Jointly Sparse Signal Recovery with Prior Info

Natalie Durgin[1], Rachel Grotheer[2], Chenxi Huang[3], Shuang Li[4], Anna Ma[5], Deanna Needell[6], and Jing Qin[7]

***Abstract*—The multiple measurement vector (MMV) problem with jointly sparse signals has been of recent interest across many fields and can be solved via $\ell_{2,1}$ minimization. In such applications, *prior information* is typically available and utilizing weights to incorporate the prior information has only been empirically shown to be advantageous. In this work, we prove theoretical guarantees for a weighted $\ell_{2,1}$ minimization approach to solving the MMV problem where the underlying signals admit a jointly sparse structure. Our theoretical findings are complemented with empirical results on simulated and real world video data.**

I. Introduction

As the amount of available data grows, it becomes demanding to design fast and scalable approaches to processing this data. In particular, the control and analysis of time-varying large scale data has recently attracted a lot of attention in applications such as wireless communication [1, 2], medical imaging hyperspectral diffuse optical tomography (hyDOT) [3, 4], and video signal recovery [5, 6]. In such applications, the goal is to recover a sequence of highly correlated signals $\{\boldsymbol{x}_t\}_{t=1}^T$ from measurements of the form

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{z}_t, \tag{I.1}$$

where the unknown signals are assumed to be sparse, i.e., few non-zero entries. Here, we use the measurement matrix $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ $(M \ll N)$ and linear measurements $\boldsymbol{y}_t \in \mathbb{R}^M$ without precise information about the additive noise $\boldsymbol{z}_t \in \mathbb{R}^M$.

Problem (I.1) can be solved using *multiple measurement vector (MMV)* approaches, which can recover all signals $\{\boldsymbol{x}_t\}_{t=1}^T$ simultaneously. Since MMV-based approaches can leverage sequential correlation information, there has been an increasing interest in developing algorithms that can solve MMV problems [7, 8, 9]. In addition to MMVs, prior information on the correlation amongst signals can also be used to improve signal recovery. In this work, we focus on the MMV model where the underlying signals admit a *jointly sparse* structure, i.e., the underlying signals $\boldsymbol{x}_t$ share the same sparse support, and *prior information* on the joint support is available.

There are a lot of applications in which jointly sparse signals and prior information are evident, e.g., hyperspectral diffuse optical tomography (hyDOT) [3, 4] and video recovery [5, 6]. In hyDOT, each of the measurement vectors corresponds to a specific wavelength at which a fixed tissue sample is imaged. In this application, non-zero entries in each $\boldsymbol{x}_t$ correspond to locations of cancerous cells. The jointly sparse structure is attributed to the fact that cancerous cells consistently have larger absorption coefficients across different wavelengths. Practitioners with field expertise may have general ideas on where cancerous cells can live and where they cannot. This prior information can be useful when trying to recover the location of cancerous cells, i.e., the joint support. In the video recovery problem, video frames have a small amount of variation between consecutive frames, thus one could expect that the support of the frames stays relatively consistent, i.e., each frame shares an approximately jointly sparse structure. In video surveillance applications, there could be areas of interest where activity is expected to occur. This knowledge can act as prior information and be useful in recovering video frames from measurements.

In this work, we provide recovery error bounds for the approximation of jointly sparse vectors in the presence of prior information using weighted $\ell_{2,1}$ minimization. In particular, we show that the average recovery error per signal improves as more measurement vectors become available and accurate prior information is provided. While weighted $\ell_{2,1}$ minimization for joint sparsity was discussed in [7, 8], our work is the first to provide

[1]Spiceworks, Austin, TX, njdurgin@gmail.com. [2]Goucher College, Baltimore, MD, rachel.grotheer@goucher.edu. [3]Yale University, New Haven, CT, chenxi.huang@yale.edu. [4]Colorado School of Mines, Golden, CO, shuangli@mines.edu. [5]University of California, San Diego, CA, annama@ucsd.edu. [6]University of California, Los Angeles, Los Angeles, CA, deanna@math.ucla.edu. [7]University of Kentucky, Lexington, KY, jing.qin@uky.edu. The authors would like to thank ICERM at Brown University for hosting the initial WiSDM workshop at which this collaboration began, as well as for funding our follow-up collaboration. Needell was funded by NSF CAREER DMS #1348721 and NSF BIGDATA #1740325. Li was supported by the NSF grants CCF-1409258, CCF-1704204, and the DARPA Lagrange Program under ONR/SPAWAR contract N660011824020. Qin is supported by the NSF grant DMS-1941197.

theoretical guarantees for signal recovery in this setting.

The remainder of this paper is organized as follows. In Section II, we discuss necessary groundwork in the *single measurement vector* (SMV) setting, i.e., when $T = 1$. In Section III, we formulate our problem by introducing the MMV model of interest and the related existing results from [10]. In Section IV, we provide theoretical guarantees for the weighted $\ell_{2,1}$ minimization in the presence of prior information. To verify our theoretical findings, simulated data and real world video recovery experiments are conducted in Section V. Finally, we provide some concluding remarks and future directions in Section VI.

II. Previous Work

We refer to the problem consisting of a single signal as the SMV problem. One of the most popular approaches to recover a sparse signal $\boldsymbol{x}$ in the SMV framework is to minimize the ℓ_1 norm of the signal subject to an inequality constraint, i.e.,

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ subject to } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \varepsilon. \quad \text{(II.1)}$$

The above ℓ_1 minimization has been widely studied under a variety of settings and applications [11, 12, 13]. Note that the parameter ε is related to the noise level or, assuming $\boldsymbol{z}_t \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, the standard deviation σ of the Gaussian noise. For other noise types, the constraints can be modified accordingly.

For $t = 1, \ldots T$, given measurements $\boldsymbol{y}_t$ from (I.1), each associated with an unknown signal $\boldsymbol{x}_t$, it is possible to recover each $\boldsymbol{x}_t$ one at a time using the ℓ_1 minimization (II.1). However, this naive approach cannot take advantage of correlation among signals at different times and thereby loses reconstruction efficiency.

In addition, one can also use *prior information* from previous signals to improve signal recovery at the next time step with the SMV model. In particular, if signals are slowly varying over time, the previous signal can be a good proxy for the next signal. A common approach for incorporating prior information on signals is to use a weighting scheme to place emphasis on entries of *importance* [14, 10]. Prior works have shown that when accurate prior information is available, the solutions to weighted optimization problems tend to outperform their unweighted counterparts [10].

III. Problem Formulation

A. Joint Sparse MMV Signal Recovery

In the MMV framework, we aim to recover multiple signals simultaneously rather than sequentially, given measurements $\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{z}_t$ for $t = 1, \ldots, T$. For notational simplicity, vectors are column-wise concatenated to form the matrices $\boldsymbol{X} \in \mathbb{R}^{N \times T}$, $\boldsymbol{Y} \in \mathbb{R}^{M \times T}$, and $\boldsymbol{Z} \in \mathbb{R}^{M \times T}$. We denote rows, columns, and entries of a matrix $\boldsymbol{M}$ by $\boldsymbol{M}_{(i,\cdot)}$, $\boldsymbol{M}_{(\cdot,j)}$, and $\boldsymbol{M}_{(i,j)}$, respectively. The joint support of $\boldsymbol{X}$ is the set $S \subset \{1, \ldots, N\}$ of indices such that $\boldsymbol{X}_{(i,\cdot)}$ contains at least one non-zero entry. We say that $\boldsymbol{X}$ is K-row sparse if $|S| = K$. Additionally, the best K-row sparse approximation of $\boldsymbol{X}$ is denoted as $\boldsymbol{X}_K = \boldsymbol{X}|_S$ with $S = \operatorname{argmin}_{|S|=K} \|\boldsymbol{X}|_S - \boldsymbol{X}\|_F^2$, that is, the matrix $\boldsymbol{X}$ with rows indexed by $\{1, \ldots, N\} \setminus S$ are replaced with zeros. We will use S_0 to denote the row support of $\boldsymbol{X}_K$.

Given a weight vector $\boldsymbol{w} \in \mathbb{R}^N$, which can be obtained from the prior information (see more details in Section III-B), we can recover each column of $\boldsymbol{X}$ one at a time with weighted ℓ_1 minimization [10] and we refer to this recovery strategy as *weighted SMV* throughout. In the case when there is no prior information, i.e., $\boldsymbol{w}$ is the all ones vector, the weighted ℓ_1 minimization reduces to ℓ_1 minimization, which is referred to as the *unweighted SMV* algorithm in this work. A more elegant way to recover the jointly sparse matrix $\boldsymbol{X}$ is to recover the whole matrix at one time. As in some previous works [7, 8], we also consider the following weighted $\ell_{2,1}$ minimization

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{2,1,\boldsymbol{w}} \text{ s.t. } \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_F \leq \varepsilon, \quad \text{(III.1)}$$

for jointly sparse MMV signal recovery when some prior information is provided. Here, the weighted $\ell_{2,1}$ norm is defined as $\|\boldsymbol{X}\|_{2,1,\boldsymbol{w}} \triangleq \sum_{i=1}^N w_i \|\boldsymbol{X}_{(i,\cdot)}\|_2$. We use the weighted $\ell_{2,1}$ norm in (III.1) because minimizing the $\ell_{2,1}$ norm promotes row sparsity. We denote program (III.1) as the *weighted MMV* algorithm and use *unweighted MMV* to denote the case when we have no prior information.

B. Weight Vector with Prior Information

Given some prior information in the form of an estimated joint support, $\widehat{S} \subset \{1, \ldots, N\}$, we set the weight vector according to

$$w_i = \begin{cases} p_i \in \mathbb{R}_+, & \text{if } i \in \widehat{S} \\ 1, & \text{else} \end{cases}, \quad \text{(III.2)}$$

as in [10]. Here, weights can take any non-negative value and are not restricted to be uniform over all entries in the estimate of the joint support. For arbitrary weights in the weighted ℓ_1 minimization, the authors of [10] show that under certain RIP conditions, the approximation of $\boldsymbol{x}$ can be bounded above by a combination of the measurement error, K-sparse approximation error, and accuracy of the prior information. In this work, we extend their results to the weighted $\ell_{2,1}$ minimization for arbitrarily weighted rows.

While the theoretical guarantees presented in this work are for generic weights p_i, the experimental section considers three different weighting schemes: (1) a fixed weight of $p_i = 0.1$, (2) a weighting scheme based on the accuracy of the prior information, i.e., $p_i = 1 - \frac{|\widehat{S} \cap S_0|}{K}$ [11], and (3) an ℓ_1-optimal weighting scheme based on [15], where the computed weight is optimal in the sense of the number of measurements needed for exact signal recovery is minimized.

IV. Theoretical Guarantees

In this section, we present the theoretical performance guarantees for the weighted $\ell_{2,1}$ minimization (Theorem 4.1). In particular, we show that, under mild conditions, the approximation error between the estimated signal $\widehat{\boldsymbol{X}}$ and the true signal $\boldsymbol{X}$ is upper bounded by terms dependent on the best K-row sparse approximation of $\boldsymbol{X}$ and the reliability of the prior information. The constants C_0' and C_1' that appear in Theorem 4.1 are made explicit in Remark 4.1. Remark 4.2 makes explicit the benefit of using MMVs.

Theorem 4.1: Let $\boldsymbol{X}_K$ denote the best K-row sparse approximation of $\boldsymbol{X} \in \mathbb{R}^{N \times T}$, and S_0 denote the row support of $\boldsymbol{X}_K$. Let $\widetilde{S}_i \subset \{1, ..., N\}$ for $i = 1, ..., n$ where $1 \leq n \leq N$ be arbitrary disjoint sets and denote $\widetilde{S} = \bigcup_{i=1}^{n} \widetilde{S}_i$. Without loss of generality, assume that the weights are ordered so that $1 \geq w_1 \geq w_2 \geq ... \geq w_N \geq 0$ and let $\omega = \sum_{i=1}^{N} w_i$. For each i, define the relative size ρ_i and accuracy α_i as $\rho_i = \frac{|\widetilde{S}_i|}{K}$ and $\alpha_i = \frac{|\widetilde{S}_i \cap S_0|}{|\widetilde{S}_i|}$. Furthermore, suppose that there exists a constant a satisfying $a > 1$, $a \in \frac{1}{K}\mathbb{Z}$ and $\sum_{i=1}^{N} \rho_i (1 - \alpha_i) \leq a$, and that the measurement matrix $\boldsymbol{A}$ has the RIP with

$$\delta_{aK} + \frac{a}{\kappa_N} \delta_{(a+1)K} < \frac{a}{\kappa_N^2} - 1,$$

where

$$\kappa_N = w_N + (1 - w_1) \sqrt{1 + \sum_{i=1}^{N} (\rho_i - 2\alpha_i \rho_i)} + \sum_{j=2}^{N} \left((w_{j-1} - w_j) \sqrt{1 + \sum_{i=j}^{N} \rho_i - 2\alpha_i \rho_i} \right).$$

Then, the minimizer $\widehat{\boldsymbol{X}}$ of the weighted $\ell_{2,1}$ minimization (III.1) obeys

$$\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_F \leq C_0' \varepsilon + C_1' K^{-\frac{1}{2}} \omega \|\boldsymbol{X}_K - \widehat{\boldsymbol{X}}\|_{2,1} + C_1' K^{-\frac{1}{2}} \left((1 - \omega) \|\boldsymbol{X}_{\widetilde{S}^c \cap \widetilde{S}_0^c}\|_{2,1} - \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} w_j \|X_{\widetilde{S}_i \cap S_0^c}\|_{2,1} \right),$$

where C_0' and C_1' are well-behaved constants that depend on the measurement matrix $\boldsymbol{A}$, the weights w_i, and the parameters ρ_i and α_i. We use S^c to denote the complement of a set S.

Remark 4.1: As in [10], the constants C_0' and C_1' are explicitly given by

$$C_0' = \frac{2\left(1 + \frac{\kappa_N}{\sqrt{a}}\right)}{\sqrt{1 - \delta_{(a+1)K}} - \frac{\kappa_N}{\sqrt{a}} \sqrt{1 + \delta_{aK}}}, \tag{IV.1}$$

and

$$C_1' = \frac{2a^{-\frac{1}{2}} \left(\sqrt{1 - \delta_{(a+1)K}} + \sqrt{1 + \delta_{aK}} \right)}{\sqrt{1 - \delta_{(a+1)K}} - \frac{\kappa_N}{\sqrt{a}} \sqrt{1 + \delta_{aK}}}. \tag{IV.2}$$

Remark 4.2: To compare the recovery bounds for SMV and MMV frameworks, we note that:

1) For any matrix $\boldsymbol{M} \in \mathbb{R}^{N \times T}$,
$$\|\boldsymbol{M}\|_{2,1} \leq \sqrt{T} \sum_{i=1}^{N} \max_t \boldsymbol{M}_{(i,j)}^2.$$
2) The tolerance on the Frobenius error in (III.1), i.e., ε, grows on the order of $\sqrt{T}$.

It can be seen that the average recovery error per signal (of T total signals) for the non-uniformly weighted $\ell_{2,1}$ minimization depends inversely on $\sqrt{T}$. Therefore, the average recovery error per signal decreases as the number of MMVs increase.

V. Experiments

In the first experiment, we compare the recovery performance of the weighted and unweighted $\ell_{2,1}$ minimization and sequentially recovering signals with weighted and unweighted ℓ_1 minimization.[1] In this experiment, we set the dimension of the data $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ to be $M = 20$, $N = 100$, and set the row sparsity $K = 10$. Entries of the noise matrix $\boldsymbol{Z}$ are drawn i.i.d. from a normal distribution with mean 0 and standard deviation $\sigma = 0.1$ so that the parameter ε in (III.1) is set as $\varepsilon = \sigma$. Each result is averaged over 50 trials. The relative error is defined as $\frac{\|\widehat{\mathbf{X}} - \boldsymbol{X}\|_F}{\|\boldsymbol{X}\|_F}$, where $\widehat{\mathbf{X}}$ is the estimation of the ground truth $\boldsymbol{X}$. In Figure 1, we fix the number of signals $T = 10$. The weights p_i introduced in (III.2) are being varied and we assume that perfect prior information is known, i.e., $\widehat{S} = S_0$, where $\widehat{S}$ and S_0 denote the given approximate joint support and the true joint support, respectively. In Figure 2, we use $p_i = 0.1$, $p_i = 1 - \frac{|\widehat{S} \cap S_0|}{K}$,

[1] We use the CVX software package [16] to solve all of the minimization programs.

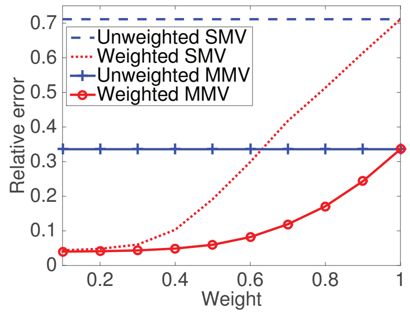

Figure 1: Performance of solving jointly sparse MMV problems using weighted and unweighted versions of SMV and MMV algorithms with different weights.

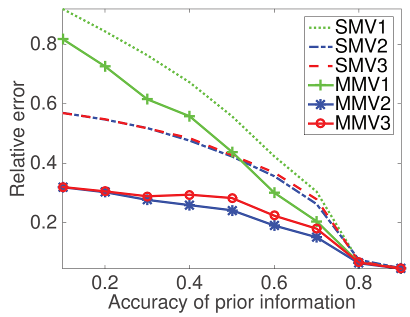

Figure 2: The performance of solving weighted ℓ_1 minimization and weighted $\ell_{2,1}$ minimization (III.1) using different weighting schemes. Scheme 1 uses $p_i = 0.1$, scheme 2 uses $p_i = 1 - \frac{|\widehat{S} \cap S_0|}{K}$, and scheme 3 uses the "optimal weight" introduced in Section III.

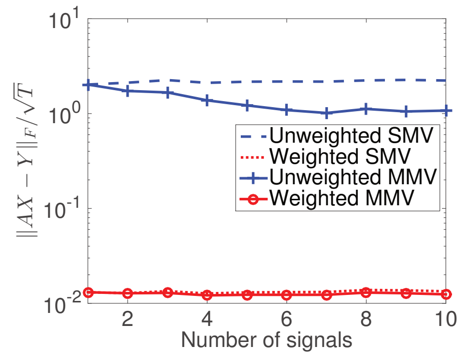

Figure 3: The scaled error: $\frac{\|AX-Y\|_F}{\sqrt{T}}$ (vertical axis) as a function of number of signals T (horizontal axis).

and an "optimal weight" from paper [15] as introduced in Section III. In Figure 3, we set $p_i = 0.1$ and assume that perfect prior information is known. Then, we present the relationship between the number of available signals T in the MMV framework and the square root of the average error per signal, i.e., $\frac{\|\boldsymbol{AX}-\boldsymbol{Y}\|_F}{\sqrt{T}}$. As indicated in our theory, the recovery error $\|\boldsymbol{AX} - \boldsymbol{Y}\|_F$ for our proposed weighted MMV algorithm does scale with $\sqrt{T}$.

In the second experiment, we compare the weighted $\ell_{2,1}$ and weighted ℓ_1 algorithms and their corresponding unweighted versions on jointly sparse video sequence reconstruction. We use a candle video which consists of 75 frames from the Dynamic Texture Toolbox in http://www.vision.jhu.edu/code/. An example of the 5th frame is shown in Figure 4 (a). We select a subset of the frames, i.e., the first 10 frames, each of which is of size 64×32. Then we create a data matrix $\boldsymbol{X} \in \mathbb{R}^{2048 \times 10}$, whose columns are a vectorization of all video frames. To further obtain a sparse representation of $\boldsymbol{X}$, we use a 2048×2048 Daubechies wavelet dictionary. An example of the 5th reconstructed frame from the Daubechies wavelet dictionary is shown in Figure 4 (b). The number of non-zero rows in the sparse coefficient matrix is $K = 393$. We set the accuracy of $\widehat{S}$ as $(K-2)/K = 0.9949$. Then, we use "scheme 2" as in Figure 2 to set the weights for weighted algorithms, i.e., $p_i = 1 - \frac{|\widehat{S} \cap S_0|}{K}$. We also use a random Gaussian sensing matrix of size 1179×2048 to compress $\boldsymbol{X}$ and add a random Gaussian noise with mean 0 and variance $\sigma = 0.02$ to obtain the measurements. The relative recovery errors for both the sparse coefficient matrix Θ and the noiseless data matrix $\boldsymbol{X}$ are shown in Table I. We present the 5th recovered frame with different algorithms in Figure 4. It can be seen that our proposed weighted MMV algorithm has a lower recovery error than the other three algorithms.

VI. Conclusion

In this work, we extend the weighted SMV algorithm, weighted ℓ_1 minimization, to the MMV framework and consider the *weighted $\ell_{2,1}$ minimization* for jointly sparse signal recovery. We provide theoretical guarantees for the weighted MMV algorithm, weighted $\ell_{2,1}$ minimization, as well as a series of experiments to show the advantage over its unweighted and/or SMV counterparts. Currently, we focus on signals with jointly sparse support motivated by time-varying signals that admit a jointly sparse structure. We leave the more general time-varying model without joint support for future work.

References

[1] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*, vol. 2. prentice hall PTR New Jersey, 1996.

	Coefficient (Θ)	Data (X)
Unweighted SMV	0.0372	0.0709
Weighted SMV	0.0351	0.0699
Unweighted MMV	0.0360	0.0703
Weighted MMV	**0.0312**	**0.0680**

Table I: Relative recovery error for both sparse coefficient matrix Θ and the noiseless data matrix $\boldsymbol{X}$.

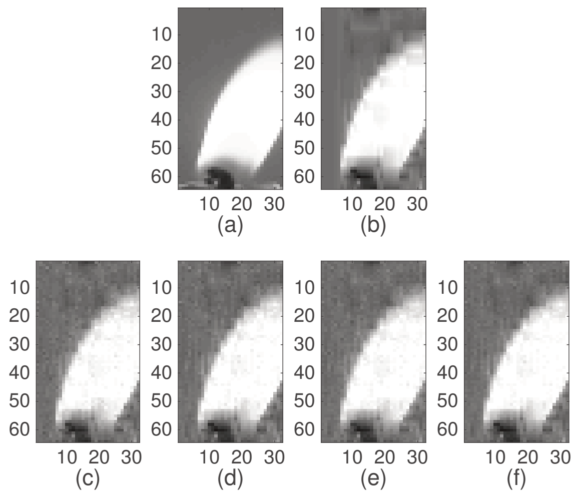

Figure 4: Candle video experiment: (a) The 5th frame of the candle video. (b) The 5th frame reconstructed by the Daubechies wavelet dictionary with relative representation error : 0.0572. (c-f) The 5th frame recovered from different algorithms with Daubechies wavelet dictionary. The relative recovery error for this frame is: (c) unweighted SMV: 0.0679, (d) weighted SMV: 0.0665, (e) unweighted MMV: 0.0668, (f) weighted MMV: 0.0645.

[2] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1527–1550, 2017.

[3] N. Durgin, R. Grotheer, C. Huang, S. Li, A. Ma, D. Needell, and J. Qin, "Fast hyperspectral diffuse optical imaging method with joint sparsity," in *2019 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, to appear.

[4] N. Durgin, R. Grotheer, C. Huang, S. Li, A. Ma, D. Needell, and J. Qin, "Sparse randomized kaczmarz for support recovery of jointly sparse corrupted multiple measurement vectors," in *Research in Data Science*, pp. 1–14, Springer, 2019.

[5] A. Majumdar and R. K. Ward, "Face recognition from video: an mmv recovery approach," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2221–2224, IEEE, 2012.

[6] J. Qin, S. Li, D. Needell, A. Ma, R. Grotheer, C. Huang, and N. Durgin, "Stochastic greedy algorithms for multiple measurement vectors," *arXiv preprint arXiv:1711.01521*, 2017.

[7] B. Ma, A. Zhang, and D. Xiang, "A unified approach to weighted $\ell_{2,1}$ minimization for joint sparse recovery," in *2nd International Conference on Teaching and Computational Science*, Atlantis Press, 2014.

[8] C. Zheng, G. Li, X.-G. Xia, and X. Wang, "Weighted $\ell_{2,1}$ minimisation for high resolution range profile with stepped frequency radar," *Electronics letters*, vol. 48, no. 18, pp. 1155–1156, 2012.

[9] S. Li, D. Yang, G. Tang, and M. B. Wakin, "Atomic norm minimization for modal analysis from random and compressed samples," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1817–1831, 2018.

[10] D. Needell, R. Saab, and T. Woolf, "Weighted ℓ_1-minimization for sparse recovery under arbitrary prior information," *Information and Inference: A Journal of the IMA*, vol. 6, no. 3, pp. 284–309, 2017.

[11] H. Mansour and R. Saab, "Recovery analysis for weighted ℓ_1-minimization using the null space property," *Applied and Computational Harmonic Analysis*, vol. 43, no. 1, pp. 23–38, 2017.

[12] D. Needell, "Noisy signal recovery via iterative reweighted ℓ_1-minimization," *arXiv preprint arXiv:0904.3780*, 2009.

[13] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[14] L. Weizman, Y. C. Eldar, and D. Ben Bashat, "Compressed sensing for longitudinal mri: an adaptive-weighted approach," *Medical physics*, vol. 42, no. 9, pp. 5195–5208, 2015.

[15] A. Flinth, "Optimal choice of weights for sparse recovery with prior information," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 4276–4284, 2016.

[16] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2009.