Approximate Message Passing with Nearest Neighbor Sparsity Pattern Learning

Xiangming Meng, Sheng Wu, Linling Kuang, Defeng (David) Huang, and Jianhua Lu, *Fellow, IEEE*

***Abstract*—We consider the problem of recovering clustered sparse signals with no prior knowledge of the sparsity pattern. Beyond simple sparsity, signals of interest often exhibits an underlying sparsity pattern which, if leveraged, can improve the reconstruction performance. However, the sparsity pattern is usually *unknown* a priori. Inspired by the idea of k-nearest neighbor (k-NN) algorithm, we propose an efficient algorithm termed approximate message passing with nearest neighbor sparsity pattern learning (AMP-NNSPL), which learns the sparsity pattern adaptively. AMP-NNSPL specifies a flexible spike and slab prior on the unknown signal and, after each AMP iteration, sets the sparse ratios as the average of the nearest neighbor estimates via expectation maximization (EM). Experimental results on both synthetic and real data demonstrate the superiority of our proposed algorithm both in terms of reconstruction performance and computational complexity.**

***Index Terms*—Compressed sensing, structured sparsity, approximate message passing, k-nearest neighbor.**

I. Introduction

Compressed sensing (CS) aims to accurately reconstruct sparse signals from undersampled linear measurements [1]–[3]. To this end, a plethora of methods have been studied in the past years. Among others, approximate message passing (AMP) [4] proposed by Donoho *et al.* is one state-of-the-art algorithm to address sparse signal reconstruction in CS. Moreover, AMP has been extended to Bayesian AMP (B-AMP) [5], [6] and general linear mixing problems [7]–[9]. While many practical signals can be described as sparse, they often exhibit an underlying structure, e.g., the nonzero coefficients occur in clusters [10]–[16]. Exploiting such intrinsic structure beyond simple sparsity can significantly boost the reconstruction performance [14]–[16]. To this end, various algorithms have been proposed, e.g., group LASSO [10], StructOMP [17], Graph-CoSaMP [18], and block sparse Bayesian learning (B-SBL) [19]–[21], etc. However, these algorithms require knowledge of sparsity pattern which is usually *unknown* a priori. To reconstruct sparse signals with unknown structure, a number of methods [22]–[28] have been developed to use various structured priors to encourage both sparsity and cluster patterns simultaneously. The main effort of these algorithms lies in constructing a hierarchical prior model, e.g., Markov tree [23], structured spike and slab [24], [25], hierarchical Gamma-Gaussian [26]–[28] to encode the structured sparsity pattern.

In this letter, we take an alternative approach and propose an efficient message passing algorithm, termed AMP with nearest neighbor sparsity pattern learning (AMP-NNSPL), to recover clustered sparse signals adaptively, i.e., without any prior knowledge of the sparsity pattern. For clustered sparse signals, if the nearest neighbors of one element are zeros (nonzeros), it will tend to be zero (nonzero) with high probability, a similar idea of k-nearest neighbor (k-NN) algorithm which assumes that data close together more likely belong to the same category [29], [30]. Therefore, instead of explicitly modeling the sophisticated sparsity pattern, AMP-NNSPL specifies a flexible spike and slab prior on the unknown signal and, after each AMP iteration, updates the sparse ratios as the average of their nearest neighbor estimates via expectation maximization (EM) [31]. In this way, the sparsity pattern is learned adaptively. Simulations results on both synthetic and real data demonstrate the superiority of our proposed algorithm both in terms of reconstruction performance and computational efficiency.

II. System Model

Consider the following linear Gaussian model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^N$ is the unknown signal, $\mathbf{y} \in \mathbb{R}^M$ is the available measurements, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the known measurement matrix, and $\mathbf{w} \in \mathbb{R}^M \sim \mathcal{N}\big(\mathbf{w}; 0, \Delta_0\mathbf{I}\big)$ is the additive noise. $\mathcal{N}\big(\mathbf{x}; \mathbf{m}, \mathbf{C}\big)$ denotes a Gaussian distribution of $\mathbf{x}$ with mean $\mathbf{m}$ and covariance $\mathbf{C}$ and $\mathbf{I}$ denotes the identity matrix. Our goal is to estimate $\mathbf{x}$ from $\mathbf{y}$ when $M \ll N$ and $\mathbf{x}$ is clustered sparse while its specific sparsity pattern is unknown a priori.

This work was partially supported by the National Nature Science Foundation of China (Grant Nos. 91338101, 91438206, and 61231011), the National Basic Research Program of China (Grant No. 2013CB329001).

X. Meng and J. Lu are with the Department of Electronic Engineering, Tsinghua University, Beijing, China. (e-mail: mengxm11@mails.tsinghua.edu.cn; lhh-dee@mail.tsinghua.edu.cn).

S. Wu and L. Kuang are with the Tsinghua Space Center, Tsinghua University, Beijing, China. (e-mail: thuraya@mail.tsinghua.edu.cn; kll@mail.tsinghua.edu.cn).

Defeng (David) Huang is with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Australia (e- mail: huangdf@ee.uwa.edu.au).

To enforce sparsity, from a Bayesian perspective, the signals are assumed to follow sparsity-promoting prior distributions, e.g., Laplace prior [32], automatic relevance determination [33], and spike and slab prior [6], [34]. In this letter we consider a flexible spike and slab prior of the form

$$p_0(\mathbf{x}) = \prod_{i=1}^{N} p_0(x_i) = \prod_{i=1}^{N} \big[(1-\lambda_i)\delta(x_i) + \lambda_i f(x_i)\big], \quad (2)$$

where $\lambda_i \in (0,1)$ is the sparse ratio, i.e., the probability of x_i being nonzero, $\delta(x_i)$ is the Dirac delta function, $f(x_i)$ is the distribution of the nonzero entries in $\mathbf{x}$, e.g., $f(x_i) = \mathcal{N}(x_i; \mu_0, \tau_0)$ for sparse Gaussian signals and $f(x_i) = \delta(x_i - 1)$ for sparse binary signals, etc.

It is important to note that in (2) we specify an individual λ_i for each entry, as opposed to a common value in [6], [34]. This is a key feature that will be exploited by the proposed algorithm for reconstruction of structured sparse signals. Up to now, it seems that no structure is ever introduced to enforce the underlying sparsity pattern. Indeed, if the sparse ratios $\lambda_i, i = 1, \ldots, N$ are learned independently, we will not benefit from the potential structure. The main contribution of this letter is a novel adaptive learning method which encourages clustered sparsity, as descried in Section III.

III. Proposed Algorithm

In this section, inspired by the idea of k-NN, we propose an adaptive reconstruction algorithm to recover clustered sparse signals without any prior knowledge of the sparsity pattern, e.g., structure and sparse ratio.

Before proceeding, we first give a brief description of AMP. Generally, AMP decouples the vector estimation problem (1) into N scalar problems in the asymptotic regime [35], [36]

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \longrightarrow \begin{cases} R_1 = x_1 + \tilde{w}_1 \\ \vdots \\ R_N = x_N + \tilde{w}_N \end{cases}, \quad (3)$$

where the effective noise $\tilde{w}_i$ asymptotically follows $\mathcal{N}(\tilde{w}_i; 0, \Sigma_i)$. The values of R_i, Σ_i are updated iteratively in each AMP iteration (see Algorithm 1) and the posterior distribution of x_i is estimated as

$$q(x_i|R_i, \Sigma_i) = \frac{1}{Z(R_i, \Sigma_i)} p_0(x_i) \mathcal{N}(x_i; R_i, \Sigma_i), \quad (4)$$

where $Z(R_i, \Sigma_i)$ is the normalization constant. From (4), the estimates of the mean and variance of x_i are

$$g_a(R_i, \Sigma_i) = \int x_i q(x_i|R_i, \Sigma_i) dx_i, \quad (5)$$

$$g_c(R_i, \Sigma_i) = \int x_i^2 q(x_i|R_i, \Sigma_i) dx_i - g_a^2(R_i, \Sigma_i). \quad (6)$$

For more details of AMP and its extensions, the readers are referred to [4]–[6], [35]. Two problems arise in traditional AMP. First, it assumes full knowledge of the prior distribution and noise variance, which is an impractical assumption. Second, it does not account for the potential structure of sparsity. In the sequel, we resort to expectation maximization (EM) to learn the unknown hyperparameters. Further, to encourage structured sparsity, we develop a nearest neighbor sparsity pattern learning rule motivated by the idea of k-NN algorithm. For lack of space, we only consider the sparse Gaussian case, $f(x_i) = \mathcal{N}(x_i; \mu_0, \tau_0)$, while generalization to other settings is possible.

The hidden variables are chosen as the unknown signal vector $\mathbf{x}$ and the hyperparameters are denoted by $\boldsymbol{\theta}$. The specific definition of $\boldsymbol{\theta}$ depends on the choice of distribution $f(x)$ in (2). In the Gaussian case, $\boldsymbol{\theta} = \{\mu_0, \tau_0, \Delta_0, \lambda_i, i = 1, \ldots, N\}$ while in the binary case, $\boldsymbol{\theta} = \{\Delta_0, \lambda_i, i = 1, \ldots, N\}$. Denote by $\boldsymbol{\theta}^t$ the estimate of hyperparameters at the tth EM iteration, then EM alternates between the following two steps [31]

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = \mathsf{E}\Big\{\ln p(\mathbf{x}, \mathbf{y}) | \mathbf{y}; \boldsymbol{\theta}^t\Big\}, \quad (7)$$

$$\boldsymbol{\theta}^{t+1} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t), \quad (8)$$

where $\mathsf{E}\{\cdot|\mathbf{y}; \boldsymbol{\theta}^t\}$ denotes expectation conditioned on observations $\mathbf{y}$ with parameters $\boldsymbol{\theta}^t$, i.e., the expectation is with respect to the posterior distribution $p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}^t)$. From (1), (2), the joint distribution $p(\mathbf{x}, \mathbf{y})$ in (7) is defined as

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x}) \prod_i (1 - \lambda_i)\delta(x_i) + \lambda_i f(x_i), \quad (9)$$

where $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{A}\mathbf{x}, \Delta_0 \mathbf{I})$. AMP offers an efficient approximation of $p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}^t)$, denoted as $q(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}^t) = \prod_i q(x_i|R_i, \Sigma_i)$, whereby the E step (7) can be efficiently calculated. Since joint optimization of $\boldsymbol{\theta}$ is difficult, we adopt the incremental EM update rule proposed in [37], i.e., we update one or partial elements at a time while holding the other parameters fixed.

After some algebra, the marginal posterior distribution of x_i in (4) can be written as

$$q(x_i|R_i, \Sigma_i) = (1 - \pi_i)\delta(x_i) + \pi_i \mathcal{N}(x_i; m_i, V_i), \quad (10)$$

where

$$V_i = \frac{\tau_0 \Sigma_i}{\Sigma_i + \tau_0}, \quad (11)$$

$$m_i = \frac{\tau_0 R_i + \Sigma_i \mu_0}{\Sigma_i + \tau_0}, \quad (12)$$

$$\pi_i = \frac{\lambda_i}{\lambda_i + (1 - \lambda_i)\exp(-\mathcal{L})}, \quad (13)$$

$$\mathcal{L} = \frac{1}{2} \ln \frac{\Sigma_i}{\Sigma_i + \tau_0} + \frac{R_i^2}{2\Sigma_i} - \frac{(R_i - \mu_0)^2}{2(\Sigma_i + \tau_0)}. \quad (14)$$

Note that for notational brevity, we have omitted the iteration index t. The mean and variance defined in (5) and (6) can now be explicitly calculated as

$$g_a\big(R_i,\Sigma_i\big)=\pi_i m_i, \tag{15}$$

$$g_c\big(R_i,\Sigma_i\big)=\pi_i\big(m_i^2+V_i\big)-g_a^2\big(R_i,\Sigma_i\big). \tag{16}$$

To learn the sparse ratios $\lambda_i, i=1,\ldots,N$, we need to maximize $Q\big(\boldsymbol{\theta},\boldsymbol{\theta}^t\big)$ with respect to λ_i. After some algebra, we obtain the standard EM update equation as $\lambda_i^{t+1}=\pi_i^t$, which, albeit simple, fails to capture the inherent structure in the sparsity pattern. To address this problem, a novel learning rule is proposed as follows

$$\lambda_i^{t+1}=\frac{1}{\big|\mathcal{N}(i)\big|}\sum_{j\in\mathcal{N}(i)}\pi_j^t, \tag{17}$$

where $\mathcal{N}(i)$ denotes the set of nearest neighbor indexes of element x_i in $\mathbf{x}$ (1) and $\big|\mathcal{N}(i)\big|$ denotes the cardinality of $\mathcal{N}(i)$. For one dimensional (1D) data , $\mathcal{N}(i)=\big\{i-1,i+1\big\}$[1] and $\big|\mathcal{N}(i)\big|=2$, while for two dimensional (2D) data, $\mathcal{N}(i)=\big\{(q,l-1),(q,l+1),(q-1,l),(q+1,l)\big\}$ and $\big|\mathcal{N}(i)\big|=4$, where (q,l) indicates the coordinates of x_i in the 2D space. Generalizations to other cases can be made.

Note that in (17), we have chosen the nearest neighbor of each element, excluding itself, as the neighboring set. The estimate of one sparse ratio is not determined by its own estimate, but rather the average of its nearest neighbor estimates. The insight for this choice is that, for clustered sparse signals, if the nearest neighbors of one element are zero (nonzero), it will be zero (nonzero) with high probability, a similar idea to k-NN. If the neighboring set is chosen as the whole elements, the proposed algorithm reduces to EM-BG-GAMP [6], [34].

The leaning of other hyperparameters follows the standard rule of EM algorithm. Maximizing $Q\big(\boldsymbol{\theta},\boldsymbol{\theta}^t\big)$ with respect to Δ_0 and after some algebra, we obtain

$$\Delta_0^{t+1}=\frac{1}{M}\sum_a\Big[\frac{\big(y_a-Z_a^t\big)^2}{\big(1+V_a^t/\Delta_0^t\big)^2}+\frac{\Delta_0^t V_a^t}{\Delta_0^t+V_a^t}\Big], \tag{18}$$

where Z_a^t and V_a^t are obtained within the AMP iteration and are defined in Algorithm 1. Similarly, maximizing $Q\big(\boldsymbol{\theta},\boldsymbol{\theta}^t\big)$ with respect to μ_0 and τ_0 results in the update equations

$$\mu_0^{t+1}=\frac{\sum_i\pi_i^t m_i^t}{\sum_i\pi_i^t}, \tag{19}$$

$$\tau_0^{t+1}=\frac{1}{\sum_i\pi_i^t}\sum_i\pi_i^t\big[\big(\mu_0^t-m_i^t\big)^2+V_i\big]. \tag{20}$$

Valid initialization of the unknown hyperparameters is essential since EM algorithm may converge to a local maximum or a saddle point of the likelihood function [31]. The sparse ratios λ_i and noise variance Δ_0 are initialized as $\lambda_i^1=0.5$ and $\Delta_0^1=\big\|\mathbf{y}\big\|_2^2/M\big(\mathrm{SNR}^0+1\big)$, respectively, where SNR^0 is suggested to be 100 [34]. For the sparse Gaussian case, active mean μ_0 and variance τ_0 are initialized as $\mu_0^1=0$, and $\tau_0^1=\big(\big\|\mathbf{y}\big\|_2^2-M\Delta_0^1\big)/\lambda_i^1\big\|\mathbf{A}\big\|_F^2$, respectively, where$\big\|\cdot\big\|_2$, $\big\|\cdot\big\|_F$ are the l_2 norm and Frobenius norm, respectively.

The proposed approximate message passing with nearest neighbor sparsity pattern learning (AMP-NNSPL) is summarized in Algorithm 1. The complexity of AMP-NNSPL is dominated by matrix-vector multiplications in the original AMP and thus only scales as $\mathcal{O}(MN)$, i.e., the proposed algorithm is computationally efficient.

Algorithm 1 AMP-NNSPL Algorithm

Input: $\mathbf{y}$ $\mathbf{A}$.

Initialization: Set $t=1$ and T_{max},ϵ_{toc}. Initialize μ_0,τ_0,Δ_0 and $\lambda_i, i=1,\ldots,N$ as in Section III. $\hat{x}_i^1=\int x_i p_0(x_i)dx_i, \nu_i^1=\int|x_i-\hat{x}_i^1|^2 p_0(x_i)dx_i, i=1,\ldots,N$, $V_a^0=1, Z_a^0=y_a, a=1,\ldots,M$.

1) Factor node update: For $a=1,\ldots,M$

$$V_a^t=\sum_i|A_{ai}|^2\nu_i^t,$$

$$Z_a^t=\sum_i A_{ai}\hat{x}_i^t-\frac{V_a^t}{\Delta_0^t+V_a^{t-1}}\big(y_a-Z_a^{t-1}\big).$$

2) Variable node update: For $i=1,\ldots,N$

$$\Sigma_i^t=\Big[\sum_a\frac{|A_{ai}|^2}{\Delta_0^t+V_a^t}\Big]^{-1},$$

$$R_i^t=\hat{x}_i^t+\Sigma_i^t\sum_a\frac{A_{ai}\big(y_a-Z_a^t\big)}{\Delta_0^t+V_a^t},$$

$$\hat{x}_i^{t+1}=g_a\left(R_i^t,\Sigma_i^t\right),$$

$$\hat{\nu}_i^{t+1}=g_c\left(R_i^t,\Sigma_i^t\right).$$

3) Update $\lambda_i^{t+1}, i=1,\ldots N$, as (17);

4) Update $\mu_0^{t+1},\tau_0^{t+1},\Delta_0^{t+1}$ as (19), (20), and (18);

5) Set $t\leftarrow t+1$ and proceed to step 1) until T_{max} iterations or $\big\|\hat{\mathbf{x}}^{t+1}-\hat{\mathbf{x}}^t\big\|_2<\epsilon_{toc}\big\|\hat{\mathbf{x}}^t\big\|_2$.

IV. Numerical Experiments

In this section, a series of numerical experiments are performed to demonstrate the performance of the proposed algorithm under various settings. Comparisons are made to some state-of-the-art methods which need no prior information of the sparstiy pattern, e.g., PC-SBL [26] and its AMP version PCSBL-GAMP [27], MBCS-LBP [28], and EM-BG-GAMP [34]. The performance of Basis Pursuit (BP) [38]–[40] is also evaluated. Throughout the experiments, we set the maximum number of

[1]For end points of 1D data, the nearest neighbor set has only one element. For edge points of 2D data, the nearest neighbor set has only two or three elements.

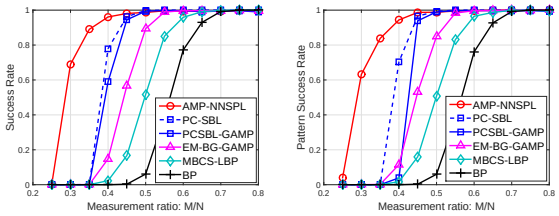


Figure 1. Success rate (left) and pattern success rate (right) vs. M/N for block-sparse signals $N = 100$, $K = 25$, $L = 4$, noiseless case.

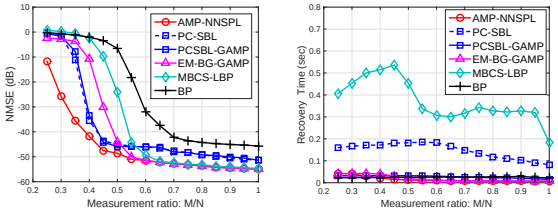


Figure 2. NMSE (left) and recovery time (right) vs. M/N for block-sparse signals $N = 100$, $K = 25$, $L = 4$, SNR = 50 dB.

iterations for AMP-NNSPL, PCSBL-GAMP, and EM-BG-GAMP to be $T_{max} = 200$, and the tolerance value of termination to be $\epsilon_{toc} = 10^{-6}$. For other algorithms, we use the defaut setting. The elements of measurement matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ are independently generated following standard Gaussian distribution and the columns are normalized to unit norm. The success rate is defined as the ratio of the number of successful trials to the total number of experiments, where a trial is successful if the normalized mean square error (NMSE) is less than -60 dB, where $\mathsf{NMSE} = 20\log_{10}(\|\hat{\mathbf{x}}-\mathbf{x}\|_2/\|\mathbf{x}\|_2)$. The pattern recovery success rate is defined as the ratio of the number of successful trials to the total number of experiments, where a trial is successful if the support is exactly recovered. A coefficient whose magnitude is less than 10^{-4} is deemed as a zero coefficient.

A. Synthetic Data

We generate synthetic block-sparse signals in a similar way as [21], [26], where K nonzero elements are partitioned into L blocks with random sizes and random locations. Set $N = 100, K = 25, L = 4$ and the nonzero elements are generated independently following Gaussian distribution with mean $\mu_0 = 3$ and variance $\tau_0 = 1$. The results are averaged over 1000 independent runs. Fig. 1 depicts the success rate and pattern recovery success rate. It can be seen that AMP-NNSPL achieves the highest success rate and pattern recovery rate at various measurement ratios. In the noisy setting, Fig. 2 shows the average NMSE and runtime of different algorithms when the signal to noise ratio (SNR) is 50 dB, where $SNR = 20\log_{10}(\|\mathbf{Ax}\|_2/\|\mathbf{w}\|_2)$. We see that AMP-NNSPL outperforms other methods both in terms of NMSE and computational efficiency.

B. Real Data

To evaluate the performance on real data, we consider a real angiogram image [18] of 100×100 pixels with sparsity around 0.12. Fig. 3 depicts the success rate in noiseless case and NMSE at $SNR = 50$ dB, respectively. The MBCS-LBP and PC-SBL algorithms are not included due to their high computational complexity. It

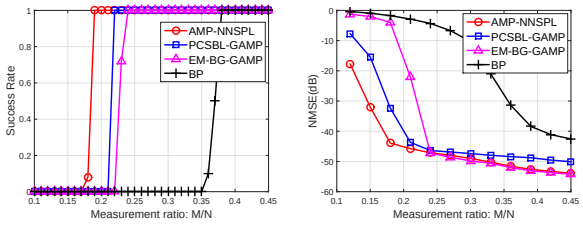
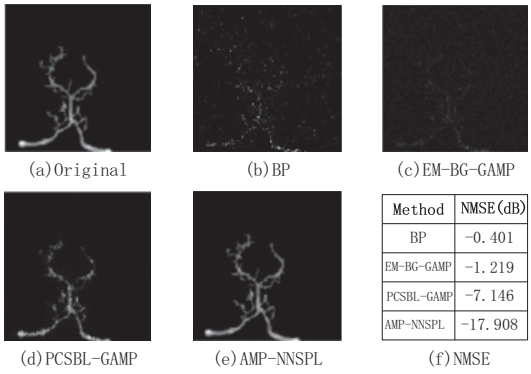


Figure 3. Success rate (left) in noiseless case and NMSE (right) at $SNR = 50$dB vs. M/N for real 2D angiogram image.

can be seen that AMP-NNSPL significantly outperforms other methods both in terms of success rate and NMSE. In particular, when $M/N = 0.12$ and $SNR = 50$ dB, typical recovery results are illustrated in Fig. 4, which shows that AMP-NNSPL achives the best reconstruction performance.



Method	NMSE(dB)
BP	-0.401
EM-BG-GAMP	-1.219
PCSBL-GAMP	-7.146
AMP-NNSPL	-17.908

(f) NMSE

Figure 4. Recovery results of real 2D angiogram image in noisy setting when $M/N = 0.12$ and $SNR = 50$ dB.

V. Conclusion

In this lettter, we propose an efficient algorithm termed AMP-NNSPL to recover clustered sparse signals when the sparsity pattern is unknown. Inspired by the k-NN algorithm, AMP-NNSPL learns the sparse ratios in each AMP iteration as the average of their nearest neighbor estimates using EM, thereby the sparsity pattern is learned adaptively. Experimental results on both synthetic and real data demonstrate the state-of-the-art performance of AMP-NNSPL.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25(2), pp. 21–30, Mar. 2008.

[3] Y. C. Eldar and G. Kutyniok, Eds., *Compressed sensing: theory and applications*. Cambridge Univ. Press, 2012.

[4] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," in *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, Nov. 2009, pp. 18 914–18 919.

[5] ——, "Message passing algorithms for compressed sensing: I. motivation and construction," in *IEEE Information Theory Workshop (ITW)*, Jan. 2010, pp. 1–5.

[6] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, 2012.

[7] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 2168–2172.

[8] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, Dec. 2011.

[9] U. S. Kamilov, S. Rangan, A. K. Fletcher, and M. Unser, "Approximate message passing with consistent parameter estimation and applications to sparse learning," *IEEE Trans. Inf. Theory*, vol. 60(5), pp. 2969–2985, May. 2014.

[10] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[11] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," DTIC Document, Tech. Rep., 2009.

[12] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, 2009.

[13] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[14] J. Huang, T. Zhang *et al.*, "The benefit of group sparsity," *The Annals of Statistics*, vol. 38, no. 4, pp. 1978–2004, 2010.

[15] 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.

[16] 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.

[17] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *The Journal of Machine Learning Research*, vol. 12, pp. 3371–3412, 2011.

[18] C. Hegde, P. Indyk, and L. Schmidt, "A nearly-linear time framework for graph-structured sparsity," in *Proceedings of The 32nd International Conference on Machine Learning*, 2015, pp. 928–937.

[19] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[20] 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.

[21] Z. Zhang and B. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. on Signal Process.*, vol. 61, no. 8, pp. 2009–2015, 2013.

[22] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, 2009.

[23] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a markov-tree prior," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3439–3448, 2012.

[24] L. Yu, H. Sun, J.-P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259–269, 2012.

[25] M. R. Andersen, O. Winther, and L. K. Hansen, "Spatio-temporal spike and slab priors for multiple measurement vector problems," *arXiv preprint arXiv:1508.04556*, 2015.

[26] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. on Signal Process.*, vol. 63, no. 2, pp. 360–372, 2015.

[27] J. Fang, L. Zhang, and H. Li, "Two-dimensional pattern-coupled sparse Bayesian learning via generalized approximate message passing," *arXiv preprint arXiv:1505.06270*, 2015.

[28] L. Yu, H. Sun, G. Zheng, and J. P. Barbot, "Model based Bayesian compressive sensing via local beta process," *Signal Processing*, vol. 108, pp. 259–271, 2015.

[29] E. Fix and J. L. Hodges Jr, "Discriminatory analysis-nonparametric discrimination: consistency properties," DTIC Document, Tech. Rep., 1951.

[30] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[31] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[32] T. Park and G. Casella, "The Bayesian lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.

[33] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.

[34] J. P. Vila and P. Schniter, "Expectation-maximization gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, 2013.

[35] A. Montanari, "Graphical models concepts in compressed sensing," *arXiv preprint arXiv:1011.4328*, 2010.

[36] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[37] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in graphical models*. Springer, 1998, pp. 355–368.

[38] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[39] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

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