A Unified Approach to Weighted $L_{2,1}$ Minimization for Joint Sparse Recovery

Binqiang Ma, Aodi Zhang, Dongyang Xiang
Naval Marine Academy, Guangzhou 510430, China

***Abstract*—A unified view of the area of joint sparse recovery is presented for the weighted $L_{2,1}$ minimization. The support invariance transformation (SIT) is discussed to insure that the proposed scheme does not change the support of the sparse signal. The proposed weighted $L_{2,1}$ minimization framework utilizes a support-related weighted matrix to differentiate each potential position, resulting in a favorable situation that larger weights are assigned at those positions where indices of the corresponding bases are more likely to be outside of the row support so that the solution at those positions are close to zero. Therefore, the weighted $L_{2,1}$ minimization prefers to allot the received energy to those positions where indices of the corresponding bases are inside of the row support, which further improves the sparseness of the solution. The simulations demonstrate that the weighted $L_{2,1}$ minimization reaches the strong recover threshold with lower SNR and fewer measurements.**

Keywords-Weighted $L_{2,1}$ minimization; sparse signal reconstruction; multiple measurement vectors (MMV)

I. INTRODUCTION

The sparse recovery can be seen as a process that allots the received energy to the corresponding bases that are subject to the given constraint, which only those bases whose indices are inside of the support are considered as a correct situation. For the regular ℓ_1 minimization every basis has same priority class when the received energy is assigned. For the weighted ℓ_1 minimization, however, every basis has different priority class that is materialized with different weights when the received energy is assigned. Those bases whose indices are inside of the support has a priority by employing small weights, while other bases whose indices are more likely to be outside of the support are refused to assign the received energy by employing large weights [1-5]. Compared with the regular ℓ_1 minimization, the weighted ℓ_1 minimization not only avoids the disadvantage of the dependence on magnitude of the regular ℓ_1 minimization but also further promotes the sparseness of the solution and improves the performance [1-5]. It is proved that, for a nontrivial class of signals, the methodology of the weighted ℓ_1 minimization can enhance the recoverable sparsity thresholds upon the regular ℓ_1 minimization [3]. In addition, Needell provided the provable results that the weighted ℓ_1 minimization improves the recovery accuracy in the noisy case [4].

Obviously, designing the support-related weighted matrix is essential in order to achieve the methodology of the weighted ℓ_1 minimization. For example, the iterative reweighted ℓ_1 minimization that was presented to deal with the Single Measurement Vector (SMV) problem employs the iterative process to appoint larger weights to those locations whose indices are more likely to be outside of the support [1]. In this paper, however, we focus on the Multiple Measurement Vectors (MMV) problem that has many applications in the areas of array processing [6-10], nonparametric spectrum analysis of time series [11], equalization of sparse communication channels [12]. We firstly define the support invariance transformation (SIT) that the support of the sparse signal is invariant when the weighted processing about the sparse signal is employed to improve the performance. We give a proposition to insure the proposed scheme is SIT in terms of the sparse signal. Next, for designing a weighted $\ell_{2,1}$ minimization scheme (the mixed norm $\ell_{2,1}$ norm can be regarded as the counterpart of the ℓ_1 norm in the MMV case), the methods of spectral analysis that can obtain the estimates of the power spectra of signal upon given bases (e.g., the Bartlett [11], [13]) is employed to gain the support-related weighted matrix. As a result, the proposed method appoints larger weights to the elements whose indices are more likely to be outside of the support of sparse signal so that the received energy is not projected onto the corresponding bases, which promotes the sparseness of the solution. The simulations demonstrate that the proposed methods outperform the ℓ_1-SVD algorithm and some existing sparse recovery algorithms.

The outline of this paper is stated as follows. In the next section, we discuss SIT in terms of the sparse signal. In Section III , we formulate the weighted $\ell_{2,1}$-SVD algorithm. In Section Ⅳ, numerical experiments are provided for illustrating the performance of the proposed methods. A conclusion is given in Section Ⅴ.

II. SUPPORT INVARIANCE TRANSFORMATION

The measurements with the MMV of time series can be written as

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{n}(t), t = 1, 2, \cdots, T. \tag{1}$$

where $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_k, \cdots, \mathbf{a}_K] \in \mathrm{F}^{M\times K}$, $\mathrm{F} = \mathbb{C}$ or $\mathbb{R}$, is an overcomplete basis matrix, the vector $\mathbf{x}(t) \in \mathrm{F}^{K\times 1}$ is the jointly-sparse signals that the indices of non-zero rows of $\mathbf{x}(t)$ do not change with various sample time t [6], the vector $\mathbf{n}(t) \in \mathrm{F}^{M\times 1}$ denotes an additive noise vector with zero-mean and variance σ^2, especially we only consider the case for $T \geq M$ in this paper. Without loss of generality, the additive noise $\mathbf{n}(t)$ is assumed to be uncorrelated with the jointly-spares signals $\mathbf{x}(t)$.

Equation (1) can be expressed in matrix form:

$$\mathbf{Y} = \mathbf{AX} + \mathbf{N}. \quad (2)$$

The support of the joint sparse signals can be defined as [14]:

$$\text{Supp} = \{k| \ \mathbf{X}_k^{(\ell_2)} \neq 0\} \square \Lambda, \quad (3)$$

where $\mathbf{X}_k^{(\ell_2)}$ denotes the kth entry of $\mathbf{X}^{(\ell_2)}$, $\mathbf{X}^{(\ell_2)}$ is a column vector whose kth elements denotes the ℓ_2 norm of kth row of $\mathbf{X}$, $\Lambda \subseteq \{,\cdots,K\}$ and its cardinality $|\Lambda| = P$. Based on the definition of the support, (2) can be rewritten as [15], [16]

$$\mathbf{Y} = \mathbf{A}_\Lambda \mathbf{X}_\Lambda + \mathbf{N}, \quad (4)$$

where $\mathbf{A}_\Lambda$ denotes the matrix composed of the columns of $\mathbf{A}$ indexed by the set Λ and $\mathbf{X}_\Lambda$ is the matrix composed of the rows of $\mathbf{X}$ indexed by the set Λ.

By solving a Least Square (LS) problem, the solution $\mathbf{X}_\Lambda$ can be obtained [5], [6], [15], [16]

$$\mathbf{X}_\Lambda = (\mathbf{A}_\Lambda)^\dagger \mathbf{Y}, \quad (5)$$

where the sign $(\cdot)^\dagger$ denotes the Moore-Penrose pseudoinverse. Therefore, the core of the sparse signal reconstruction is how to determine the row support Λ. In practice, we can use some transformations that do not change the set Λ to obtain some gains. For example, in the MMV case the joint SVD processing achieves the joint-time processing that does not change the support of the jointly-sparse signals, meanwhile it reduces the number of problems from T to P for $T \geq M > P$.

Definition 1: Suppose that $\{\mathbf{x}(t), t = 1,\cdots,T\}$ is a sequence of jointly-sparse signals with a common support Λ and $\mathbf{X}$ is its matrix form. A transformation $\text{T}(\mathbf{X})$ is called as Support Invariance Transformation in terms of $\mathbf{X}$, if and only if $\text{Supp}(\text{T}(\mathbf{X})) \equiv \text{Supp}(\mathbf{X})$, where $\text{Supp}(\text{T}(\mathbf{X}))$ denote the support of $\text{T}(\mathbf{X})$.

In the following subsection, we prove the weighting processing is a SIT.

Proposition 1: The transformation $\text{T}_W(\mathbf{X}) = \mathbf{W}\mathbf{X}^{(\ell_2)}$ is a SIT in terms of $\mathbf{X}$, where $\mathbf{W}$ is a diagonal matrix and its diagonal elements $w_k > 0$.

Proof: Obviously, $w_k \mathbf{X}_k^{(\ell_2)} \neq 0$ for $\mathbf{X}_k^{(\ell_2)} \neq 0$ and $w_k \mathbf{X}_k^{(\ell_2)} = 0$ for $\mathbf{X}_k^{(\ell_2)} = 0$, i.e., $\text{Supp}(\text{T}_W(\mathbf{X})) = \Lambda$, where $\text{T}_W(\mathbf{X}) = \mathbf{W}\mathbf{X}^{(\ell_2)}$. □

As an example, the iterative reweighted transformation in [1] is a SIT in terms of $\mathbf{X}$ for $T = 1$. Its weights can be written as $w_k = 1/(|x_k| + \varepsilon)$, where $|x_k|$ and w_k denote the absolute value of the kth coefficient of the sparse signal and the corresponding weight value, the parameter ε can avoid the undefined weight value when x_k is the zero-valued component [1].

Based on the Proposition 1, we conclude that the transformation $\text{T}_W((\mathbf{X}))$ does not change the support of the jointly-sparse signals. Using this conclusion, we propose a unified approach to the weighted $\ell_{2,1}$ minimization to handle the recovery problem of the jointly-sparse signals in the following section.

III. WEIGHTED $\ell_{2,1}$-SVD ALGORITHM

In [7-9] the truncated Singular Value Decomposition (SVD) was exploited to hold the principal components on the measurements $\mathbf{Y}$:

$$\mathbf{Y}_{\text{SV}} = \mathbf{U\Sigma D}_P = \mathbf{YVD}_P, \quad (6)$$

where $\mathbf{Y} = \mathbf{U\Sigma V}^{\text{H}}$, the sign $(\cdot)^{\text{H}}$ denotes the conjugate transpose, the non-zero entries of $\mathbf{\Sigma} \in \square^{M\times T}$ are equal to the singular values of $\mathbf{Y}$; the columns of $\mathbf{U} \in \text{F}^{M\times M}$ and $\mathbf{V} \in \text{F}^{T\times T}$ are, respectively, left singular vectors and right singular vectors for corresponding singular values; $\mathbf{D}_P = [\mathbf{I}_P; \mathbf{0}]$ where $\mathbf{I}_P$ is a $P \times P$ identity matrix and $\mathbf{0}$ is a $(T - P) \times P$ matrix of zeroes. Consequently, the jointly-sparse signals $\mathbf{X}$ is transformed with $\text{T}_{SVD}(\mathbf{X})$ to reduce the computation complexity, where $\text{T}_{SVD}(\mathbf{X}) = \mathbf{X}_{\text{SV}} = \mathbf{XVD}_P$.

For the MMV case, the correlation matrix of the measurements can be obtained by the result of the SVD of the measurements. Then we try to exploit the Bartlett and Capon spectra estimate that utilize the correlation matrix or its inverse matrix to design the weighted matrix. The correlation matrix of the measurements can be written with the following equation

$$\mathbf{R} = \text{E}\{\mathbf{y}(t)\mathbf{y}^{\text{H}}(t)\} \square \frac{1}{T}(\mathbf{U\Sigma V})(\mathbf{U\Sigma V})^{\text{H}} = \frac{1}{T}\mathbf{U\Sigma}_{\mathbf{M}}^2\mathbf{U}^{\text{H}}, \quad (7)$$

where it is noted that $\mathbf{U}$ and $\mathbf{V}$ are the unitary matrix, the first M columns of $\mathbf{\Sigma}$ constitutes the diagonal matrix $\mathbf{\Sigma}_{\mathbf{M}}$.

The inverse of $\mathbf{R}$ can be expressed as:

$$\mathbf{R}^{-1} \square T(\mathbf{U\Sigma}_{\mathbf{M}}^2\mathbf{U}^{\text{H}})^{-1} = T\mathbf{U\Sigma}_{\mathbf{M}}^{-2}\mathbf{U}^{\text{H}}. \quad (8)$$

The output power of the Bartlett and Capon filers is given by [13]

$$p(\mathbf{a}_k) = \text{E}\{|\ \mathbf{h}^{\text{H}}\mathbf{y}(t)|^2\} = \text{E}\{|\ x_k(t)|^2\} + \sigma^2\mathbf{h}^{\text{H}}\mathbf{h}, \quad (9)$$

where $\mathbf{h}$ denotes the response of the filter, especially $\mathbf{h}_b = \mathbf{a}_k / M$ for Bartlett filter [13] and $\mathbf{h}_c = \mathbf{R}^{-1}\mathbf{a}_k / (\mathbf{a}_k^{\text{H}}\mathbf{R}^{-1}\mathbf{a}_k)$ for Capon filter [11], [13], $x_k(t)$ is the kth row of the joint-sparse signals $\mathbf{x}(t)$. Broadly speaking, the classical Bartlett and Capon methods can be interpreted as computing the value of the power of the measurements at the given basis $\mathbf{a}_k$. Therefore, the expected output power of the Bartlett filter at the given basis $\mathbf{a}_k$ can be written as [13]

$$p_b(\mathbf{a}_k) = \frac{1}{M^2}\mathbf{a}_k^{\mathrm{H}}\mathbf{R}\mathbf{a}_k \approx \frac{1}{TM^2}\mathbf{a}_k^{\mathrm{H}}\mathbf{U}\mathbf{\Sigma}_{\mathrm{M}}^2\mathbf{U}^{\mathrm{H}}\mathbf{a}_k. \quad (10)$$

Similarly, the expected output power of the Capon filter at the given basis can be express as [13]

$$p_c(\mathbf{a}_k) = \frac{1}{\mathbf{a}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{a}_k} \approx \frac{1}{\mathbf{a}_k^{\mathrm{H}}\mathbf{U}\mathbf{\Sigma}_{\mathrm{M}}^{-2}\mathbf{U}^{\mathrm{H}}\mathbf{a}_k}. \quad (11)$$

We define the weight on the kth row of the joint-sparse signals as

$$w_k = \frac{1}{\sqrt{Tp(\mathbf{a}_k)}} = \frac{1}{\sqrt{T(\mathrm{E}\{|x_k(t)|^2\} + \sigma^2\mathbf{h}^{\mathrm{H}}\mathbf{h})}}, \quad (12)$$

where $p(\mathbf{a}_k) = p_b(\mathbf{a}_k)$ for Bartlett filter or $p(\mathbf{a}_k) = p_c(\mathbf{a}_k)$ for Capon filter. Obviously, the item $\sigma^2\mathbf{h}^{\mathrm{H}}\mathbf{h}$ in (12) avoids the infinite weight phenomenon that is infeasible in practice. And then we can design a weighted matrix $\mathbf{W}$,

$$\mathbf{W} = diag\{w_1, \cdots, w_k, \cdots, w_K\}. \quad (13)$$

Because each weight $w_k > 0$ is the inverse of $\sqrt{Tp(\mathbf{a}_k)}$, larger weights are appointed to those elements whose indices are more likely to be outside of the support. It can avoid the phenomena of the spurious peaks that some energy is allotted to those positions where the indices of those bases should be outside of the support.

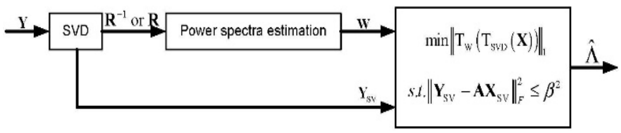

Figure 1. The entire flow of the proposed algorithm.

Now, we can find the sparse solution by the unified approach:

$$\min \| \mathrm{T}_W(\mathrm{T}_{SVD}(\mathbf{X}))\|_1, \ s.t. \| \mathbf{Y}_{\mathrm{SV}} - \mathbf{A}\mathbf{X}_{\mathrm{SV}}\|_{\mathrm{F}}^2 \le \beta^2, \quad (14)$$

where $\|\cdot\|_1 \triangleq \sum_i |[\cdot]_i|$ and $\beta^2 \ge \| \mathbf{N}\mathbf{V}\mathbf{D}_P\|_{\mathrm{F}}^2$ [8] is a regularization parameter (see [8] for details to determine the regularization parameter β^2). In practice, the indices of the P peaks in the solution $\mathbf{X}_{\mathrm{SV}}^{(\ell_2)}$ are chose as the estimate of the support set. We illustrate the entire flow of the proposed algorithm in Fig.1. We call the weighted $\ell_{2,1}$-SVD that uses the Bartlett filter as BW-$\ell_{2,1}$-SVD. Similarly, the weighted $\ell_{2,1}$-SVD is called as CW-$\ell_{2,1}$-SVD if the Capon filter is used, which was proposed in [17].

IV. SIMULATIONS

In this section, we use several experimental results to demonstrate our weighted $\ell_{2,1}$ minimization scheme for the MMV case. We consider a sparse matrix $\mathbf{X} \in \mathbb{C}^{K\times T}$ with P non-zero rows that their indices are chosen randomly, and the amplitude of each non-zero element is chosen randomly from a symmetric Bernoulli ± 1 distribution. The overcomplete basis matrix $\mathbf{A} \in \mathbb{C}^{M\times K}$ is a random matrix with i.i.d. Gaussian entries, and its columns are normalized. We select the indices corresponding to P peaks in the solution $\mathbf{X}_{\mathrm{SV}}^{(\ell_2)}$ as the estimate of the support Λ. Specially, the estimate of the row support is regarded as a successful estimate if and only if it is fully consistent with the true row support. We explore the strong recover threshold that a recover scheme can obtain the row support with certainty [5]. For example, for fixed M, K, P, and T, the lowest SNR that always obtains the correct estimate of the row support is called the strong recover threshold of SNR for the $\ell_{2,1}$ minimization. For illustrating the advantages of the proposed weighted scheme we compare the strong recover threshold obtained using the proposed algorithms with those of M-FOCUSS [6] and ℓ_1-SVD algorithm [7-9].

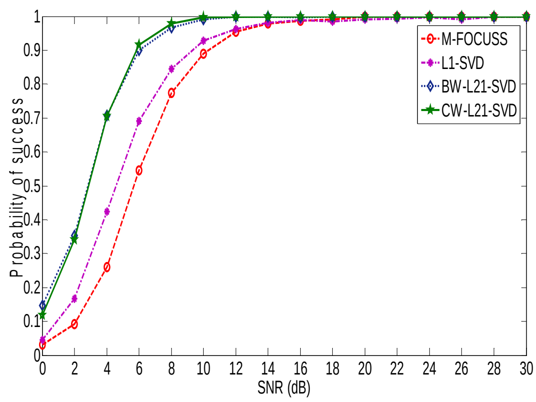

Figure 2. Probability of success versus SNR. Each point is average of 1000 Monte-Carlo trials and the number of time samples is 50.

In Fig.2, we consider the strong recover threshold of SNR. The simulation conditions are set as follow: $M = 10$, $K = 30$, $P = 5$, and $T = 50$. The strong recover threshold of SNR is 12dB for the proposed weighted $\ell_{2,1}$ minimization under the given conditions, while ℓ_1-SVD that solves a regular ℓ_1 minimization reach the strong recover threshold of SNR when SNR is more than 20dB, i.e., the requirement of SNR can be reduced due to the weighted scheme effect. It is worth mentioning that M-FOCUSS that solves a reweighted ℓ_2 minimization reach the strong recover threshold with higher SNR than the proposed methods.

Similarly, Candes *et al* also observed that reweighted ℓ_1 minimization is more powerful in recovering sparse signals than the FOCUSS algorithm because the unweighted ℓ_2 minimization has not the natural tendency of sparsity-promoting while unweighted ℓ_1 minimization does [1].

V. CONCLUSION

In this paper, we first define SIT and prove the weighted scheme does change the support. Then, we designed a unified framework for the weighted $\ell_{2,1}$ minimization to improve the performance of the ℓ_1-SVD algorithm. We used the strong recover threshold as the criteria to compare the proposed weighted $\ell_{2,1}$-SVD with other algorithms. Several advantages can be gained by using the proposed weighted scheme, e.g., decreased the requirement of SNR for reaching the strong recover threshold.

REFERENCES

[1] E.J. Candes, M.B. Wakin, and S.P. Boyd, "Enhancing sparsity by reweighted L_1 minimization," Journal of Fourier Analysis and Applications, vol. 14, no. 5, pp.877–905, 2008.

[2] D. Wipf and S. Nagarajan, "Iterative reweighted L_1 and L_2 methods for finding sparse solutions," IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 2, pp. 317–329, 2010.

[3] W. Xu, M.A. Khajehnejad, A.S. Avestimehr, and B. Hassibi, "Breaking through the thresholds: an analysis for iterative reweighted L_1 minimization via the grassmann angle framework," in 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2010), Mar. 2010, pp. 5498–5501.

[4] D. Needell, "Noisy signal recovery via iterativereweighted L_1 minimization," in Conference Recordof the Forty-Third Asilomar Conference on Signals, Systems and Computers, 2009, pp. 113–117.

[5] M.A. Khajehnejad, W. Xu, A.S. Avestimehr, and B. Hassibi, "Analyzing weighted L_1 minimization for sparse recovery with nonuniform sparse models," IEEE Transactions on Signal Processing, vol. 59, no. 5, pp.1985–2001, 2011.

[6] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problemswith multiple measurement vectors," *IEEE Transactionson Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[7] D.M. Malioutov, M. Cetin, and A.S. Willsky, "Source localization by enforcing sparsity through a laplacian prior: an svd-based approach," in *IEEE Workshop on Statistical Signal Processing, 2003*, 2003, pp. 573–576.

[8] D. Malioutov, M. Cetin, and A.S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[9] D.M. Malioutov, "A sparse signal reconstruction perspectivefor source localization with sensor arrays,"M.S. thesis, Mass. Inst. Technol., Cambridge, MA, 2003.

[10] J. Zheng, M. Kaveh, and H. Tsuji, "Sparse spectral fitting for direction of arrival and power estimation," in *IEEE/SP 15th Workshop on Statistical Signal Processing, 2009*, 2009, pp. 429–432.

[11] P. Stoica and R.L. Moses, *Spectral analysis of signals*, Pearson/Prentice Hall, Upper Saddle River, NJ, 2005.

[12] I.J. Fevrier, S.B. Gelfand, and M.P. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 927–937, 1999.

[13] S.S. Haykin and K.J.R Liu, *Handbook on array processing and sensor networks*, Wiley Online Library, Hoboken, NJ, 2009.

[14] E. Berg and M.P. Friedlander, "Joint-sparse recovery from multiple measurements," *Arxiv preprint arXiv: 0904.2051*, 2009.

[15] M. Mishali and Y.C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4692–4702, 2008.

[16] B. Mailh`e, R. Gribonval, P. Vandergheynst, and F. Bimbot, "Fast orthogonal sparse approximation algorithms over local dictionaries," *Signal Processing*, vol. 91, pp. 2822–2835, 12 2011.

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