

An Iteratively Reweighted Method for Recovery of Block-Sparse Signal with Unknown Block Partition

Qi He, Jun Fang, *Member, IEEE*, Zhi Chen, *Member, IEEE*, and Shaoqian Li, *Fellow, IEEE*
National Key Laboratory on Communications, University of Electronic Science and Technology of China
heqi.tech@hotmail.com; {JunFang, chen zhi, lsq}@uestc.edu.cn

Abstract—In this paper, a new iteratively reweighted least squares method is proposed for recovery of block-sparse signals with unknown cluster patterns. In many practical applications, sparse signals have block-sparse structures with nonzero coefficients occurring in clusters, while the prior information of the cluster pattern is usually unavailable. To address this issue, we propose an element-overlapping log-sum functional to encourage the sparseness and the cluster pattern simultaneously. The algorithm is developed by iteratively minimizing a convex surrogate function that majorizes the original objective function, which results in an iteratively reweighted process that alternates between estimating the sparse signal and refining the weights of the surrogate function. Convergence of the iterations to a local minimum of the penalty function is also guaranteed. Numerical results are provided to illustrate the effectiveness of the proposed method.

Index Terms—Block-sparse signal recovery, iteratively reweighted, element-overlapping log-sum functional.

I. INTRODUCTION

Sparse signal recovery and compressed sensing have drawn significant attention in recent years [1]. The basic form of the sparse signal recovery is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^{M \times 1}$ is a measurement vector, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is a sampling matrix with $M < N$, and $\mathbf{x} \in \mathbb{R}^{N \times 1}$ is a sparse signal with only K nonzero coefficients. Besides sparsity, signals usually exhibit block-sparse structures in many applications, such as multi-band signals [2], [3], audio signals [4], and gene expression analysis [5]. This cluster pattern can be utilized to considerably enhance the recovery performance.

A number of algorithms, such as block-OMP [6], group LASSO [7], mixed ℓ_2/ℓ_1 norm-minimization [8], group basis pursuit [9], and block-sparse Bayesian learning (BSBL) [10] were proposed for recovery of block-sparse signals. However, these algorithms require knowledge of block partition *a priori*, which is usually unavailable in practice. StructOMP [11] does not need to know the block partition, but it requires the number of nonzero coefficients of the sparse signal. To address this difficulty, a few algorithms that do not require the knowledge of the block partition were developed recently. In [12], a Boltzman machine-based greedy pursuit (BM-MAP-OMP) method which employs the Boltzmann machine

as a prior on the support was developed to recover block-sparse signals. In [13], a cluster-structured Markov chain Monte Carlo (CluSS-MCMC) algorithm was proposed by employing a hierarchical Bayesian “spike-and-slab” prior model to encourage sparsity and promote a block-sparse structure simultaneously. A modified block sparse Bayesian learning (referred to as EBSBL) method was proposed in [14] to address the scenario where the block partition is unknown, in which an expanded model is introduced by assuming that the original sparse signal is a superposition of a number of overlapping blocks. In [15], a pattern-coupled hierarchical Gaussian prior model was proposed to exploit the statistical pattern dependencies among neighboring coefficients. Numerical results show that the pattern-coupled sparse Bayesian learning (PC-SBL) method renders competitive performance for block-sparse signal recovery.

In this paper, a new element-overlapping log-sum penalty functional is proposed for recovery of block-sparse signals with unknown cluster patterns. Unlike the conventional log-sum functional where each coefficient is associated with an individual logarithmic function, consecutive neighboring coefficients are grouped together to encourage structured-sparse solutions. Meanwhile, overlapping elements are used in different logarithmic functions to provide a flexible framework to model any block-sparse patterns. An iteratively reweighted method is developed by resorting to the majorization-minimization (MM) approach for block-sparse signal recovery. The proposed method is developed by iteratively decreasing a surrogate function that majorizes the original objective function. Numerical results are provided to illustrate the effectiveness of the proposed method.

II. ELEMENT-OVERLAPPING LOG-SUM FUNCTIONAL

The following log-sum minimization has been extensively used for sparse signal recovery [16], [17].

$$\min_{\mathbf{x}} \sum_{i=1}^N \log(|x_i| + \epsilon), \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \quad (2)$$

where x_i denotes the i th element of \mathbf{x} , and $\epsilon > 0$ is a regularization parameter to ensure that the function is well-defined. It was shown empirically [17] and theoretically [18] that log-sum minimization presents universal superiority over conventional ℓ_1 -type methods. Nevertheless, the above log-sum functional has no potential to encourage structured-sparse solutions. To promote a block-sparse solution, one

This work was supported by the National Science Foundation of China under Grant 61172114, and by the Sichuan Province Youth Fund under Grant 2014JQ0009, and by the Major Research Plan of the National Natural Science Foundation of China under Grant 91438118.

may wish to group together those coefficients which share a same sparsity pattern and place a sparse prior on each group, just like the group LASSO [7] and the mixed ℓ_2/ℓ_1 -norm method [8] do. This, however, requires the knowledge of block partition of the signal to determine which coefficients should be put together, which is usually unavailable in practice.

To address this difficulty, we propose to minimize an element-overlapping log-sum functional as follows

$$\min_{\mathbf{x}} L(\mathbf{x}) \triangleq \sum_{i=0}^N \log(x_i^2 + x_{i+1}^2 + \epsilon), \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \quad (3)$$

where x_0 and x_{N+1} are created to simplify our expression and both are set equal to 0. Clearly, this log-sum functional encourages a structured-sparse solution since pairs of neighboring coefficients are grouped together and assigned to each logarithmic function. On the other hand, due to the use of overlapping elements in different logarithmic functions, the above log-sum functional does not impose any pre-defined structures on the signal to be recovered, thus providing a flexible framework to model any cluster patterns.

III. PROPOSED ALGORITHM

In this section, a bounded optimization algorithm, belonging to the general class of majorization-minimization (MM) approaches, is developed to minimize the new penalty function $L(\mathbf{x})$ in (3). The main idea is to iteratively decrease a simple surrogate function, by which $L(\mathbf{x})$ is upper-bounded. An appropriate choice of such a surrogate function has a convex quadratic form and can be given by

$$Q(\mathbf{x}, \boldsymbol{\gamma}) \triangleq \sum_{i=0}^N \left(\frac{x_i^2 + x_{i+1}^2 + \epsilon}{\gamma_i} + \log \gamma_i - 1 \right), \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \quad (4)$$

where $\boldsymbol{\gamma} \triangleq [\gamma_0, \dots, \gamma_N]^T$, and $\gamma_i > 0, \forall i$.

It can be readily verified that

$$Q(\mathbf{x}, \boldsymbol{\gamma}) - L(\mathbf{x}) \geq 0 \quad (5)$$

with equality if and only if $\gamma_i = x_i^2 + x_{i+1}^2 + \epsilon, \forall i$.

Now we arrive at finding the local minimum of the surrogate function (4) with regard to \mathbf{x} and $\boldsymbol{\gamma}$.

For fixed $\boldsymbol{\gamma}$,

$$Q(\mathbf{x}; \boldsymbol{\gamma}) = \mathbf{x}^T \boldsymbol{\Gamma}^{-1} \mathbf{x} + \text{const}, \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \quad (6)$$

where

$$\boldsymbol{\Gamma} = \text{diag}\{\Gamma_{ii}\}, \quad \Gamma_{ii} = \frac{1}{(\gamma_{i-1})^{-1} + (\gamma_i)^{-1}} \quad (7)$$

By resorting to the standard method of Lagrange multipliers, the optimal solution of (6) can be given explicitly (see [19] for more details)

$$\hat{\mathbf{x}} = \boldsymbol{\Gamma} \mathbf{A}^T (\mathbf{A} \boldsymbol{\Gamma} \mathbf{A}^T)^{-1} \mathbf{y} \quad (8)$$

For fixed \mathbf{x} , (4) is minimized by

$$\hat{\gamma}_i = x_i^2 + x_{i+1}^2 + \epsilon, \quad \forall i \quad (9)$$

Consequently, by iteratively updating \mathbf{x} with (7) (8) and updating $\boldsymbol{\gamma}$ with (9), the surrogate function $Q(\mathbf{x}, \boldsymbol{\gamma})$ are reduced or left unchanged, which results in a non-increasing value of the penalty function $L(\mathbf{x})$ as well. This procedure can be shown as below

$$\begin{aligned} L(\hat{\mathbf{x}}^{(t)}) &= Q(\hat{\mathbf{x}}^{(t)}, \hat{\boldsymbol{\gamma}}^{(t)}) \\ &\leq Q(\hat{\mathbf{x}}^{(t)}, \hat{\boldsymbol{\gamma}}^{(t-1)}) \\ &\leq Q(\hat{\mathbf{x}}^{(t-1)}, \hat{\boldsymbol{\gamma}}^{(t-1)}) \\ &= L(\hat{\mathbf{x}}^{(t-1)}) \end{aligned} \quad (10)$$

where $\hat{\mathbf{x}}^{(t)}$ and $\hat{\boldsymbol{\gamma}}^{(t)}$ denote the estimation of \mathbf{x} and $\boldsymbol{\gamma}$ in the (t) th iteration, respectively. According to Global Convergence Theorem [19], $L(\mathbf{x})$ is guaranteed to converge monotonically to a local minimum (or a saddle point). A tiny perturbation leads to a local minimum when a saddle point is reached, which is very rare to occur [20].

By substituting (7) and (9) in (6), we eliminate the parameter $\boldsymbol{\gamma}$ and obtain

$$\Gamma_{ii} = \left(\frac{1}{x_{i-1}^2 + x_i^2 + \epsilon} + \frac{1}{x_i^2 + x_{i+1}^2 + \epsilon} \right)^{-1} \quad (11)$$

Thus, each iteration of the proposed method reduces to updating \mathbf{x} with (8) and $\boldsymbol{\Gamma}$ with (11) sequentially, which operates in the same way as the iteratively reweighted least squares (IRLS) method [21], with Γ_{ii}^{-1} seen as the ‘‘weight’’ of x_i . Note that in IRLS method each coefficient x_i is assumed independent from each other and the weight of x_i is determined by the previous estimation of x_i individually. However, as can be seen from (11), Γ_{ii}^{-1} , the ‘‘weight’’ of x_i , not only involves the previous estimation of x_i , but also its immediate neighboring coefficients x_{i-1} and x_{i+1} . Since the amplitude of the weight reflects how much we want to encourage a coefficient to become a zero component [20], it means that the neighboring coefficients x_{i-1} and x_{i+1} also have an impact on the sparsity pattern of x_i .

In this way, a coupling mechanism among neighboring coefficients is established, and the iterative reweighted algorithm has the potential to encourage block-sparse solutions. This coupling effect enables to recover block-sparse signals in a more reliable way. Note that due to the pattern coupling, sporadic recovery errors which misidentify a nonzero coefficient x_i (located in a nonzero block) as an isolated zero component are almost impossible to happen. It can be explained as follows, in the (t) th iteration, even if $\hat{x}_i^{(t)}$ is small, its associated weight $(\hat{\Gamma}_{ii}^{-1})^{(t)}$ will not become arbitrarily large due to the neighboring nonzero coefficients $\hat{x}_{i-1}^{(t)}$ and $\hat{x}_{i+1}^{(t)}$. Hence, in the $(t+1)$ th iteration the coefficient $\hat{x}_i^{(t+1)}$ will not be suppressed to be zero.

In the iterative procedure, undesirable local minima are likely to be reached when $\epsilon \rightarrow 0$. To address this issue, similarly to [21], we use a slowly decreasing sequence $\{\epsilon\}$ in the minimizing procedure, instead of a constant ϵ . This strategy provides a stable coefficient estimate at the very beginning iterations, and does not deflect the local minimum point at the last iterations when ϵ is so small that can be

ignored. Our experimental evidence shows that this strategy is far more effective in avoiding undesirable local minima troubles.

For clarity, we summarize the proposed method in the noiseless situation as follows.

- 1) Given an initialization of $\Gamma^{(0)}$.
- 2) At iteration $t = 1, 2, \dots$: obtain $\hat{\mathbf{x}}^{(t)}$ by (8), obtain $\hat{\Gamma}^{(t)}$ by (11), and decrease ϵ by some strategy, in turn.
- 3) Stop if $\|\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{x}}^{(t-1)}\|_2^2 < \epsilon$, where ϵ is a constant tolerance value; otherwise go to Step 2.

IV. EXTENSION TO THE NOISY CASE

In the preceding discussions, we covered the proposed functional and corresponding algorithm in the noiseless scenario. Next we propose an expanded version of our method in the noisy scenario.

Similarly to (1), the basic model in the noisy scenario is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \quad (12)$$

where $\mathbf{w} \in \mathbb{R}^{M \times 1}$ is an unknown multivariate Gaussian noise vector. The corresponding element-overlapping log-sum functional is proposed as

$$\min_{\mathbf{x}} L(\mathbf{x}) \triangleq \sum_{i=0}^N \log(x_i^2 + x_{i+1}^2 + \epsilon) + \lambda \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad (13)$$

where $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$ that represents the fitting error is added to make the function more robust to noise [20], [22]. λ is a *regularization parameter* that controls the tradeoff between the sparsity of the solution and the quality of fit. From (4), (5) in Section II, we can directly obtain the surrogate function in the noisy scenario,

$$Q(\mathbf{x}, \gamma) \triangleq \sum_{i=0}^N \left(\frac{x_i^2 + x_{i+1}^2 + \epsilon}{\gamma_i} + \log \gamma_i - 1 \right) + \lambda (\mathbf{A}\mathbf{x} - \mathbf{y})^T (\mathbf{A}\mathbf{x} - \mathbf{y}) \quad (14)$$

For fixed γ ,

$$Q(\mathbf{x}; \gamma) = \mathbf{x}^T \mathbf{\Gamma}^{-1} \mathbf{x} + \lambda (\mathbf{A}\mathbf{x} - \mathbf{y})^T (\mathbf{A}\mathbf{x} - \mathbf{y}) + \text{const} \quad (15)$$

where $\mathbf{\Gamma}$ still satisfies the identities in (7). By taking the derivative of (15) and setting it to zero, we obtain the optimal solution of (15)

$$\hat{\mathbf{x}} = \lambda \hat{\Sigma} \mathbf{A}^T \mathbf{y} \quad (16)$$

where $\hat{\Sigma} = (\mathbf{\Gamma}^{-1} + \lambda \mathbf{A}^T \mathbf{A})^{-1}$.

For fixed \mathbf{x} , the optimal γ is obtained by (9) as well.

The parameter $\mathbf{\Gamma}$ is deduced as the same as in the noiseless scenario. Consequently, for each iteration in the noisy situation, we update \mathbf{x} by (16), update $\mathbf{\Gamma}$ by (11), and decrease ϵ by some strategy sequentially. Convergence to a local minimum or a saddle point is also guaranteed similarly to the noiseless scenario.

As $\mathbf{\Gamma}$ is updated in the same way as the noiseless scenario, the minimizing procedure with noise relates the sparsity patterns of neighboring coefficients to each other as well.

It can be concluded that neighboring elements are related to each other in $\mathbf{\Gamma}$ -space, and this pattern-coupling effect feedback to \mathbf{x} -space by (16) in the noisy scenario. Therefore, with noise the benefits brought by pattern-coupling effect between adjacent coefficients still exist.

For each iteration, the optimal solution $\hat{\mathbf{x}}$ in the noisy scenario, formulated in (16), can be re-expressed as

$$\hat{\mathbf{x}} = \mathbf{\Gamma} \mathbf{A}^T (\lambda^{-1} \mathbf{I}_M + \mathbf{A} \mathbf{\Gamma} \mathbf{A}^T)^{-1} \mathbf{y} \quad (17)$$

which is identical to its noiseless counterpart (8) when λ approaches infinity. An infinite λ represents that the quality of fit dominates the recovery procedure, which is the same as in the noiseless scenario.

Furthermore, the choice of λ has great impact on the performance [20]. Nevertheless, to our knowledge, determining a proper value for λ still remains an implementation-level problem. In addition, it appears to be more appropriate having a value dependent on the iterations rather than limiting λ to an arbitrarily fixed value. Several approaches, e.g. Modified L-Curve Method, have been proposed in [22] to choose a proper λ dynamically. However, we propose a different empirical strategy here. As revealed in [16], a correspondence relationship exists between iteratively reweighted class methods and sparse Bayesian learning (SBL) methods [23]. Inspired by this insight, we use the inverse of the variance estimation of the noise vector \mathbf{w} in SBL methods as the estimation of λ , in each iteration, shown as below.

$$\hat{\lambda}^{(t)} = \frac{M}{\left\| \mathbf{A} \hat{\mathbf{x}}^{(t)} - \mathbf{y} \right\|_2^2 + \hat{\lambda}_{(t-1)}^{-1} \sum_{i=1}^N (1 - \hat{\Sigma}_{ii}^{(t)} / \hat{\Gamma}_{ii}^{(t-1)})} \quad (18)$$

where $\hat{\lambda}^{(t)}$ indicates the estimation of λ in the (t) th iteration. Empirical results show the superiority of our λ -estimation strategy over other methods.

V. SIMULATION RESULTS

In this section, we carry out experiments to illustrate the performance of the proposed method, also referred to as the pattern-coupled iteratively reweighted least squares (PC-IRLS) method, and its comparisons with other existing methods. In the following experiments, similarly to [21], we empirically update the factor ϵ . At the beginning, ϵ is set to a relatively large value of 1. ϵ is reduced by a factor of 10 when the change of the optimal solution (8) or (16) in relative 2-norm from the previous iteration is less than $\sqrt{\epsilon}/100$. This process is continued until ϵ attains a minimum of 10^{-8} .

As intra-block correlation exists in many practical applications, experiments are carried out respectively when the intra-block correlation equals 0 and equals 0.95. We compare the proposed method with two recently developed methods, termed EBSBL [14] and PCSBL [15], for the block-sparse signal recovery. The methods SBL [23] and IRLS [21] are also included for comparison. As performances of the methods BM-MAP-OMP [12] and CluSS-MCMC [13], both of which do not need the knowledge of block partitions as

well, were found less well than EBSBL or PCSBL [14], [15], neither of them is added to the comparisons in this paper.

A. noiseless scenario

We first consider the block-sparse signal recovery in the noiseless scenario. Similarly to [14], [15], the K nonzero coefficients in \mathbf{x} are partitioned into L blocks with random block sizes and random locations. For each trial, the block number L is drawn from $[2, \frac{2K}{h} - 2]$ uniformly, where h denotes the average size of nonzero blocks in \mathbf{x} . h is also used as a parameter for EBSBL.

In the noiseless situation, the recovery performance is evaluated in term of the success rate, which is computed as the percentage of successful trials in the total 400 independent trials. A trial is considered successful if the normalized squared error $\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 / \|\mathbf{x}\|_2^2$ is no greater than 10^{-6} . For SBL-based methods, the estimation of noise variance is fixed to a tiny value of 10^{-10} in order to yield satisfactory performance in the noiseless scenario.

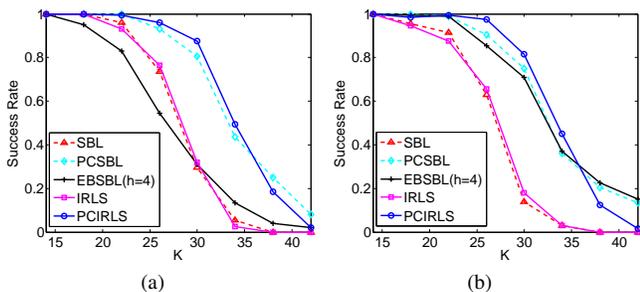


Fig. 1. Success Rates of respective algorithms vs. the sparsity level K in the noiseless scenario, where $N = 100, M = 50$ and $h = 4$. The intra-block correlation is set to be (a) 0 and (b) 0.95.

From Fig.1(a), we can see that PCIRLS and PCSBL significantly outperform the other three algorithms when intra-block correlation does not exist. It proves that the coupling of neighboring elements indeed brings benefit for block-sparse signal recovery. EBSBL perform less well in Fig.1(a) for the intra-block correlation it exploits does not exist. When elements within each nonzero block are highly correlated, as shown in Fig.1(b), PCIRLS and PCSBL exhibit similar performances with EBSBL because PCIRLS and PCSBL work by exploiting analogous neighboring relationship.

B. noisy scenario

In this subsection we consider the noisy scenario. \mathbf{x} is generated in the same way as the noiseless scenario, and the white Gaussian noise vector \mathbf{w} is added such that the signal-to-noise ratio (SNR), which is defined as $\text{SNR}(\text{dB}) \triangleq 20\log_{10}(\|\mathbf{A}\mathbf{x}\|_2 / \|\mathbf{w}\|_2)$, is constant for each trial. The normalized mean squared error (NMSE), which is calculated by averaging normalized squared errors of 400 runs, is used to evaluate the recovery performance. For PCIRLS, the λ -updating strategy in (18) is adopted instead of using a fixed λ , which tends to produce a result heavily dependent on the choice of λ .

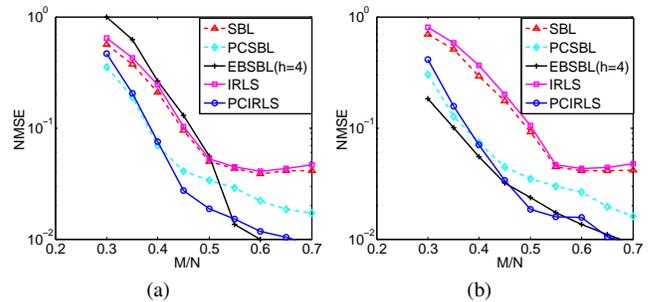


Fig. 2. NMSE of respective algorithms vs. M/N in the noisy scenario, where $N = 100, K = 20, h = 4$ and $\text{SNR} = 20 \text{ dB}$. The intra-block correlation is set to be (a) 0 and (b) 0.95.

It can be observed from Fig.2 that among all the algorithms PCIRLS exhibits one of the best performances when the ratio M/N is high and good performance when M/N is low, both with and without intra-block correlation.

C. audio data scenario

Experiments on real-world audio data are carried out in this subsection. Audio signals have block-sparse structures in certain basis, such as discrete cosine transform (DCT) basis. Similarly to [15], we consider a clean piano signal¹. For each of the total 400 trials, we randomly select a short-term segment that consists of $N = 200$ data samples from the audio signal. A compressing matrix $\mathbf{Q} \in \mathbb{R}^{M \times N}$ is generated randomly for each trial as well. The measurement matrix \mathbf{A} can be expressed as $\mathbf{A} = \mathbf{Q}\Psi$, where $\Psi \in \mathbb{R}^{N \times N}$ represents the DCT basis. The short-term segment is then reconstructed by respective algorithms. It can be observed from Fig.3 that PCIRLS provides one of the best performances for recovery of audio signals.

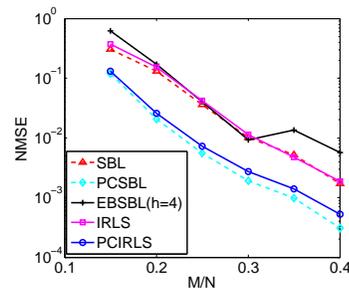


Fig. 3. NMSE of respective algorithms vs. M/N for audio signal recovery where $N = 200$.

VI. CONCLUSION

We proposed a pattern-coupled iteratively reweighted least squares method, originated from an element-overlapping log-sum functional, to recover block-sparse signals with unknown block partitions. This method exploits the relationship between adjacent elements to encourage structured sparsity, ignoring details of cluster patterns. Convergency of the corresponding optimization procedure is guaranteed. Simulation results demonstrate that the proposed method provides state-of-the-art performance for block-sparse signal recovery with unknown block partition.

¹Available at <http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html>

REFERENCES

- [1] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer Science & Business Media, 2010.
- [2] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, 2008.
- [3] —, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, 2009.
- [4] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 101–111, 2003.
- [5] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. R. Statist. Soc. B*, vol. 67, no. 1, pp. 91–108, 2005.
- [6] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, 2010.
- [7] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Statist. Soc. B*, vol. 68, pp. 49–67, 2006.
- [8] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.
- [9] E. V. D. Berg and M. Friedlander, "Probing the Pareto Frontier for Basis Pursuit Solutions," *SIAM J. Scientif. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.
- [10] Z. Zhang and B. D. Rao, "Sparse Signal Recovery With Temporally Correlated Source Vectors Using Sparse Bayesian Learning," *IEEE J. Sel. Topics. Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.
- [11] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 3371–3412, 2011.
- [12] T. Peleg, Y. C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2286–2303, 2012.
- [13] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Process.*, vol. 92, no. 1, pp. 259–269, 2012.
- [14] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, 2013.
- [15] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, 2015.
- [16] D. Wipf and S. Nagarajan, "Iterative Reweighted ℓ_1 and ℓ_2 Methods for Finding Sparse Solutions," *IEEE J. Sel. Topics. Signal Process.*, vol. 4, no. 2, pp. 317–329, 2010.
- [17] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [18] Y. Shen, J. Fang, H. Li, and S. Member, "Exact Reconstruction Analysis of Log-Sum Minimization for Compressed Sensing," *IEEE Signal Process. Letters*, vol. 20, no. 12, pp. 1223–1226, 2013.
- [19] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, 1999.
- [20] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, 1997.
- [21] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 3869–3872, 2008.
- [22] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 760–770, 2003.
- [23] M. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.