(p) = (\chi^2_{nm})^{-1}(p).$$

Then with probability $1 - p, X_0 \in S_{\epsilon}$.

Proof. Note that $\sum_{i=1}^n \|A_i x_i - y_i\|_2^2 = \sum_{i=1}^n \sum_{j=1}^m \xi_{ij}^2 \sim \chi^2_{nm}$ then

$$P(\sum_{i=1}^{n} \|A_{i}x_{i} - y_{i}\|_{2}^{2} \le \epsilon) = P(\sum_{i=1}^{n} \sum_{j=1}^{m} \xi_{ij}^{2} \le \epsilon) \ge 1 - p$$

3. CONVEX JOINT SPARSE MINIMIZATION

3.1. Convex optimization formulation

The optimization problem in (1) is difficult to solve due to its combinatorial nature. An effective solution approach is to substitute the original objective function with the mixed $\ell_{p,q}$ norm of X [7, 8, 9], thereby transforming the original problem into a more tractable convex optimization problem.

The mixed $\ell_{p,q}$ norm of X is defined as $||X||_{p,q} = (\sum_{i=1}^{k} ||x^i||_q^p)^{1/p}$, where x^i is the *i*th row of X.

Using this approach with p = 1 and q = 2, we transform the original problem in (1) into the convex optimization problem below.

minimize
$$||X||_{1,2}$$

subject to $\sum_{l=1}^{n} ||A_l x_l - y_l||_2^2 \leq \epsilon$, (2)

3.2. Algorithmic solution

There are many approaches to solving convex problems. In this paper, we propose an algorithm based on the optimization transfer approach. The main idea of the optimization transfer approach is to find an appropriate surrogate function such that it is easier to minimize the surrogate function than the objective function. Furthermore, minimizing the surrogate function also leads to minimizing the intended objective function. We begin with the definition of a surrogate function.

Definition 3.1. Function $G(x, \tilde{x})$ is a surrogate function for F(x) if the conditions: (i) $G(x, \tilde{x}) \ge F(x)$ and (ii) G(x, x) = F(x) are satisfied.

Lemma 3.1. Let $G(x, \tilde{x})$ be a surrogate function for F(x), then F(x) is nonincreasing under the update

$$\tilde{x}^{t+1} = \arg\min G(x, \tilde{x}^t), \tag{3}$$

$$\textit{Proof.} \ F(x^{t+1}) \leqslant G(x^{t+1},x^t) \leqslant G(x^t,x^t) = F(x^t). \quad \Box$$

Lemma 3.1 states that running the update rule above for the surrogate function $G(x, \tilde{x})$ iteratively will lead to minimizing the objective function F(x). Thus, selecting the appropriate surrogate function for which the update can be computed efficiently is essential for the optimization transfer algorithms.

We proceed with the derivation of a surrogate function based on which, we develop an optimization transfer algorithm for solving the optimization problem in (2). The Lagrangian function for this problem is given by

$$\mathcal{L}(X;\lambda) = \|X\|_{1,2} + \lambda (\sum_{l=1}^{n} \|A_l x_l - y_l\|_2^2 - \epsilon).$$
(4)

First we note that

$$||X||_{1,2} \leq \frac{1}{2} \left[\operatorname{Tr} \left(X^T D X \right) + ||\tilde{X}||_{1,2} \right],$$
 (5)

where $D \in \mathbb{R}^{k \times k}$ is a diagonal matrix of the form

$$D = Diag(\frac{1}{\|\tilde{x}^1\|_2}, \dots, \frac{1}{\|\tilde{x}^k\|_2})$$

The inequality (5) follows from the fact that for any two points z and \tilde{z} ,

$$|z| \leq \frac{1}{2} \left(\frac{z^2}{|\widetilde{z}|} + |\widetilde{z}| \right). \tag{6}$$

Since $\operatorname{Tr}(X^T D X) = \sum_{i=1}^n x_i^T D x_i$, we have $\mathcal{L}(X; \lambda) \leq \mathcal{U}(X, \tilde{X}; \lambda)$ where

$$\mathcal{U}(X, \tilde{X}; \lambda) = \frac{1}{2} \sum_{i=1}^{n} x_i^T D x_i + \lambda \sum_{l=1}^{n} \|A_l x_l - y_l\|_2^2 + C \quad (7)$$

and $C = \frac{1}{2} \|\tilde{X}\|_{1,2} - \lambda \epsilon$ is independent of X. It is not hard to see that $\mathcal{U}(X, \tilde{X}; \lambda)$ satisfies the definition of a surrogate function for $\mathcal{L}(X; \lambda)$. Furthermore, the update step in Lemma 3.1 can be done efficiently by differentiating $\mathcal{U}(X, \tilde{X}; \lambda)$ with respect to each column x_i separately and set it to zero to obtain:

$$\tilde{x_i}^{t+1} = \left(D(X^t) + 2\lambda A_i^T A_i\right)^{-1} 2\lambda A_i^T y_i.$$
(8)

Note that D depends on X^t .

Now, for a given λ , the update rule in (8) can be used to determine the corresponding optimal $X(\lambda)$. In practice, the true optimal λ^* is unknown. We now describe an algorithm to determine the optimal x_i^* by adjusting λ towards the true λ^* .

If $\lambda^* = 0$, then $\mathcal{L}(X^*; 0) = ||X^*||_{1,2} = 0$, which implies $X^* = \mathbf{0}$ and $\sum_{l=1}^n ||y_l||_2^2 \leq \epsilon$. The last condition enables us to obtain the optimal $X^* = \mathbf{0}$ immediately. On the other hand, when λ^* is strictly greater than 0, then by complementary slackness condition, $\sum_{l=1}^n ||A_l x_l^* - y_l||_2^2 - \epsilon = 0$. Therefore, we decrease λ when $\sum_{l=1}^n ||A_l x_l^* - y_l||_2^2 < \epsilon$ to penalize the difference. Otherwise, we increase λ . The pseudo code for the proposed algorithm is provided in Algorithm 1.

4. SIMULATION RESULTS

To evaluate the performance of an algorithm and verify that selection of ϵ is correct, i.e., produces correct row-sparsity recovery, we constructed a synthetic data set following the additive Gaussian noise model. To produce A, we generate each entry in an independent fashion according to N(0,1)and normalized each column such that the l_2 of each column is 1. We produced X^0 with row-sparsity of three by setting all entry in 3 random rows to γ/\sqrt{n} so that the l_2 norm of the all non-zero rows of X^0 is greater than γ in Prop. 2.2. Each y_i Algorithm 1 Row-sparsity recovery algorithm

Input: $(A, y, \lambda_{min}, \lambda_{max}, \epsilon, \delta)$. Output: X^* . 1: **if** $\sum_{l=1}^{n} \|y_l\|_2^2 \le \epsilon$) **then** $X^* = \mathbf{0}$, Terminated 2: 3: else while $\|\sum_{l=1}^{n} \|A_l x_l^* - y_l\|_2^2 - \epsilon \| > \delta$ do 4: $\lambda = (\lambda_{min} + \lambda_{max})/2$ 5: Find $X^*(\lambda)$ using the update rule in (8) 6: if $\sum_{l=1}^{n} \|A_l x_l^* - y_l\|_2^2 < \epsilon$ then $\lambda_{max} = \lambda$ 7: 8: else 9٠ 10: $\lambda_{min} = \lambda$ end if 11: end while 12: 13: end if

was computed by $A_i x_i + \xi_i$, where $\xi_i(j) \sim \mathcal{N}(0, 1)$. Finally, ϵ was calculated using Prop. 2.4. The vectors $\{y_i\}$, matrices $\{A_i\}$, and ϵ were provided as input to our algorithm.

To evaluate the accuracy of the method, we scanned through a range of values of ϵ and computed the sparsity of $X^*(\epsilon)$. In Fig. 1, we present a plot of the row-sparsity of X^* as a function of epsilon. Additionally, we present two lines. The first, is a vertical line indicating the value of ϵ computed via Prop. 2.4. The second line is a horizontal line indicating the true row-sparsity of X^0 . We observe that the sparsity of X^* at the proposed ϵ matches the sparsity of X^0 . Moreover, we observe that the proposed ϵ is close to the LHS of the line segment corresponding to the correct sparsity. This indicates the potential of the proposed ϵ to find exact row-sparsity solution with a low data fit error value.



Fig. 1. Row-sparsity of the solution X^* as a function of ϵ . The vertical line corresponds the value of ϵ proposed by Proposition 2.4 and the horizontal line marks the true sparsity value.

5. CONCLUSION

We presented a variant of the MMV problem whose common sparsity pattern is obfuscated by an additive noise. Specifically, we derived the conditions for perfect reconstruction of the original sparsity pattern. Based on these, we presented a tuning-free algorithm for recovering jointly sparse solutions which is based on optimization transfer. Our simulations indicate the capability of the proposed approach to recover the correct joint sparsity pattern.

6. REFERENCES

- E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal recovery from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, February 2006.
- [2] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, March 2006.
- [3] I.J. Fevrier, S.B. Gelfand, and M.P. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *Communications*, *IEEE Transactions on*, vol. 47, no. 6, pp. 927–937, 1999.
- [4] M. Mishali and Y.C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4692– 4702, 2008.
- [5] Wei Deng, W. Yin, and Yin Zhang, "Group sparse optimization by alternating direction method," 2011.
- [6] D.L. Donoho, M. Elad, and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6–18, 2006.
- [7] E. Berg and M.P. Friedlander, "Joint-sparse recovery from multiple measurements," *Arxiv preprint* arXiv:0904.2051, 2009.
- [8] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2477–2488, 2005.
- [9] Y.C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 5302– 5316, 2009.