A Framework for Compressive Sensing of Asymmetric Signals Using Normal and Skew-Normal Mixture Prior

Sheng Wang, *Student Member, IEEE*, and Nazanin Rahnavard, *Member, IEEE*

***Abstract*—In this work, we are interested in the compressive sensing of sparse signals whose significant coefficients are distributed asymmetrically with respect to zero. To properly address this problem, we develop a framework utilizing a two-state normal and skew normal mixture density as the prior distribution of the signal. The significant and insignificant coefficients of the signal are represented by skew normal and normal distributions, respectively. A novel approximate message passing-based algorithm is developed to estimate the signal from its compressed measurements. A fast gradient-based estimator is designed to infer the density of each state. Experiment results on simulated data and two real-world tests, i.e., multi-input multi-output (MIMO) communication system and weather sensor network, confirm that our proposed technique is powerful in exploiting asymmetrical feature, and outperforms many sophisticated methods.**

***Index Terms*—Compressive sensing, asymmetrical signal, mixture model, approximate message passing.**

I. Introduction

COMPRESSIVE sensing (CS) [1], [2] is a powerful technique for solving certain ill-posed linear inverse problems. In compressive sensing, the signal is sampled by a linear projection as,

$$\underline{y} = A\underline{x} + \underline{e}, \tag{1}$$

where $\underline{x} \in \mathbb{R}^{N\times 1}$ is the signal of interest, $A \in \mathbb{R}^{M\times N}$ is the known sampling matrix with $M \ll N$, $\underline{y} \in \mathbb{R}^{M\times 1}$ is the observed measurements, and $\underline{e} \in \mathbb{R}^{M\times 1} \sim \mathcal{N}(\underline{0}, \sigma_e^2 \mathrm{I}_{M\times M})$ is the measurement white Gaussian noise.

It is noted that $M \ll N$ in (1), therefore, the inverse problem has infinitely many solutions. Compressive sensing enables reliable reconstruction of the signal under the conditions that the signal is sufficiently sparse/compressible, and the sampling matrix satisfies certain properties [1]. Here by sparse/compressible, we intend that $K \ll N$ entries of the signal have significant magnitudes, with the remaining entries being insignificant, and the ratio K/N is referred to as the *sparsity rate* in the literature.

A great number of applications have been inspired by the success of compressive sensing. For instance, CS found great utilization in MIMO broadcast network [3], orthogonal frequency division multiplexing (OFDM) system [4] [5], heterogeneous cellular networks (HetNets) [6], distributed data storage [7], multiple description coding (MDC) [8], multimedia communication [9], [10], network fault identification [11], [12], and time series analysis [13].

In this work, we are particularly interested in the reconstruction task. To reconstruct the signal from its under-sampled projection, sparsity promoting algorithms are employed. Among the many, basis pursuit [1] and LASSO [14] are probably the most classic techniques. By casting the problem as a convex optimization problem, they yield close to optimal performance at $\mathcal{O}(N^3)$ complexity. The sparse reconstruction task can be treated from a Bayesian aspect as well, where the distribution of the signal is modelled by a mixture of a few density components. In [15] and [16], the signal is modelled by a mixture of Laplace densities, and the coefficients are inferred by approximate message passing (AMP). In [17], two types of mixture models, *i.e.,* a Bernoulli-Gaussian mixture and a two-state Gaussian mixture, are utilized as the prior distributions of the wavelet transform coefficients of images.

As can be seen, the density components in these studies [15]–[18] are symmetrically distributed around their means. In practice, the underling density of the signal coefficients could be asymmetric.

One example can be found in multibeam opportunistic random-beamforming network [3], where the Base Station, equipping with Q antennas, schedules its downlink transmission based on the Signal to Interference and Noise Ratio (SINR) of the mobile users.

Specifically, the SINR of n-th user on q-th Base Station antenna is calculated as [3],

$$\mathrm{SINR}_{n,q} = \frac{|\underline{g}_n^{\mathsf{T}} \underline{\psi}_q|^2}{1/\rho_n + \sum_{l\neq q} |\underline{g}_n^{\mathsf{T}} \underline{\psi}_l|^2} \geq 0, \tag{2}$$

where ρ_n is a positive scaler standing for the Signal to Noise Ratio (SNR), $\underline{g}_n \in \mathbb{C}^{Q\times 1}$ is the complex Gaussian Channel vector, and $\underline{\psi}_q \in \mathbb{R}^{Q\times 1}$ is the random beam generated at Base Station. As can be seen in (2), SINR is strictly positive, and is therefore asymmetric about zero.

Manuscript received April 9, 2015; revised August 12, 2015 and September 24, 2015; accepted September 28, 2015. Date of publication October 8, 2015; date of current version December 15, 2015. This work was supported by the National Science Foundation under Grant ECCS-1418710. The associate editor coordinating the review of this paper and approving it for publication was M. Xiao.

S. Wang is with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078 USA (e-mail: sheng.wang@okstate.edu).

N. Rahnavard is with the Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816 USA (e-mail: nazanin@eecs.ucf.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2015.2488651

Another example can be found in sensor networks, where certain type of weather data, let us say outside air temperature, when subtracted from the historical average, is asymmetrically positive or negative when the disrupting weather phenomena is heat or cool, respectively. Additionally, it is found that in microarray time course data analysis from biomedical research [19], the gene expressions involved in embryo are more often developed with an increasing trend.

Therefore, distributions including normal and Laplace in this case may not be a proper model to capture all the salient features, and dealing with asymmetric signals calls for more sophisticated approaches. Two related work can be found in [20] and [21]. In [20], a normal density mixture is employed as the prior distribution, and a powerful algorithm is put forward to estimate the signal while learning the mixture via Expectation Maximization [22]. In [21], an effective technique is developed to handle non-negative sparse signals by modelling the signal with a non-negative normal mixture.

While being highly effective in general, both [20] and [21] have limitations. For example, the mixture using normal components in [20] is known to be sensitive to outliers, and the performance degrades with smaller sample size [23]. Meanwhile, the work [21] is designed exclusively for non-negative signals, and is not capable in handling signals with both positive and negative significant elements.

Given these limitations, we are aiming to solve for asymmetrical signals, and to develop a new and more generalized framework. To this end, a two-state normal and skew normal mixture density is proposed. The significant coefficients of the signal are represented by a skew normal density, which is more general than the normal one, and comes with more flexibility in dealing with the asymmetric features. A message passing algorithm is developed to estimate the signal from the measurements. A fast gradient-based estimator is designed to infer the density of each state.

The performance of our proposed technique is examined under a variety of tests, including phase transition, noisy reconstruction, support set recovery rate, and runtime tests. Furthermore, our technique finds promising applications in MIMO communication system, as well as weather sensor network. We show that in MIMO communication system, given the same amount of feedback overhead, our technique is able to realize a significantly higher achievable sum rate. Besides, in weather sensor network application, the disrupting weather phenomena can be successfully learned by our proposed technique.

Overall, experimental results of both simulated and real-world tests show that, our technique can effectively exploit the asymmetric feature of the signal, while being competitively efficient in solving large scale problems.

The remainder of this paper is organized as follows. A review of loopy belief propagation and approximate message passing is provided in Section II. The signal model and our approximate message passing algorithm utilizing the two-state normal and skew normal mixture density are detailed in Section III. Gradient-based parameter estimation is detailed in Section IV. The complexity of our technique is analyzed in Section V. Experimental results are summarized in Section VI, and Section VII concludes the paper.

II. Overview of Approximate Message Passing

A. AMP: Fidelity and Complexity

In the reconstruction phase of compressive sensing, the task is to find the solution that complies with the measurements $\underline{y}$, while providing the best consistency with the prior sparsity model.

A variety of methods with varying balance between reconstruction fidelity and complexity are developed[1]. For instance, LASSO [1], [14] is known to achieve the least reconstruction distortion, with computational complexity $\mathcal{O}(N^3)$. On the other hand, while the fast scheme such as iterative thresholding [25], [26] has much less complexity, it requires more samples, *i.e.,* greater M, to achieve a similar fidelity of LASSO [1], [14].

Approximate message passing [15], [16], [27] is a powerful technique that realizes the best of reconstruction fidelity and computational complexity. In [27], it is shown that the performance of AMP can be precisely modelled by the *State Evolution*, where under the conditions that the signal is adequately sparse and the number of measurement is sufficient, the formal *Mean Square Error* [28] of AMP reduces to 0. Moreover, with the complexity dominated by multiplying the sampling matrix (of size M-by-N), with a vector (of size N-by-1), AMP is one of the fastest techniques in compressive sensing community [27].

B. Estimate Marginal Posterior by Message Passing

AMP is built on the success of message passing. Message passing, also known as belief propagation decoding [18], [29], allows efficient approximation of the marginal posterior $P(x_n|\underline{y})$, for $n = 1, \ldots, N$, by exchanging messages between variable nodes $\underline{x}$ and check nodes $\underline{y}$, where the messages carry the probability distribution of the corresponding variable nodes.

Specifically, let $\nu^{\,i}_{x_n \to y_m}(x_n)$ be the message sent from variable node x_n to check node y_m at i-th iteration, and denote $\nu^{\,i}_{y_m \to x_n}(x_n)$ as the reverse, with both messages encoding the belief, namely probability density function (*pdf*), of x_n.

The message from variable node x_n to check node y_m, $\nu^{\,i}_{x_n \to y_m}(x_n)$, is calculated as the product of the prior distribution of x_n, *i.e.,* $f(x_n)$, and all incoming messages to x_n from check nodes $\underline{y}$, with the exception of the one from y_m [18],

$$\nu^{i}_{x_n \to y_m}(x_n) \cong f(x_n) \prod_{u \in \{1,\ldots,M\} \setminus m} \nu^{i-1}_{y_u \to x_n}(x_n), \tag{3}$$

where $\cong$ denotes identity up to a normalization constant.

The message from check node y_m to variable node x_n, $\nu^{i}_{y_m \to x_n}(x_n)$, is evaluated based on the product of the constraint on y_m, and all incoming messages of y_m from variable nodes $\underline{x}$, with the exception of the one from x_n. Under white Gaussian noise environment, the constraint on y_m is

$$con(y_m, \underline{x}) = \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{(y_m - A^{\intercal}_{m\cdot}\underline{x})^2}{2\sigma_e^2}\right), \tag{4}$$

where $A_{m\cdot}$ represents the m-th row of A, and superscript $^{\intercal}$ denotes vector and matrix transpose. In what comes next, A_{mn}

[1]See [24] and reference therein for a survey of popular methods for the compressive sensing reconstruction task.

is the entry at the m-th row and n-th column of A. Similarly, $A_{\cdot n}$ denote the n-th column of A.

Since $con(y_m, \underline{x})$ involves all variable nodes $\underline{x}$, the product is then marginalized by sum over all $\underline{x}$ but x_n [18], *i.e.,*

$$\nu^i_{y_m\to x_n}(x_n) \cong \underbrace{\int\cdots\int}_{N-1} con(y_m,\underline{x}) \prod_{\substack{t=1\\ t\neq n}}^{N} \nu^i_{x_t\to y_m}(x_t)\, \underbrace{\mathrm{d}x_1\cdots\mathrm{d}x_t\cdots\mathrm{d}x_N}_{t\in\{1,\ldots,N\}\backslash n}. \quad (5)$$

C. Minimal-Mean-Squared-Error (MMSE) Inference by Approximation

In classic message passing [18], message is represented in the form of $1/\sigma_S$ uniform samples of the corresponding *pdf*, where σ_S is the standard deviation of the insignificant coefficients. Therefore, a storage proportional to MN/σ_S is needed at each iteration.

While being reasonably effective in some cases, this procedure calls for considerably large memory space, and is not satisfactorily efficient under large signal dimensionality and very small σ_S.

AMP, on the other hand, is more efficient. Specifically, with adequately large M and N, message in AMP is approximated[2] by Gaussian density, which is further parameterized by the corresponding mean and variance [15]–[17], *i.e.,*

$$\nu^i_{x_n\to y_m}(x_n) \approx \mathcal{N}\left(x_n; \mu^i_{x_{nm}}, \sigma^{2\,i}_{x_{nm}}\right), \quad (6)$$

where

$$\mu^i_{x_{nm}} = \int_{-\infty}^{\infty} x_n \nu^i_{x_n\to y_m}(x_n)\,\mathrm{d}x_n, \quad (7)$$

$$\sigma^{2\,i}_{x_{nm}} = \int_{-\infty}^{\infty} (x_n-\mu^i_{x_{nm}})^2 \nu^i_{x_n\to y_m}(x_n)\,\mathrm{d}x_n, \quad (8)$$

and

$$\nu^i_{y_m\to x_n}(x_n) \approx \mathcal{N}\left(x_n; \mu^i_{y_{mn}}, \sigma^{2\,i}_{y_{mn}}\right), \quad (9)$$

in which

$$\mu^i_{y_{mn}} = \frac{1}{A_{mn}} \times \left(y_m - \sum_{t\in\{1,\ldots,N\}\backslash\{n\}} A_{mt}\mu^i_{x_{tm}}\right), \quad (10)$$

$$\sigma^{2\,i}_{y_{mn}} = \frac{1}{A^2_{mn}} \times \left(\sigma^2_e + \sum_{t\in\{1,\ldots,N\}\backslash\{n\}} A^2_{mt}\sigma^{2\,i}_{x_{tm}}\right). \quad (11)$$

Following the notation in [15] and [16], define the mean operator $\mathbb{F}(\kappa,\varsigma)$, and variance operator $\mathbb{G}(\kappa,\varsigma)$ as,

$$\mathbb{F}(\kappa,\varsigma) = \mathrm{E}_{f_{v\to c}}(X), \quad (12)$$

$$\mathbb{G}(\kappa,\varsigma) = \mathrm{Var}_{f_{v\to c}}(X), \quad (13)$$

where the *pdf* of X is $f_{v\to c}(x) \cong \mathcal{N}(x;\kappa,\varsigma) f(x)$, with $f(x)$ denoting the prior distribution of X.

[2]In this work, $\approx$ is utilized to denote approximation.

Therefore, with the above approximation, and combining the product term in (3) which are all Gaussian, $\nu^{i+1}_{x_n\to y_m}$ can be written as,

$$\begin{aligned}\nu^{i+1}_{x_n\to y_m}(x_n) &\cong \mathcal{N}(x_n;\kappa^i_{nm},\varsigma^i_n) f(x_n)\\ &\cong \mathcal{N}\left(x_n;\mu^{i+1}_{x_{nm}},\sigma^{2\,i+1}_{x_{nm}}\right),\end{aligned} \quad (14)$$

where

$$\kappa^i_{nm} = \sum_{\substack{u=1\\ u\neq m}}^{M} A_{un}\mu^i_{y_{un}}, \quad \varsigma^i_n = \frac{1}{M}\sum_{u=1}^{M} A^2_{un}\sigma^{2\,i}_{y_{un}}, \quad (15)$$

$$\mu^{i+1}_{x_{nm}} = \mathbb{F}(\kappa^i_{nm},\varsigma^i_n), \quad \sigma^{2\,i+1}_{x_{nm}} = \mathbb{G}(\kappa^i_{nm},\varsigma^i_n). \quad (16)$$

As can be seen, following the above update rule, variable node x_n sends a *unique* pair of $\left(\mu^i_{x_{nm}}, \sigma^{2\,i}_{x_{nm}}\right)$ to y_m, for $m = 1,\ldots,M$. In turn, check node y_m sends a *unique* pair of $\left(\mu^i_{y_{mn}}, \sigma^{2\,i}_{y_{mn}}\right)$ to x_n, for $n = 1,\ldots,N$. As a result, the memory requirement scales with $2MN$.

The work [15] and [16] further show that, with mild accuracy compromise, the storage requirement can be further reduced by first order approximation.

Specifically, by first order approximation, it is intended that variable node x_n sends a *uniform* pair to all check nodes, *i.e.*, $\nu^i_{x_n\to y_m} = \mathcal{N}\left(\mu^i_{x_n}, \sigma^{2\,i}_{x_n}\right)$, for $m = 1,\ldots,M$. Similarly, check node y_m sends a *uniform* pair to all variable nodes, *i.e.,* $\nu^i_{y_m\to x_n} = \mathcal{N}\left(\mu^i_{y_m}, \varsigma^i\right)$, for $n = 1,\ldots,N$, in which [15]–[17],

$$\mu^i_{x_n} = \mathbb{F}(\kappa^{i-1}_{x_n}, \varsigma^{i-1}), \quad (17)$$

$$\sigma^{2\,i}_{x_n} = \mathbb{G}(\kappa^{i-1}_{x_n}, \varsigma^{i-1}), \quad (18)$$

$$\mu^i_{y_m} = y_m - \sum_{n=1}^{N} A_{mn}\mu^{i-1}_{x_n} + \frac{\mu^{i-1}_{y_m}}{M}\sum_{n=1}^{N}\mathbb{F}'(\kappa^{i-1}_{x_n}, \varsigma^{i-1}), \quad (19)$$

$$\varsigma^i = \sigma^2_e + \frac{1}{M}\sum\nolimits_{n=1}^{N} \sigma^{2\,i}_{x_n}, \quad (20)$$

$$\kappa^{i-1}_{x_n} = \sum\nolimits_{m=1}^{M} A_{mn}\mu^{i-1}_{y_m} + \mu^{i-1}_{x_n}, \quad (21)$$

with $\mathbb{F}'(\kappa^{i-1}_{x_n}, \varsigma^{i-1})$ being the first order derivative of $\mathbb{F}(\kappa^{i-1}_{x_n}, \varsigma^{i-1})$ with respect to $\kappa^{i-1}_{x_n}$. After convergence of message passing, the MMSE estimate of the signal is formed as $\hat{\underline{x}} = [\mu^{I_d}_{x_1},\ldots,\mu^{I_d}_{x_N}]^\intercal$, where I_d represents the last message passing iteration.

III. Approximate Message Passing Based on Normal and Skew Normal Mixture Density

A. Skew Normal Density

In this work, we are aiming to estimate sparse signals whose significant coefficients are distributed asymmetrically with respect to zero.

Signals with this asymmetrical feature can be either *right-skewed*, or *left-skewed*. Specifically, for *right-skewed*, the majority of the significant coefficients are of positive sign, with the remaining few being negative. Similarly, for *left-skewed*, the majority of the significant coefficients are of negative sign, with the remaining few being negative.

As discussed, due to the symmetry, neither normal nor Laplace densities could encapsulate the asymmetric nature of signals with such prior information.

In this work, we employ a normal and skew normal density mixture as the prior distribution of such signals. More specifically, the distribution of the significant coefficients is modelled by a skew normal density, the probability density function of which was formally defined in [30] as,

$$\mathcal{SN}(x;\xi,\omega,\alpha) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right), \qquad (22)$$

where $\xi, \omega,$ and α represent the location, scale, and shape parameters, $\phi(\cdot)$ and $\Phi(\cdot)$ denote the *pdf* and the cumulative density function (*cdf*) of the standard normal distributed random variable, respectively.

Compared to the normal *pdf*, a noteworthy aspect of (22) is the additional term $\Phi\left(\alpha\frac{x-\xi}{\omega}\right)$, which controls the skewness of the density. It is readily seen that (22) reduces to a normal density when α is set to 0, and approaches to positive/negative half normal density in the limits $\alpha \to \pm\infty$.

Fig. 1a and 1b show two curves of the skew normal densities with (ξ, ω, α) being set to $(0, 100, -10)$ and $(0, 100, 10)$, respectively. It can be seen that both of these two densities are asymmetric with respect to $x = 0$, where the density with negative shape parameter α in Fig. 1a is left-skewed, and the density with positive α in Fig. 1b is right-skewed. Besides, as compared to the non-negative normal density [21], the skew normal density is more flexible in accommodating both positive and negative elements.

Similar to [18], the distribution of the insignificant coefficients is modelled by normal density. Meanwhile, we consider the case where the location parameter $\xi = 0$. Overall, the *pdf* of the signal can be written as,

$$f(x) \cong (1-\lambda)\times\mathcal{N}(x;0,\sigma_S^2) + \lambda\times\mathcal{SN}(x;0,\omega_L,\alpha_L), \quad (23)$$

where $\lambda = K/N$ denotes the *sparsity rate*. For convenience, let $\underline{\Theta} = [\sigma_S^2, \omega_L, \alpha_L]$ be the characterizing parameters set of the mixture.

B. System Diagram

Our proposed method consists of two functionality modules, with each module being iterative. Fig. 2 is the system diagram of our technique.

The first module, as shown in the left of Fig. 2, involves the estimation of the signal using the approximation message passing with skew normal and normal mixture density. After the message passing completes, the estimation $\underline{\hat{x}}$ is then fed to the second module, where the parameters of the mixture, $\underline{\hat{\Theta}} = [\hat{\sigma}_S^2, \omega_L^*, \alpha_L^*]$, are inferred. These two modules execute alternatively and repeatedly until convergence is achieved.

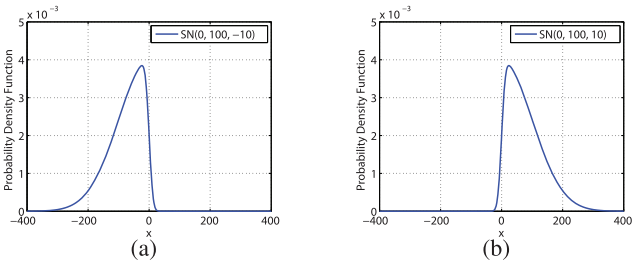


Fig. 1. Skew Normal Density. (a) Left-skewed with $\alpha = -10$. (b) Right-skewed with $\alpha = 10$.

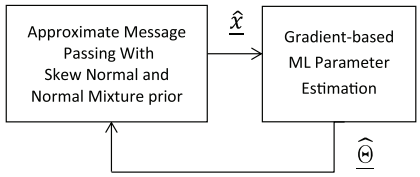


Fig. 2. System Diagram.

C. Message Passing

In this section, we will detail our message passing algorithm utilizing the proposed skew normal and normal mixture density.

Specifically, given (6) to (11), and recalling the skew normal and normal mixture density (23), the message from x_n to y_m at $(i+1)$-th iteration can be written as,

$$\begin{aligned}\nu_{x_n\to y_m}^{i+1}(x_n) &\cong \mathcal{N}(x_n;\kappa_{nm}^i,\varsigma_n^i)f(x_n)\\ &= (1-\lambda)\mathcal{N}\left(x_n;\kappa_{nm}^i,\varsigma_n^i\right)\mathcal{N}(x_n;0,\sigma_S^2)\\ &\quad+\lambda\mathcal{N}\left(x_n;\kappa_{nm}^i,\varsigma_n^i\right)\mathcal{SN}(x_n;0,\omega_L,\alpha_L), \qquad (24)\end{aligned}$$

where $\kappa_{nm}^i = \sum_{\substack{u=1\\u\neq m}}^{M} A_{un}\mu_{y_{un}}^i$, and $\varsigma_n^i = \frac{1}{M}\sum_{u=1}^{M} A_{un}^2\,\sigma_{y_{un}}^{2\,i}$.

With the above, the next step is to approximate (24) by normal density (6). This calls for the evaluation of the mean and variance of $\nu_{x_n\to y_m}^{i+1}$. For our specific problem, in which the prior density is a normal and skew normal mixture, one needs to analyze the product $\mathcal{N}(x|\kappa,\varsigma)\mathcal{SN}(x|0,\omega_0,\alpha_0)$ in (24). Therefore, ***Lemma 1*** to ***Lemma 3*** are derived below.

Lemma 1: Let $U \sim \mathcal{N}(\mu,\sigma^2)$ be a Gaussian random variable. We have $\mathrm{E}(\Phi(hU+k)) = \Phi(\frac{k+h\mu}{\sqrt{1+h^2\sigma^2}})$ for any $h, k \in \mathbb{R}$.

Proof: ***Lemma 1*** is a direct extension of ***Lemma 2*** in [30], which states that $\mathrm{E}(\Phi(hV+k)) = \Phi(\frac{k}{\sqrt{1+h^2}})$ for $V \sim \mathcal{N}(0,1)$. By change of variable, $V = \frac{U-\mu}{\sigma}$, ***Lemma 1*** follows. ■

Corollary 1: Let $G(x) = \mathcal{N}(x;\kappa,\varsigma)\mathcal{SN}(x;0,\omega_0,\alpha_0)$ be the product of the *pdf* of normal and skew normal densities, then $C_0\int_{-\infty}^{\infty} G(x)\mathrm{d}x = 1$ for a $C_0 \in \mathbb{R}^+$.

Proof: To prove ***Corollary 1***, it is sufficient to show that $\int_{-\infty}^{\infty} G(x)\mathrm{d}x$ has a finite value. Recalling $G(x) \geq 0$ and $\Phi(x) \leq 1$ for $x \in \mathbb{R}$, it is derived that,

$$\int_{-\infty}^{\infty} G(x)\mathrm{d}x < 2\int_{-\infty}^{\infty} \mathcal{N}(x;\kappa,\varsigma)\mathcal{N}(x;0,\omega_0^2)\mathrm{d}x < \sqrt{\frac{2}{\pi\omega_0^2}}\int_{-\infty}^{\infty}\mathcal{N}(x;\kappa,\varsigma)\mathrm{d}x = \sqrt{\frac{2}{\pi\omega_0^2}}. \quad (25)$$

■

Additionally, $\int_{-\infty}^{\infty} G(x)\mathrm{d}x$ is found to be,

$$\int G(x)\mathrm{d}x = \int \mathcal{N}(x;\kappa,\varsigma)\mathcal{SN}(x;0,\omega_0,\alpha_0)\mathrm{d}x \quad (26)$$

$$= \int \frac{1}{\pi\omega_0\sqrt{\varsigma}}\exp\left(-\frac{(x-\kappa)^2}{2\varsigma}-\frac{x^2}{2\omega_0^2}\right)\Phi\left(\frac{\alpha_0 x}{\omega_0}\right)\mathrm{d}x \quad (27)$$

$$= \frac{1}{\pi\omega_0\sqrt{\varsigma}}\exp\left(\frac{-\kappa^2}{2(\varsigma+\omega_0^2)}\right)\int\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\Phi\left(\frac{\alpha_0 x}{\omega_0}\right)\mathrm{d}x \quad (28)$$

$$= \sqrt{\frac{2}{\pi(\varsigma+\omega_0^2)}}\exp\left(\frac{-\kappa^2}{2(\varsigma+\omega_0^2)}\right)\int\mathcal{N}(\mu,\sigma^2)\Phi\left(\frac{\alpha_0 x}{\omega_0}\right)\mathrm{d}x, \quad (29)$$

where $\mu = \frac{\kappa\omega_0^2}{\varsigma+\omega_0^2}$, $\sigma^2 = \frac{\varsigma\omega_0^2}{\varsigma+\omega_0^2}$, and all integrals are from $-\infty$ to ∞.

Applying ***Lemma 1*** on the integral term in (29), it is derived that, $C_0 = \upsilon\exp(\gamma)\,\Phi(\eta)^{-1}$, in which $\upsilon = \left(0.5\pi(\varsigma+\omega_0^2)\right)^{1/2}$, $\gamma = \frac{\kappa^2}{2(\varsigma+\omega_0^2)}$, $\eta = \frac{h\mu}{\sqrt{1+h^2\sigma^2}}$, $h = \frac{\alpha_0}{\omega_0}$.

Lemma 2. Let the *pdf* of the random variable X be $C_0 \times \mathcal{N}(x;\kappa,\varsigma)\mathcal{SN}(x;0,\omega_0,\alpha_0)$. The moment generating function of X is found to be,

$$M_X(t) = \exp\left(\mu t+\frac{\sigma^2 t^2}{2}\right)\Phi^{-1}(\eta)\Phi\left(\eta+\frac{h\sigma^2 t}{\sqrt{1+h^2\sigma^2}}\right). \quad (30)$$

Proof:

$$M_X(t) = C_0\int\exp(tx)\mathcal{N}(x;\kappa,\varsigma)\mathcal{SN}(x;0,\omega_0,\alpha_0)\mathrm{d}x \quad (31)$$

$$= \Phi(\eta)^{-1}\int\exp(tx)\mathcal{N}(\mu,\sigma^2)\Phi\left(\frac{\alpha_0 x}{\omega_0}\right)\mathrm{d}x \quad (32)$$

$$= \frac{\exp\left(\mu t+\frac{t^2\sigma^2}{2}\right)}{\Phi(\eta)}\int\mathcal{N}(\mu+t\sigma^2,\sigma^2)\Phi\left(\frac{\alpha_0 x}{\omega_0}\right)\mathrm{d}x \quad (33)$$

$$= \exp\left(\mu t+\frac{\sigma^2 t^2}{2}\right)\Phi^{-1}(\eta)\Phi\left(\eta+\frac{h\sigma^2 t}{\sqrt{1+h^2\sigma^2}}\right), \quad (34)$$

where (32) holds due to ***Corollary 1***, and (34) holds due to ***Lemma 1***, and all integrals are from $-\infty$ to ∞. ■

With the moment generating function $M_X(t)$, the mean and variance of the density function $C_0 \times G(x)$ are derived.

Lemma 3. Let the *pdf* of the random variable X be $C_0 \times \mathcal{N}(x;\kappa,\varsigma)\mathcal{SN}(x;0,\omega_0,\alpha_0)$. Then the mean and variance are given by

$$\mathrm{E}(X) = \mu+\frac{\theta\sigma^2}{\sqrt{2\pi}}\,\Phi^{-1}(\eta)\exp\left(-\frac{1}{2}\eta^2\right), \quad (35)$$

TABLE I
MESSAGE PASSING PARAMETERS FOR $\mathbb{F}(\kappa,\varsigma)$ AND $\mathbb{G}(\kappa,\varsigma)$

$\mu_1 = \kappa\rho_S$,	$\mu_2 = \mu_0 + \frac{\theta\sigma_0^2}{\sqrt{2\pi}}\Phi^{-1}(\eta)\exp\left(-\frac{1}{2}\eta^2\right)$,	
$\sigma_1^2 = \varsigma\rho_S$,	$\sigma_2^2 = \mu_0^2+\sigma_0^2-\mu_2^2+(\mu_2-\mu_0)\rho_0$,	
$p_1 = (1-\lambda)\frac{C}{C_1}$,	$p_2 = \lambda\frac{C}{C_2}$,	$C_1 = \upsilon_S\beta$,
$C_2 = \upsilon_L\exp(\gamma)\Phi(\eta)^{-1}$,	$\mu_0 = \kappa\rho_L$,	$\sigma_0^2 = \varsigma\rho_L$,
$\rho_S = \frac{\sigma_S^2}{\varsigma+\sigma_S^2}$,	$\rho_L = \frac{\omega_L^2}{\varsigma+\omega_L^2}$,	$\rho_0 = \frac{2\mu_0+\mu_0 h^2\sigma_0^2}{1+h^2\sigma_0^2}$,
$\gamma = \frac{1}{2\sigma_0^2}\left(\kappa^2\rho_L-\mu_0^2\right)$,	$\eta = \frac{h\mu_0}{\sqrt{1+h^2\sigma_0^2}}$,	$\theta = \frac{h}{\sqrt{(1+h^2\sigma_0^2)}}$,
$h = \frac{\alpha_L}{\omega_L}$,	$\beta = \exp\left(\frac{\kappa^2}{2(\varsigma+\sigma_S^2)}\right)$,	$\upsilon_S = \sqrt{2\pi(\varsigma+\sigma_S^2)}$,
$\upsilon_L = \sqrt{\frac{\pi(\varsigma+\omega_L^2)}{2}}$,	$C = \frac{C_1C_2}{\lambda C_1+(1-\lambda)C_2}$.	

TABLE II
MESSAGE PASSING PARAMETERS FOR $\mathbb{F}'(\kappa,\varsigma)$

$\delta = \rho_L - \frac{\theta\sigma_0^2}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\eta^2\right)\left(\tau_0\Phi^{-2}(\eta)+\eta\rho_L\theta\Phi^{-1}(\eta)\right)$.	
$\tau = (1-\lambda)\tau_2\upsilon_L+\lambda(\tau_1\Phi(\eta)+\beta\tau_0)\upsilon_S$,	
$\zeta_1 = \frac{\upsilon_L\tau_2\beta_0-\upsilon_L\exp(\gamma)\tau}{\beta_0^2}$,	$\zeta_2 = \frac{\upsilon_S\beta_0(\tau_1\Phi(\eta)+\beta\tau_0)-\upsilon_S\beta\Phi(\eta)\tau}{\beta_0^2}$,
$\tau_1 = \frac{\beta\kappa}{\varsigma+\sigma_S^2}$,	$\tau_2 = \frac{(\kappa-\mu_0)\exp(\gamma)\rho_L}{\sigma_0^2}$,
$\tau_0 = \frac{\exp(-0.5\eta^2)\rho_L\theta}{\sqrt{2\pi}}$,	$\beta_0 = (1-\lambda)\upsilon_L\exp(\gamma)+\lambda\beta\upsilon_S\Phi(\eta)$.

and

$$\mathrm{Var}(X) = \mu^2+\sigma^2+(\mathrm{E}(X)-\mu)\,\rho-(\mathrm{E}(X))^2, \quad (36)$$

respectively, where $\theta = \frac{h}{\sqrt{1+h^2\sigma^2}}$, and $\rho = \frac{2\mu+\mu h^2\sigma^2}{1+h^2\sigma^2}$.

Using ***Lemma 1*** to ***Lemma 3*** and omitting the iteration superscript i and subscripts n and m for coefficients, (24) can be approximated by normal density as,

$$\nu_{x\to y}(x) \cong \mathcal{N}(\mu_x,\sigma_x^2), \quad (37)$$

$$\mu_x = \mathbb{F}(\kappa,\varsigma) = p_1\mu_1+p_2\mu_2, \quad (38)$$

$$\sigma_x^2 = \mathbb{G}(\kappa,\varsigma) = p_1\left(\mu_1^2+\sigma_1^2\right)+p_2\left(\mu_2^2+\sigma_2^2\right)-(p_1\mu_1+p_2\mu_2)^2, \quad (39)$$

where $\mu_1, \sigma_1^2, p_1, \mu_2, \sigma_2^2, p_2$ are calculated in Table I.

Omitting the iteration superscripts and coefficient subscripts, $\mathbb{F}'(\kappa_{x_n}^{i-1},\varsigma^{i-1})$ in (19) is calculated as,

$$\mathbb{F}'(\kappa,\varsigma) = (1-\lambda)\left(\mu_1\zeta_1+\frac{C}{C_1}\rho_S\right)+\lambda\left(\mu_2\zeta_2+\frac{C}{C_2}\delta\right), \quad (40)$$

in which ζ_1, ζ_2 and δ can be calculated as Table II.

Therefore, similar to the approximate message passing (17)-(21) for arbitrary prior density, our approximate message passing utilizing the proposed normal and skew normal density (23) is concluded as (38), (39), (19), (20) and (21), where $\mathbb{F}'(\kappa_{x_n}^{i-1},\varsigma^{i-1})$ in (19) is calculated as (40).

IV. GRADIENT BASED PARAMETER ESTIMATION

We now detail the parameter estimation for the density of each state. To estimate the parameters, we fit the reconstruction

of AMP to the proposed normal and skew normal prior density model (23). It is expected that, the prior density model, and the learned parameters can regularize later AMP reconstructions.

Our strategy is *divide-and-conquer*. First of all, the reconstruction is divided into two sets, *i.e.,* large state set and small state set, according to the sparsity rate[3] $\lambda = K/N$. Specifically, Let T be the set of K largest coefficients of $\underline{\hat{x}} = [\mu_{x_1}^{I_d}, \ldots, \mu_{x_N}^{I_d}]$. Meanwhile, denote T^c as the set containing the remaining $N-K$ coefficients.

For the small state, its variance can be estimated as the unbiased sample variance, *i.e.,*

$$\hat{\sigma}_S^2 = \frac{1}{N-K-1} \sum_{\hat{x}_i \in T^c} \hat{x}_i^2. \tag{41}$$

Given the large state set T, the parameters are estimated by maximizing the log-likelihood of the large state set T, with respect to ω_L and α_L, *i.e.,*

$$\omega_L^*, \alpha_L^* = \underset{\omega_L, \alpha_L \in \mathbb{R}}{\arg\max}\, \ell(T; \omega_L, \alpha_L), \tag{42}$$

where

$$\ell = K \log \frac{2}{\omega_L} - \frac{1}{2} \sum_{\hat{x}_i \in T} \left(\frac{\hat{x}_i}{\omega_L} \right)^2 + \sum_{\hat{x}_i \in T} \log \left(\Phi\left(\alpha_L \frac{\hat{x}_i}{\omega_L} \right) \right). \tag{43}$$

Besides, the gradients of ω_L and α_L with respect to (47) are found to be,

$$\frac{d\ell}{d\omega_L} = -\frac{K}{\omega_L} + \sum_{\hat{x}_i \in T} \frac{\hat{x}_i^2}{\omega_L^3} - \frac{\alpha_L}{\omega_L^2} \sum_{\hat{x}_i \in T} \frac{\phi\left(\alpha_L \frac{\hat{x}_i}{\omega_L} \right)}{\Phi\left(\alpha_L \frac{\hat{x}_i}{\omega_L} \right)} \hat{x}_i, \tag{44}$$

$$\frac{d\ell}{d\alpha_L} = \frac{1}{\omega_L} \sum_{\hat{x}_i \in T} \frac{\phi\left(\alpha_L \frac{\hat{x}_i}{\omega_L} \right)}{\Phi\left(\alpha_L \frac{\hat{x}_i}{\omega_L} \right)} \hat{x}_i. \tag{45}$$

With (43) and gradients (44) (45), one can choose from a variety of solvers, including *trust-region-reflective* [31], [32], *interior-point* [33] algorithms to find the optimum ω_L^* and α_L^*, after which $\hat{\underline{\Theta}} = [\hat{\sigma}_S^2, \omega_L^*, \alpha_L^*]$ is fed back to the approximate message passing (24).

It should be noted that (43) is not convex in general. As a result, the proposed gradient estimator can only find local solutions, and a good initialization strategy becomes consequential for our task.

In this work, we find that initializing ω_L and α_L such that the expected mean and variance of the skew normal density match the sample mean and variance of the large state coefficients of $\underline{\hat{x}}$ works satisfactorily. Therefore, ω_L is initialized at,

$$\omega_0 = \sqrt{\mu_T^2 + \sigma_T^2}, \tag{46}$$

[3]As [18], the sparsity rate, $\lambda = K/N$, is assumed to be known at the reconstruction stage.

TABLE III
AVERAGE RUNNING TIME (IN NANOSECONDS, 10^{-9} SECONDS) OF FREQUENTLY EVALUATED FUNCTIONS

add	multiply	divide	square	ϕ	Φ	ϕ/Φ
5.93	6.08	6.24	6.39	30.42	34.94	80.5

and the initial value of α_L can be found by solving the following,

$$\frac{\alpha_0^2}{1+\alpha_0^2} = \frac{\pi}{2} \frac{\mu_T^2}{\mu_T^2 + \sigma_T^2}, \tag{47}$$

where μ_T and σ_T^2 are the sample mean and variance of large state set T.

Give the reconstruction, the noise variance can be estimated based on the residual, *i.e.,*

$$\hat{\sigma}_e^2 = \frac{1}{M} \sum_{m=1}^{M} (y_m - A_{m\cdot} \underline{\hat{x}})^2. \tag{48}$$

V. COMPLEXITY ANALYSIS

Thanks to the efficient AMP framework, and together with the fast gradient-based parameter estimation, our proposed technique is highly computationally effective.

Similar to [15], [16], [27], the complexity of each message passing iteration in our AMP module is dominated by multiplying sampling matrix $A \in \mathbb{R}^{M \times N}$ with vector $\underline{\hat{x}} \in \mathbb{R}^{N \times 1}$. Besides, as can be seen in (44) and (45), the parameter estimation module involves only vector operations. This makes the complexity of our proposed technique dominated by the AMP module.

It is worth pointing out that in our derivation, evaluating functions including $\phi(\cdot)$, $\Phi(\cdot)$, as well as their division $\phi(\cdot)/\Phi(\cdot)$, will incur sizable computation overhead. Table III summarizes the running time of several frequently evaluated functions in our scheme.

The test is implemented in Matlab [34] and is performed on a computer with dual core 2.67 GHz CPUs, and 8 GB of 1333 MHz RAM, where the input argument of each function is a scaler, and the results are the average of 10^8 random and independent trials.

It can be seen in Table III, compared to scaler addition, evaluating $\phi(\cdot)$ and $\Phi(\cdot)$ are generally 5 to 6 times slower, while the division $\phi(\cdot)/\Phi(\cdot)$ is about 14 times slower.

Therefore, as a rule of thumb, a Floating Point Operations (FLOP) proportional to $10M \times (2N-1) \approx 20MN$ is expected at each iteration. As will be seen in the test, the runtime of our proposed method scales decently as the signal dimensionality N increases, making it one of most efficient techniques in the community.

VI. SIMULATIONS

In this section, the performance of our proposed method is evaluated under phase transition, noisy reconstruction, support set recovery, and runtime tests. Besides, our technique is examined under two real world applications, *i.e.,* MIMO communication and weather sensor network.

The sampling matrix A is generated from standard Gaussian ensemble, with each column being normalized to unit norm, *i.e.*, $\|A_{\cdot n}\|_2 = 1$, for $n = 1, \ldots, N$.

In reconstruction, our method alternates between approximate message passing and parameter estimation. Unless otherwise specified, these two modules execute up to 8 times, or stopped when consecutive normalized reconstruction difference $\|\underline{\hat{x}}^{new} - \underline{\hat{x}}^{old}\|_2^2 / \|\underline{\hat{x}}^{new}\|_2^2 \leq 10^{-6}$. In approximate message passing, $\mu_{x_n}^1$ is initialized at 0 for $n = 1, \ldots, N$, $\mu_{y_m}^1$ is set to y_m for $m = 1, \ldots, M$. Besides, ς^1 (20) is set to 10^4 to make the inference robust. The message passing is executed up to 50 iterations, or until the convergence, which is claimed when $\|\underline{\hat{\mu}}^{i+1} - \underline{\hat{\mu}}^{i}\|_2$ is less than 10^{-7}, where $\underline{\hat{\mu}}^i = [\mu_{x_1}^i, \ldots, \mu_{x_N}^i]$.

In estimating the parameters, we employ the classic *trust-region-reflective* [31], [32] as the optimizer, where ω_L is bounded by $[0, \infty]$. Additionally, α_L is bounded by $[-15, 15]$ for numerical stability. The optimization is terminated after 500 iterations, or when the consecutive log-likelihood difference $\leq 10^{-6}$, whichever comes earlier.

A. Phase Transition

In the first test, the proposed method is examined under the empirical phase transition test [28]. The support set of the signal is generated uniformly at random, namely, index $n = 1, \ldots, N$ is sampled with a uniform probability $\lambda = K/N$. In generating the magnitude of the significant coefficients, two cases are considered.

In the first case, the significant coefficients are generated identically and independently from normal distribution $\mathcal{N}(0, \sigma^2)$, where the standard deviation σ follows a prior uniform distribution $\mathcal{U}[5, 25]$. Besides, the significant coefficients are made strictly non-negative by taking the absolute values.

In the second case, the significant coefficients are generated identically and independently from uniform distribution $\mathcal{U}[b_l, b_u]$, where the lower bound follows a prior uniform distribution $b_l \sim \mathcal{U}[-20, 0]$, and the upper bound follows a prior uniform distribution $b_u \sim \mathcal{U}[0, 200]$.

The insignificant coefficients are generated from normal distribution with mean 0 and variance 10^{-4}. In the first execution of approximate message passing module, $\underline{\Theta} = [\sigma_S^2, \omega_L, \alpha_L]$ is set to $[10^{-5}, 50, 0]$, which in later executions, will be updated at the solution found by the gradient based parameter estimator $\underline{\hat{\Theta}} = [\hat{\sigma}_S^2, \omega_L^*, \alpha_L^*]$. The signal length is set to $N = 1000$. Meanwhile, M/N is set from 0.05 to 0.5 at steps of 0.025. For each value of M/N, K/M is varied from 0 to 1 at steps of 0.025. 500 independent trials are executed for each combination of M/N and K/M, and the Normalized Square Error (NSE), evaluating as $\text{NSE} \triangleq \|\underline{\hat{x}} - \underline{x}_{true}\|_2^2 / \|\underline{x}_{true}\|_2^2$ with $\underline{x}_{true}$ denoting the ground truth, is recorded for each trial.

As in [28], the maximum value of K/M, up to which the corresponding success rate is $\geq 50\%$ is registered. Besides, a success trial is defined as the one with $\text{NSE} \leq 10^{-4}$.

The performance is compared with two AMP based techniques, namely EMGMAMPMOS [20], and EMNNAMP [21]. Besides, SPGL1 [35], and CVX [36] are included in the comparison to solve LASSO [14]. It should be noted, CVX [36] is aided with side information, where the optimization is constrained with upper bound being the maximum of $\underline{x}_{true}$, and lower bound being the minimum of $\underline{x}_{true}$. Additionally, several powerful Bayesian and greedy algorithms, including Sparse Bayesian Learning (SBL) [37], Bayesian Compressive Sensing (BCS) [38], and Orthogonal Matching Pursuit (OMP) [39] are also included in the tests. Furthermore, since the sparsity rate λ is assumed to be known in our scheme, for fairness, the sparsity ratio in EMGMAMPMOS [20], EMNNAMP [21], and OMP [39] are fixed to $\lambda = K/N$. The simulation results are plotted in Figures 3a and 3b.

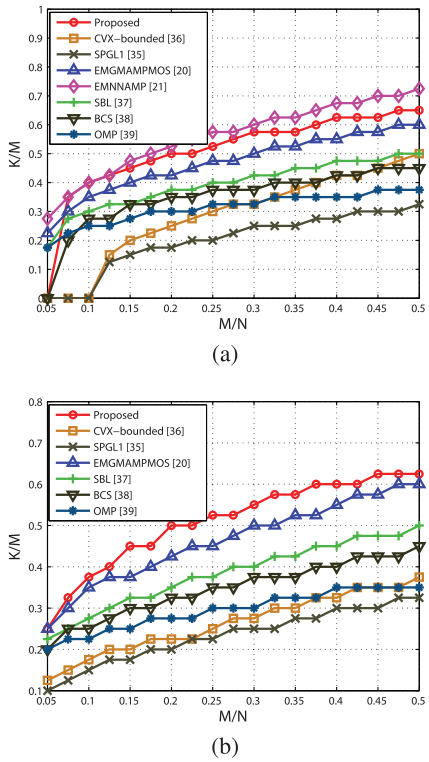


Fig. 3. Phase transition test. (a) Significant coefficients are strictly non-negative. (b) Significant coefficients are a mix of positive and negative elements. The signal length is set to $N = 1000$.

As can be seen in Fig. 3a where all significant coefficients are strictly non-negative, EMNNAMP [21] gives the benchmark phase transition curve by taking advantage of the non-negative normal density mixture. It is also noted that in Fig. 3a, although without any prior of the non-negativity, our proposed scheme is capable of exploiting the asymmetric feature of the significant coefficients, and provides very competitive performance.

Since EMNNAMP [21] is designed exclusively for non-negative signals, its plot is omitted in Fig. 3b, where the significant coefficients consist of both positive and negative components. As can be seen in Fig. 3b, comparing to many sophisticated techniques, our method provides the most competitive performance. This shows our technique can effectively exploit the underlining skewness of the signal, while being sufficiently flexible to accommodate both positive and negative elements.

B. Noisy Reconstruction

In the second test, the performance of our technique is examined under noisy environments. The significant coefficients are generated in ways similar to previous phase transition test. To make the reconstruction more challenging, unlike the phase transition test where the magnitudes of insignificant coefficients are negligible, in this test, the insignificant coefficients are generated from normal distribution with mean 0 and variance 0.5. It should be noted that similar setups, referred to as *heavy-tailed tests* [4], can be found in [20] where Student's-t and log-normal prior densities are utilized to generate the signals.

The length of the signal is set to $N = 500$. The number of significant coefficients K is set to 50, and the number of samples M is set to 125. The noise vector $\underline{e}$ is sampled from Gaussian density, *i.e.,* $\underline{e} \sim \mathcal{N}(\underline{0}, \sigma_e^2 \mathrm{I}_{M\times M})$, and is added to the measurement. The variance of the noise, σ_e^2, is adjusted such that $\mathrm{SNR} = 10\log_{10}(\|A\underline{x}_{true}\|_2^2/\|\underline{e}\|_2^2)$ is varied from 10 dB to 30 dB at 2 dB increments. Meanwhile, $\underline{\Theta} = [\sigma_S^2, \omega_L, \alpha_L]$ are set to [1, 50, 0] at the first execution of approximate message passing module. The noise variance is initialized at 1, and is estimated as (48) in later reconstruction iterations.

The performance of our proposed technique, EMGMAMPMOS [20], EMNNAMP [21], SPGL1 [35], and CVX (bounded) [36], SBL [37], BCS [38], and OMP [39] are compared and the results are summarized in Fig. 4, where each data point is the average of 500 independent trials. As can be seen, our technique yields superior results in both Fig. 4a and 4b.

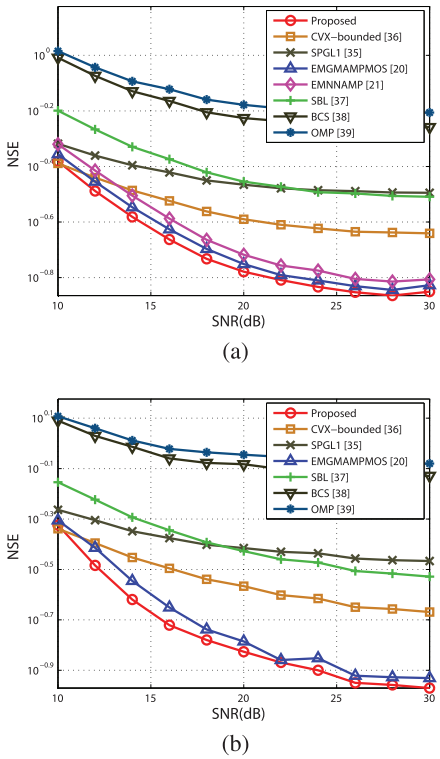


Fig. 4. NSE vs. SNR. (a) Significant coefficients are strictly non-negative. (b) Significant coefficients are a mix of positive and negative elements.

C. Support Set Recovery

In this test, the capability of support set recovery is examined. As previous tests, two types of signals are generated, *i.e.,* strictly non-negative, and mix of positive and negative, where for each type of signals, the parameters characterizing both significant state and insignificant state, as well as the initialization of $\underline{\Theta} = [\sigma_S^2, \omega_L, \alpha_L]$, are set identical to those of phase transition test. The measurement is noiseless.

The length of signal is set to $N = 500$, and the number of random samples is fixed at $M = 125$. We gradually vary the number of significant coefficients K by adjusting K/M from 0.025 to 1, at steps of 0.025, where for each value of K, 500 independent random trials are performed. Besides, the support set recovery rate is calculated by counting the trail with correct recovery of support set, *i.e.,* the trial whose estimated support set matches exactly with the ground truth. Since not all techniques are able to yield strictly sparse solutions, a threshold of 0.1 is applied to get the estimated support from the raw reconstruction.

We compare our proposed technique with EMGMAMPMOS [20], EMNNAMP [21], SPGL1 [35], and CVX (bounded) [36], SBL [37], BCS [38], and OMP [39], and the results are plotted in Figures 5a and 5b. As can be seen, for each method, the support set recovery rate decays with increasing K. Yet, thanks to ability of exploiting the asymmetrical feature of the signal, our proposed technique is capable of providing reliable support recovery over a decently large region of K in both Figures 5a and 5b.

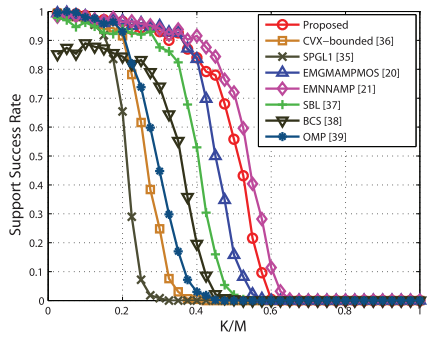
[4]By default, the *heavy-tailed tests* in EMGMAMPMOS [20] assumes a symmetrical signal, and the means of the density components are fixed to 0. For fairness, we turn on the update of means for EMGMAMPMOS.

D. Runtime

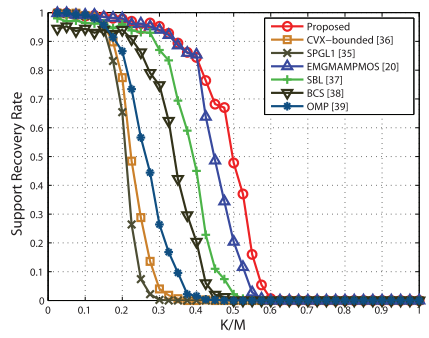
We are now testing the Runtime of our proposed technique. In this test, the length of signal, N, is varied from 500 to 5000, at steps of 500. Meanwhile, without lose of generality, we fix $M/N = 0.5$, and $K/M = 0.4$, for all values of N. Signals are generated such that all significant coefficients are strictly positive, where the characterizing parameters of the densities, as well as the initialization of $\underline{\Theta} = [\sigma_S^2, \omega_L, \alpha_L]$, are set similar as phase transition test. The test is performed on a computer with hex core 2.0 GHz CPUs, and 32 GB of 1333 MHz RAM.

We compare the runtime of our technique with EMGMAMPMOS [20], EMNNAMP [21], SPGL1 [35], and CVX (bounded) [36], SBL [37], BCS [38], and OMP [39], where all methods are implemented with Matlab [34].

The mean runtime are plotted in Fig. 6, where each data point is the average of 50 independent trials. Clearly, similar to the two AMP relatives, namely EMGMAMPMOS [20] and EMNNAMP [21], our proposed technique is computationally effective. This advantage is most remarkable under relatively large signal dimensionality. For example, when $N = 5000$, our technique yields an average runtime of 6.204 Seconds (sec), which is more than 650 times faster than CVX, and 6.27 times faster than SPGL1. Besides, comparing to OMP, our technique runs 26 times faster. Moreover, our technique has advantage over Bayesian algorithms, with BCS and SBL being 3.73 and 299 times slower.



(a)



(b)

Fig. 5. Support Recovery Rate vs. K/M (a) Significant coefficients are strictly non-negative. (b) Significant coefficients are a mix of positive and negative elements.

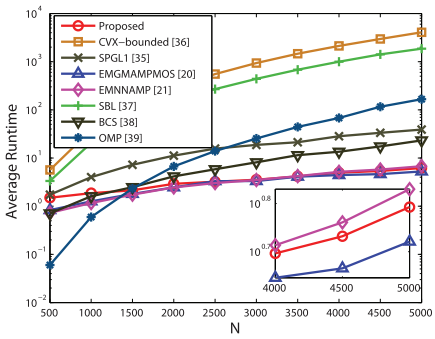


Fig. 6. Signal Length N vs. Average Runtime (in seconds).

E. Communication Throughput Test

In this test, the performance of our technique is evaluated under real world communication application, *i.e.,* multibeam opportunistic random-beamforming network [3], [40]. In [40], a communication scheme is proposed to study the multiuser capacity in MIMO broadcast channels. By broadcasting random beams to all users and requesting their SINRs, Base Station is able to dedicate its antennas to users with stronger SINR. Therefore, compared to randomly allocating antennas, higher sum-rate capacity can be achieved, at the cost of the associated feedback overhead [40]. It is further shown that in [3], similar sum-rate capacity can be realized with substantially reduced feedback overhead, by requesting those SINRs above certain threshold from the corresponding users.

The uplink feedback of SINRs is then modelled by compressive sensing. Specifically, as (10) of [3], the reception at Base station can be represented as $\underline{y} = B\underline{x} + \underline{e}$, where $\underline{x} \in \mathbb{R}^{N\times 1}$ is the SINRs of all users [3], [40], and $\underline{e} \in \mathbb{R}^{M\times 1}$ is the white Gaussian noise vector, with independent and identically distributed (*i.i.d.*) Gaussian element $e_m \sim \mathcal{N}(0, \sigma_e^2)$. Besides, the matrix $B = [b_{\cdot 1}, \ldots, b_{\cdot N}] \in \mathbb{R}^{M\times N}$ is formed by concatenating signature sequence vectors [3], *i.e.,* $b_{\cdot n} \in \mathbb{R}^{M\times 1}$, with its entry following Gaussian distribution $\mathcal{N}(0, 1/M)$. Additionally, in this context, N is denoting the number of single-antenna mobile users, M is denoting the number of time slots designated at the uplink feedback channel.

With the reception of under-sampled SINRs, LASSO is then applied at Base Station to get an estimate of SINRs [3], after which, the Base Station antenna can be allocated to user with the strongest SINR. This process is repeated Q times, such that all Q antennas at BS are allocated[5].

In this test, SINRs are generated as (2). Specifically, similar to [3], in the downlink, the q-th Base Station antenna broadcasts random beams $\underline{\psi}_q \in \mathbb{R}^{Q\times 1}$ to all users, with $[\underline{\psi}_1, \ldots, \underline{\psi}_Q] \in \mathbb{R}^{Q\times Q}$ being orthonormal and $q = 1, \ldots, Q$. Meanwhile, elements of channel vector $\underline{g}_n \in \mathbb{C}^{Q\times 1}$ are *i.i.d.* following complex Gaussian distribution, with zero mean and unit variance. Additionally, a homogeneous communication environment is considered [3], where the SNR is set to $\rho_n = 10$ in (2) for $n = 1, \ldots, N$.

$\underline{x} \in \mathbb{R}^{N\times 1}$ is generated by keeping the largest S elements out of the N SINRs, and making the remaining $N - K$ elements to 0. Besides, the available time slots in the uplink channel is set to $M = 10 \times S$, where $S = 10$. Additionally, the number of antennas at Base Station is set to $Q = 4$. The uplink channel is noisy, with $\sigma_e = 0.5$.

At the Base Station, we apply our proposed method to reconstruct the SINRs. The achievable sum rate is calculated similar to (29) of [3], which is proportional to the support set recovery rate. It should be noted in [3], *exact model recovery rate* [41] of LASSO [14] is employed as the support set recovery rate. In our test, the support set recovery rate is empirical, where a threshold of $0.2 \times \rho_n$ is applied to get the estimated support from the raw reconstruction.

Similar to previous tests, we compare our technique with EMGMAMPMOS [20], EMNNAMP [21], SPGL1 [35], and CVX (bounded) [36], SBL [37], BCS [38], and OMP [39]. Besides, $\underline{\Theta} = [\sigma_S^2, \omega_L, \alpha_L]$ is initialized similar to phase transition test, and the noise variance is estimated as (48).

Fig. 7 shows the achievable sum rate obtained from the variant methods, where each data point is averaged on 500 independent trials, and the number of users N is varied from 100 to 1000, at steps of 100. It can be seen in Fig. 7, the achievable sum rate increases when more users are involved. Meanwhile, given the same feedback overhead, *i.e.,* $S = 10$ and $M = 10 \times S$, our proposed technique yields competitive achievable sum rate by taking advantage of the asymmetrical property of SINRs. Moreover, comparing to [3] where LASSO (SPGL1 [35]) is used to reconstruct SINRs, our proposed technique provides

[5]For ease of notation, we omit the superscripts denoting the BS antennas of B, $\underline{x}$, $\underline{y}$, and $\underline{\hat{x}}$.

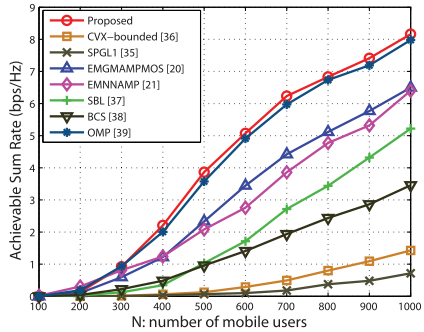


Fig. 7. Achievable Sum-Rate (bps/Hz) vs. Number of Mobile Users N.

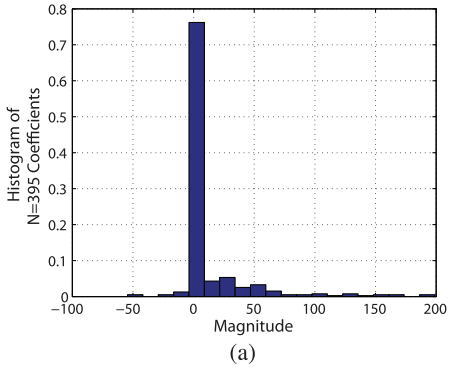
constantly higher achievable sum-rate. It is also noticed that in Fig. 7, although has being developed for decades, OMP gives very competitive results in this test.

F. Weather Data Test

In the last test, we are interested in evaluating our proposed technique with a dataset collected from a real weather sensor network. The data is referred to as *cooling degree day departure from normal* [42]. Cooling degree day is derived from outside air temperature, and is widely used in estimating the energy needed to cool a structure [42]. The phrase *departure from normal* suggests that a 30-year historical average is subtracted from the data. Our data is obtained from Automated Surface Observing System (ASOS) [43], and is accessible at National Climate Data Center [42].

The data is of length $N = 395$, and has $K = 143$ nonzero coefficients. As can be seen in the histogram plotted in Fig. 8a, the nonzero coefficients are asymmetrically positive. As previous tests, the data is down-sampled by projection with a Gaussian random sampling matrix A. The measurement is noisy, and the noise variance σ_e^2 is adjusted such that the SNR is varied from 10 dB to 30 dB at 2 dB increments. For each value of SNR, 100 realizations of random sampling matrix A are generated, and $M = 2K$. For each trial, our method performs approximate message passing decoding and parameter estimation up to 8 times, or stopped $\|\hat{\underline{x}}^{new} - \hat{\underline{x}}^{old}\|_2^2/\|\hat{\underline{x}}^{new}\|_2^2 \leq 10^{-2}$. Additionally, $\underline{\Theta} = [\sigma_S^2, \omega_L, \alpha_L]$ is initialized similar to phase transition test, and the noise variance is estimated as (52).

We compare our technique with EMGMAMPMOS [20], SPGL1 [35], and CVX (bounded) [36], SBL [37], BCS [38], and OMP [39]. Since the significant coefficients contain negative elements, EMNNAMP [21] is excluded from the test. Fig. 8b summarizes the reconstruction NSE as SNR varies. Overall, our scheme provides satisfactory results in most of the range. It is noteworthy that, although not being designed for asymmetrical signals, BCS [38] gives very competitive results by exploiting the sparsity of the signal in this test.



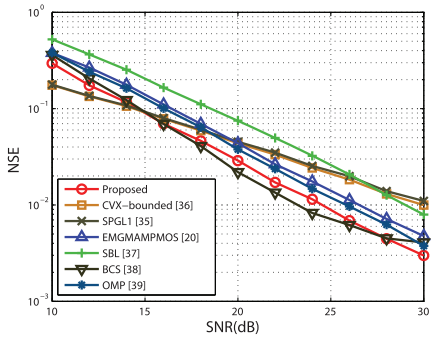


Fig. 8. Temperature Data Test (a) Histogram of the temperature data. (b) NMSE vs. SNR.

VII. Conclusion

In this work, the compressive sensing of the sparse signals whose significant coefficients are distributed asymmetrically with respect to zero is analyzed. To properly capture the asymmetry, a two-state normal and skew normal mixture density is proposed to model the density of the signal. The significant and insignificant coefficients of such signals are represented by a skew normal distribution and a normal distribution, respectively. An approximate message passing algorithm is then designed to take inference of the signal from the compressive sensing measurement while providing fitting to the model. A gradient-based parameter estimator is put forward to infer the underlining density of each component. Experiment results on simulated data and two real-world data, *i.e.,* MIMO communication system and weather sensor network, show our proposed technique can effectively exploit the asymmetrical feature, and provides competitive results compared to the state-of-the-art techniques.

However, it should be noted that, our proposed technique is most effective with signals which can be well represented by the two state normal and skew normal density mixture model. As an interesting extension of this work, an AMP based algorithm with multi-state skew normal density mixture can be developed. Additionally, aside from the numerical study, rigorous theoretical analysis of our technique is an important topic for future research.

Acknowledgments

S. Wang would like to thank Mrs. Weizhou Xiao for her considerable help in proofreading the paper. The authors would like to thank anonymous reviewers for their valuable inputs that have greatly improved the quality of the paper.

REFERENCES

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

[2] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[3] M. E. Eltayeb, T. Y. Al-Naffouri, and H. R. Bahrami, "Compressive sensing for feedback reduction in MIMO broadcast channels," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3209–3222, Sep. 2014.

[4] P. Cheng *et al.*, "Channel estimation for OFDM systems over doubly selective channels: A distributed compressive sensing based approach," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4173–4185, Oct. 2013.

[5] T. Al-Naffouri, A. A. Quadeer, and G. Caire, "Impulse noise estimation and removal for OFDM systems," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 976–989, Mar. 2014.

[6] N. M. Gowda and A. P. Kannu, "Interferer identification in HetNets using compressive sensing framework," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4780–4787, Nov. 2013.

[7] A. Talari and N. Rahnavard, "Cstorage: Distributed data storage in wireless sensor networks employing compressive sensing," in *Proc. IEEE Global Telecommun. Conf.*, 2011, pp. 1–5.

[8] D. Valsesia, G. Coluccia, and E. Magli, "Graded quantization for multiple description coding of compressive measurements," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1648–1660, May 2015.

[9] S. Pudlewski, A. Prasanna, and T. Melodia, "Compressed-sensing-enabled video streaming for wireless multimedia sensor networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 1060–1072, Jun. 2012.

[10] S. Wang, B. Shahrasbi, and N. Rahnavard, "SRL1: Structured reweighted l1 minimization for compressive sampling of videos," in *Proc. Int. Symp. Inf. Theory*, 2013, pp. 301–305.

[11] U. Nakarmi and N. Rahnavard, "BCS: Compressive sensing for binary sparse signals," in *Proc. IEEE Mil. Commun. Conf.*, 2012, pp. 1–5.

[12] S. Wang and N. Rahnavard, "Binary compressive sensing via sum of L-1 norm and L-infinity norm regularization," in *Proc. IEEE Mil. Commun. Conf.*, 2013, pp. 1616–1621.

[13] B. Shahrasbi, A. Talari, and N. Rahnavard, "TC-CSBP: Compressive sensing for time-correlated data based on belief propagation," in *Proc. IEEE Conf. Inf. Sci. Syst.*, 2011, pp. 1–6.

[14] R. Tibshirani, "Regression shrinkage and selection with the lasso," *J. Roy. Stat. Soc.*, vol. B (Methodological), pp. 267–288, 1996.

[15] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. IEEE Inf. Theory Workshop*, 2010, pp. 1–5.

[16] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: II. Analysis and validation," in *Proc. IEEE Inf. Theory Workshop*, 2010, pp. 1–5.

[17] 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, Jul. 2012.

[18] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

[19] A. Farcomeni and S. Arima, "A Bayesian autoregressive three-state hidden Markov model for identifying switching monotonic regimes in microarray time course data," *Stat. Appl. Genet. Mol. Biol.*, vol. 11, no. 4, 2012.

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

[21] J. Vila and P. Schniter, "An empirical-Bayes approach to recovering linearly constrained non-negative sparse signals," *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4689–4703, Sep. 2014.

[22] A. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from imcomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, vol. 39, pp. 1–17, 1977.

[23] M. Svensén and C. M. Bishop, "Robust Bayesian mixture modelling," *Neurocomputing*, vol. 64, pp. 235–252, 2005.

[24] A. Maleki and D. L. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 330–341, Apr. 2010.

[25] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, pp. 629–654, 2008.

[26] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no.3, pp. 265–274, 2009.

[27] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[28] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.*, vol. 367, no. 1906, pp. 4273–4293, 2009.

[29] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Exploring Artificial Intelligence in the New Millennium*. G. Lakemeyer and B. Nebel Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003, vol. 8, pp. 236–239.

[30] A. Azzalini, "A class of distributions which includes the normal ones," *Scand. J. Stat.*, vol. 12, pp. 171–178, 1985.

[31] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, pp. 418–445, 1996.

[32] T. F. Coleman and Y. Li, "On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds," *Math. Programm.*, vol. 67, no. 2, pp. 189–224, 1994.

[33] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. Optim.*, vol. 9, no. 4, pp. 877–900, 1999.

[34] The MathWorks Inc. *MATLAB 2013a*. Natick, MA, USA: MathWorks Inc., 2013.

[35] E. V. D. Berg and M. P. Friedlander., "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[36] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. D. Blondel, S. P Boyd, and H. Kimura Eds., 2008, pp. 95–110 [Online]. Available: http://link.springer.com/chapter/10.1007/978-1-84800-155-8_7

[37] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[38] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

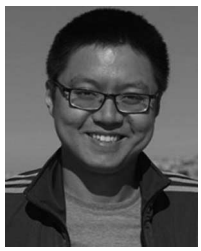
[39] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals Syst. Comput*, Pacific Grove, CA, USA, Nov. 1993, pp. 40–44.

[40] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[41] E. J. Candès and Y. Plan, "Near-ideal model selection by ℓ-1 minimization," *Ann. Statist.*, vol. 37, no. 5A, pp. 2145–2177, 2009.

[42] National Climatic Data Center. (2014, Oct. 28). *Automated Surface Observing System* [Online]. Available: http://www1.ncdc.noaa.gov/pub/download/asos/

[43] National Weather Service. (2014, Sep. 15). *NWS ASOS Program* [Online]. Available: http://www.nws.noaa.gov/asos/



Sheng Wang (S'15) received the B.S. degree in instrumentation and control engineering from Hefei University of Technology, Hefei, Anhui, China, in 2008, and the M.S. degree in instrumentation and control engineering from Tianjin University, Tianjin, China, in 2010. Since 2010, he has been pursuing the Ph.D. degree at the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, USA. His research interests include compressive sensing, statistical signal processing, and novel error control-coding techniques.

Nazanin Rahnavard (S'97–M'10) received the B.S. and M.S. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1999 and 2001, respectively, and the Ph.D. degree from the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2007. Currently, she is an Associate Professor with the Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL, USA. Her research interests include the communications, networking, and signal processing areas. She serves on the Editorial Board of the *Journal on Computer Networks* (Elsevier) and on the Technical Program Committee of several prestigious international conferences. She was the recipient of NSF CAREER Award in 2011. She was also the recipient of the 2007 Outstanding Research Award from the Center of Signal and Image Processing, Georgia Tech.