# SIDE INFORMATION-AIDED COMPRESSED SENSING RECONSTRUCTION VIA APPROXIMATE MESSAGE PASSING

Xing Wang, Jie Liang \*

School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada

## ABSTRACT

In this paper, the side information (SI)-aided compressed sensing reconstruction is considered, where a sparse signal is observed via a noisy underdetermined linear system, and a SI is available during the reconstruction. We develop a SI-aided approximate message passing (SI-AMP) algorithm to solve the problem. Based on the corresponding state evolution formula, the asymptotic prediction performance and noise-sensitivity analysis of the scheme are derived. Simulation results are presented to verify the efficiency of the proposed method.

*Index Terms*— Side information, approximate message passing, phase transition, prediction

# 1. INTRODUCTION AND RELATION TO PRIOR WORK

Recently, compressed sensing of sparse signals has been extensively studied. Many algorithms have been developed for the reconstruction of these signals, including, *e.g.*, convex optimization [1], greedy method [2], and iterative thresholding [3]. To analyze and compare the performances of different reconstruction algorithms, restricted isometry property (RIP) [1] and coherence property [2] are two possible tools. However, they can only provide loose bounds on the reconstruction error, and cannot be used to study the exact performances of an algorithm.

Another approach is to analyze the performance of CS algorithms via estimation theory. In [4], using the replica method that has been widely used in statistical physics, a sharp prediction is derived for the performance of the  $\ell_1$ -regularized least squares problem, or LASSO [5]. However, the replica assumption is not rigorous and it cannot be checked for specific problems.

In [6,7], a fast approximate message passing (AMP) algorithm is developed, based on the classic message passing framework [8]. The AMP is rigorous and can predict the performance of the algorithm accurately. More importantly, it offers a unified framework to exploit further information about the original signal, *e.g.*, structural priors [7], and Gaussian mixture distribution [9].

In many applications, there exists an initial estimation of the sparse signal x as a side information (SI) for reconstruction. For example, adjacent frames in a video sequence are usually very similar. Therefore, an estimation of the current frame can be obtained using motion estimation [10]. Similarly, in multiview video systems, videos in neighboring views also exhibit strong correlation. Therefore, disparity estimation and depth-based image rendering techniques can be used to obtain a prediction of a view from neighboring views [11–14].

There have been some approaches that attempt to exploit various SI in CS. One example is the CS problem with partially known support [15]. In addition, although bounds on reconstruction error are derived in [15], the exact performance of the algorithm is still unknown. Another relevant scheme is to recover the estimation error instead of the sparse signal [16], by assuming that the prediction error between the SI and the sparse signal is sparser than the signal itself. However, this method lacks theoretical analysis. It is also possible that the prediction error is denser than the original signal, if the initial estimation has poor quality. In [14], a squared-errorconstrained penalty term is introduced to the compressed sensing of multiview images. It also considers a more general case, where the variances of the prediction errors are different at different entries.

In [17], the belief-propagation-based CS framework (BPCS) in [18] is used to exploit the SI from neighboring cameras in multiview image systems, although the SI is only used as the starting point for belief propagation.

In this paper, a new model is proposed that in addition to compressed measurements, we can still get a noisy version of the sparse signal. The most relevant framework to ours is the sparsityconstrained dynamic system estimation scheme proposed in [19] and dynamic compressed sensing via approximate message passing (DCS-AMP) proposed in [20,21]. In [19], a prediction of the signal is obtained from the state evolution model, and the norm of the prediction error is added as a penalty term in the objective function of LASSO or BPDN method [5, 22]. In [20, 21], the authors model the sparse signal as Bernoulli-Gaussian distribution and the correlation between active amplitudes in differetn time slots following stationary steady-state Gaussian-Markov process. Expectation maximization (EM) and AMP are applied to learn the hidden parameters and do inference. Note that although the model in [20], [21] is similar to ours, the proposed algorithms rely on the sequential data, can't be applied to solve the problem discussed here directly and is not the main focus of this work.

Such model is prevalent in hybrid distributed sensor network, where because of the low sampling cost of compressed sensing based hardware, some sensors apply compressed sensing based hardware to sense the environment which is sparse in certain basis, e.g., AIC (analog-to-information converter) [23], while other sensors use traditional ADC (analog-to-digital converter). After the data are collected, each sensor sends their data to the central unit to recover the sparse signal jointly. The problem is how to efficiently fuse the compressed measurements and the noisy Nyquist samples. To the best of our knowledge, such problem hasn't been addressed before, although some other papers discussed distributed compressed sensing [24].

Moreover, an important question we address in this paper is : Can me measure how much this SI improves the performance of estimation compared to the classical case without this prior. Several papers have considered similar problems [15], [25]. In [15], the authors have provided guarantees on the performance of similar problems. However, the results are usually inconclusive because

<sup>\*</sup>This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under grant RGPIN312262. Emails: {xingw, jiel}@sfu.ca.

of the loose constant involved in the analyses. Kamilov *et al.* [25] have tried to provide a theoretical understanding of expectation maximization based algorithms [20, 21]. However, the complete understanding of the expectation maximization employed in such methods is not available yet.

In this paper, we develop a fast AMP-based framework to solve the SI-aided LASSO problem, denoted as SI-AMP. We then study its state evolution, which enables the accurate prediction of the MSE performance of the SI-AMP algorithm. We also derive the noisesensitivity analysis of SI-AMP, and show that different from the conventional CS framework, the MSE of the reconstruction using SI-AMP is bounded everywhere int the sampling space and the phase transition doesn't exist any more, thanks to the SI. Simulation results show that the SI-AMP algorithm can achieve better reconstruction than the conventional method. Due to space limitation, some detailed derivations of the main results in the paper are skipped, and the details can be found in the journal version [26].

#### 2. THE SI-AIDED LASSO PROBLEM

In compressed sensing, the observation is given by y = Ax + w, where  $x \in \mathbb{R}^n$  is a k-sparse signal, *i.e.*, with k nonzero entries ( $k \ll n$ ),  $A \in \mathbb{R}^{m \times n}$  is a known linear measurement matrix, with m < n, and  $w \in \mathbb{R}^m$  is the additive noise, often assumed to be white Gaussian with variance  $\sigma^2$ , *i.e.*,  $w \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ . In this paper, the entries of A follow i.i.d. Gaussian distribution  $\mathcal{N}(0, 1/m)$ , and A has zero column mean and unit column norm. The following ratios are also frequently used in the paper.

$$\delta = m/n, \ \varepsilon = k/n, \ \rho = \varepsilon/\delta = k/m.$$
 (1)

The LASSO is a frequently used algorithm to reconstruct sparse signals, which is an  $\ell_1$ -regularized least square optimization [5]:

$$\hat{x}(\lambda,\tau_s) = \operatorname*{arg\,min}_{z \in \mathbb{R}^n} \left( \frac{1}{2} \left\| y - Az \right\|_2^2 + \lambda \left\| z \right\|_1 \right).$$
(2)

An important fact is that the solution of LASSO is equivalent to the soft thresholding algorithm in wavelet denoising [27],

$$\hat{x}^{\lambda}(i) = \eta(y(i); \lambda), \tag{3}$$

where the soft thresholding operation with threshold  $\theta$  is

$$\eta(x;\theta) = \begin{cases} x - \theta & \text{if } x > \theta, \\ 0 & \text{if } -\theta \leqslant x \leqslant \theta, \\ x + \theta & \text{if } x < -\theta. \end{cases}$$
(4)

The threshold in (3) can be chosen as  $\lambda = \alpha \sigma$ , where  $\alpha$  is a scaling parameter. The MSE of the soft thresholding algorithm can thus be written as

$$\operatorname{mse}(\sigma^2; p, \alpha) \equiv E\{[\eta(X + \sigma Z; \alpha \sigma) - X]^2\},$$
(5)

where the expectation is with respect to independent random variables  $Z \sim \mathcal{N}(0, 1)$  and  $X \sim p$ .

For k-sparse X with  $\varepsilon = k/n$  as in (1), we define the following set of probability measures with small non-zero probability,

$$\mathcal{F}_{\varepsilon} \equiv \{p : p \text{ is a probability measure with } p(\{0\}) \ge 1 - \varepsilon\}.$$
 (6)

Thanks to the scale invariance of soft thresholding method [6], we only need to focus on  $\sigma = 1$  when studying the noise sensitivity

of soft thresholding method. In this case, the minimax threshold MSE of the method is defined as [6]

$$M^{\pm}(\varepsilon) = \inf_{\alpha > 0} \sup_{p \in \mathcal{F}_{\varepsilon}} \operatorname{mse}(1; p, \alpha)$$
(7)

which is the minimal MSE of the worst distribution in  $\mathcal{F}_{\varepsilon}$ , where  $\pm$  means a nonzero estimate can take either sign.

In terms of Bayesian theory, the LASSO framework is equivalent to adding a zero-mean Laplace prior distribution of x and then finding the maximum a posteriori (MAP) estimation.

In this paper, we assume that a SI or initial estimation of x, denoted by  $\tilde{x}$ , is available during reconstruction, which can be seen as a noisy version of x. The error of the SI,  $e = \tilde{x} - x$ , is assumed to be white Gaussian with variance  $\sigma_s^2$ , *i.e.*,  $e \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I})$ . In this case, if we apply the MAP theory to estimate x, the solution becomes the following SI-aided LASSO (SI-LASSO) framework [26], where an  $\ell_2$ -norm term is added to incorporate the SI. This provides a theoretical justification to the framework in [19].

$$\hat{x}(\lambda,\tau_s) = \operatorname*{argmin}_{z \in \mathbb{R}^n} \left( \frac{1}{2} \|y - Az\|_2^2 + \lambda \|z\|_1 + \frac{\tau_s}{2} \|\tilde{x} - z\|_2^2 \right).$$
(8)

The parameters  $\lambda$  and  $\tau_s$  are closely related to  $\sigma_s^2$ , the noise variance of the SI. The conventional LASSO is a special case of SI-LASSO with  $\tau_s = 0$ . Another extreme case is when the SI equals to the original signal. In this case, we should set  $\tau_s = \infty$ , and the final reconstruction result is simply  $\tilde{x}$ .

# 3. SI-AIDED APPROXIMATE MESSAGE PASSING

The LASSO problem can be solved by several methods, such as interior method and gradient method. In this paper, we are interested in the AMP algorithm [6, 7], which enjoys several advantages, *e.g.*, low complexity and the capability of predicting the performance.

For Gaussian matrix A, conventional message passing algorithms such as the min-sum method needs to calculate 2mn messages in each iteration. However, in AMP [6, 7], by employing quadratic approximation and the properties of the soft thresholding and the Gaussian matrix A for large values of m and n, the expression of each message can be simplified, and the number of messages can be reduced to m + n. The final formulae can be expressed as

$$\hat{x}_{0}^{t} = x^{t} + A^{T}r^{t},$$

$$x^{t+1} = \eta(\hat{x}_{0}^{t}; \theta_{t}),$$

$$r^{t} = y - Ax^{t} + \frac{\|x^{t}\|_{0}}{m}r^{t-1} \equiv y - Ax^{t} + b_{t}r^{t-1},$$
(9)

where the *n* entries in  $x^t$  are the estimate of *x* in the *t*-th iteration, and the *m* entries in  $r^t$  are the residual with respect to the observation.  $\hat{x}_0^t$  is the un-thresholded updated estimation of *x*.  $\eta(:,:)$  is the soft thresholding in (4), and  $b_t$  is a forgetting or reaction factor when updating the residual. It turns out that the AMP algorithm is very similar to the iterative soft thresholding algorithm [27]. The only difference is the introduction of the term  $b_t r^{t-1}$ . However, this term significantly improves the performance of the algorithm.

In addition to lower complexity, the AMP also allows us to predict the final reconstruction performance of LASSO if necessary parameters are given, by solving the fixed-point equation of state evolution derived from the formulae above.

To derive the SI-aided AMP, we start by modifying the local function of each variable node to be  $\lambda |x_i| + \frac{\tau_s}{2} (x_i - \tilde{x}_i)^2$ . After

this, following the derivation of AMP, the proposed SI-AMP algorithm can be obtained as (details in [26])

$$\hat{x}_{0}^{t} = \frac{u_{t}}{1+u_{t}}\tilde{x} + \frac{1}{1+u_{t}}(x^{t} + A^{T}r^{t}), 
x^{t+1} = \eta(\hat{x}_{0}^{t}; \theta_{t}), 
r^{t} = y - Ax^{t} + \frac{\|x^{t}\|_{0}}{m(1+u_{t-1})}r^{t-1} \equiv y - Ax^{t} + b_{t}r^{t-1},$$
(10)

Comparing (9) and (10), the SI-AMP introduces a weighted average of the SI  $\tilde{x}$  and  $x^t + A^T r^t$ , and the weighting parameter is controlled by  $u_t$ . In addition, the reaction factor  $b_t$  is also a function of  $u_{t-1}$ . This enables SI-AMP to fully exploit the SI and improve both the theoretical and empirical performances of AMP. When there is no SI,  $u_t = 0$ , SI-AMP reduces to AMP.

The following proposition shows that SI-AMP provides a very general solution for the SI-LASSO problem in Eq. (8) [26].

**Proposition 3.1** Let  $(x^*, r^*)$  be the fixed point of the SI-AMP algorithm given by (10) for fixed  $\theta_t = \theta$ ,  $u_t = u$ , and  $b_t = b$ . Then  $x^*$  is also a minimum of the SI-LASSO problem in (8) with

$$\lambda = (1+u)\theta(1-b),$$
  

$$\tau_s = u(1-b).$$
(11)

When there is no SI (u = 0), the result above reduces to Prop. 5.1 in [7] for LASSO.

## 3.1. SI-AMP State Evolution

One attractive feature of AMP is that it gives the formula for the evolution of the MSE of the estimation during the iteration [7], which allows us to accurately predict the final MSE by solving a fixed-point equation. We now derive the MSE state evolution of the SI-AMP. We will also show that there is also a state evolution for the SI weighting parameter  $u_t$  in (10), which is not available in the AMP.

The unthresholded estimator in (10) includes the SI, residual  $r^t$  and interference  $x^t$ . Let sri be its estimation variance, which can be found to be (Appendix A in [26])

$$\operatorname{sri}(q_t^2, u_t; \delta, \sigma, \sigma_s) = \left(\frac{u_t}{1+u_t}\right)^2 \sigma_s^2 + \left(\frac{1}{1+u_t}\right)^2 \left(\sigma^2 + \frac{q_t^2}{\delta}\right), (12)$$

where  $q_t^2$  is the variance of the thresholded estimator  $x^t$ .

ı

~

The optimal  $u_t$  should minimize the sri above. Taking derivative of  $u_t$  and setting it to 0, the optimal  $u_t$  can be found to be

$$u_t = \frac{\sigma^2 + q_t^2/\delta}{\sigma_s^2}.$$
(13)

Therefore, if the quality of the SI is higher than  $x^t + A^T r^t$ , we will have  $u_t > 1$ , the SI will have more weights in constructing the next unthresholded estimate in (10). As the decrease of  $q_t^2$ , it is possible to have  $u_t < 1$ , *i.e.*, the impact of the SI will be gradually reduced.

To find  $q_{t+1}^2$ , the variance of the thresholded estimate  $x_{t+1}$  in (10), we need to use (5). Let

$$\Psi(q_t^2, u_t; \delta, \sigma, \sigma_s, \alpha, p) \equiv \mathsf{mse}(\mathsf{sri}(q_t^2, u_t; \delta, \sigma, \sigma_s); p, \alpha).$$
(14)

Since  $(\delta, \sigma, \sigma_s, \alpha, p)$  are fixed, we can use  $(q_t^2, u_t)$  to represent the state of the algorithm in the *t*-th iteration. The state evolution for  $q_t^2$  can thus be written as

$$q_{t+1}^2 = \Psi(q_t^2, u_t). \tag{15}$$

Eq. (13) can be viewed as the state evolution of  $u_t$ , which is not available in AMP.

To find the final MSE, we substitute  $u_t$  in Eq. (13) to Eq. (12), sri can be simplified into

$$\operatorname{sri}(q_t^2) = \frac{\sigma_s^2(\sigma^2 + q_t^2/\delta)}{\sigma_s^2 + \sigma^2 + q_t^2/\delta}.$$
(16)

When the algorithm converges, the steady-state or fixed-point condition of the state evolution in (15) is

$$q_*^2 = \Psi(q_*^2, \frac{\sigma^2 + q_*^2/\delta}{\sigma_s^2}) = \text{mse}(\text{sri}(q_*^2); p, \alpha).$$
(17)

Define sri $(q_*^2)$  as a new variable  $\xi^2$ , and plug (17) into (16), we can get the following fixed-point equation for  $\xi^2$ .

$$\xi^2 = \frac{\sigma_s^2(\sigma^2 + \operatorname{mse}(\xi^2; p, \alpha)/\delta)}{\sigma_s^2 + \sigma^2 + \operatorname{mse}(\xi^2; p, \alpha)/\delta} \equiv F(\xi^2, \alpha).$$
(18)

With an appropriate choice of  $\alpha$  (Prop. 4.3 in [26]), the fixedpoint equation has a unique solution  $\xi_*$ , which is the final MSE performance of the SI-AMP algorithm.

## 4. NOISE SENSITIVITY ANALYSIS OF SI-AMP

The noise sensitivity phase transition is a curve in the  $(\rho, \delta)$  plane, indicating the sensitivity of a CS reconstruction method to the noise variance in the measurement [6], where  $\rho = k/m$  and  $\delta = m/n$ , as defined in (1). For many classical CS algorithms, the MSE under the phase transition curve is bounded, and the MSE is unbounded above the curve. In this section, we show that because of the SI, the phase transition diminishes, and the MSE is bounded everywhere in the sampling space and the phase transition property of classical CS problem is just a special case where the variance of the SI is  $\infty$ .

In this section, several minimax risks are used.  $M^{\pm}(\delta\rho)$  is defined in (7).  $M^{b}(\delta,\rho)$  and  $M^{*}(\delta,\rho,\gamma_{s}^{2})$  are the minimax risk of LASSO and SI-LASSO respectively. The parameter  $\gamma_{s}^{2} = \sigma_{s}^{2}/\sigma^{2}$  plays an important role, which measures the relative quality of the SI with respect to the quality of underdetermined linear measurements.

The following result shows that for any point in the sampling space, the minimax risk of SI-LASSO is bounded.

**Proposition 4.1** For any point in the sampling space, i.e.,  $\rho < 1/\delta$ , i.e.,  $\varepsilon = \delta \rho \leq 1$ , the SI-LASSO minimax risk is bounded, and  $M^*$  is given by

$$M^{*}(\delta,\rho,\gamma_{s}^{2}) = \frac{2\delta\gamma_{s}^{2}M^{\pm}(\delta\rho)}{G(\delta,\rho,\gamma_{s}^{2}) + \sqrt{G(\delta,\rho,\gamma_{s}^{2})^{2} + 4\delta\gamma_{s}^{2}M^{\pm}(\delta\rho)}},$$
(19)
where  $G(\delta,\rho,\gamma_{s}^{2}) = \delta\gamma_{s}^{2} + \delta - \gamma_{s}^{2}M^{\pm}(\delta\rho).$ 

The corresponding parameter selection rules for minimax risk case are listed in Proposition 5.1 of [26].

It is easy to verify that  $\partial M^*(\delta, \rho, \gamma_s^2)/\partial \gamma_s^2$  is monotonically increasing, so  $M^*(\delta, \rho, \gamma_s^2)$  is an increasing function of  $\gamma_s^2$ . Since SI-AMP reduces to AMP when  $\gamma_s^2 = \infty$ , this means that the minimax bound of SI-LASSO is no greater than that of LASSO, *i.e.*,

$$M^*(\delta,\rho,\gamma_s^2) \leqslant M^b(\delta,\rho) = \frac{M^{\pm}(\rho\delta)}{1 - M^{\pm}(\rho\delta)/\delta},$$
 (20)

where  $M^{b}(\delta, \rho)$  is the bound of LASSO minimax risk.

δ	ρ	$h^*$	$\lambda^*$	$\tau^*$	fMSE	eMSE	eMSE	fMSE	eMSE	eMSE
					(SI-AMP)	(SI-OWLQN)	(SI-AMP)	(AMP)	(OWLQN)	(AMP)
0.1000	0.0950	2.8279	2.5845	0.9953	0.0329	0.0316	0.0326	0.1362	0.1193	0.1276
0.1000	0.1800	2.7988	2.2226	0.9920	0.0577	0.0582	0.0576	2.0626	1.9584	3.1587
0.1000	1.9000	2.6558	0.9194	0.9513	0.4054	0.4045	0.4057	UB	UB	UB
0.250	0.1340	2.5808	2.0253	0.9954	0.0858	0.0913	0.0875	0.3738	0.3689	0.3659
0.2500	0.2540	2.5287	1.6631	0.9924	0.1447	0.1453	0.1483	5.5758	6.6647	5.6803
0.2500	1.9000	2.2759	0.5112	0.9729	0.6185	0.6252	0.6264	UB	UB	UB
0.5000	0.1930	2.3619	1.5123	0.9954	0.1815	0.1837	0.1839	0.8528	0.8447	0.8559
0.5000	0.3660	2.2846	1.1402	0.9930	0.2908	0.2957	0.2903	12.4266	15.6648	12.1652
0.5000	1.9000	1.2526	0.0469	0.9859	0.6887	0.6890	0.6962	UB	UB	UB

Table 1. Empirical and predicted MSEs of different methods for points in the sampling space

When there is no SI, the formal MSE noise sensitivity above the phase transition is infinite. However, this is no longer the case in the presence of the SI, as we can at least assign  $\tau_s$  to  $\infty$  while keeping  $\lambda$  to be finite, and the formal MSE noise sensitivity is thus bounded by  $\gamma_s^2$ . We can do even better by exploiting the measurement and the sparsity of the original signal, which still follows the statement above.

There are three extreme cases here: First, let  $\gamma_s^2 = \infty$ , then Equ. (19) degrades to the formulae of bounded MSE below the phase transition boundary of AMP. Second, if  $\gamma_s^2 = 0$ , i.e., the side information is exactly the same as the unknown sparse signal, we don't need to do reconstruction, and MSE is 0 which coincides with Equ. (19) where  $\gamma_s^2 = 0$ . Last, if  $\delta = 0$ , which means there is no compressed measurements, solving the minimization problem (8) is equal to do denoising in scalar case, and the minimax MSE is  $M^{\pm}(\rho\delta)\sigma_s^2$ , which also agrees with the denoising of scalars introduced in the background part of [6].

### 5. NUMERICAL EXPERIMENTS

We first evaluate the accuracy of the MSE of the SI-LASSO predicted by SI-AMP, and compare SI-LASSO with the LASSO. The entries of  $x_0$  are randomly chosen from  $\{+1, 0, -1\}$  with probabilities  $P(x_{0,i} = +1) = P(x_{0,i} = -1) = 0.064$ . The variance of the noise w is  $\sigma^2 = 0.2$ . The SI  $\tilde{x}$  is obtained by adding to  $x_0$  a zero-mean Gaussian noise vector e with variance  $\sigma_s^2 = 0.2\gamma_s^2$ .

In this experiment, two values of n are tested, n = 200 and n = 2000 respectively. For n = 200, the empirical results of LASSO are obtained by CVX [28], which is an interior point-based convex solver written in Matlab. The empirical results of SI-LASSO are obtained by modifying CVS to incorporate the SI. For n = 2000, the empirical results of LASSO are obtained by OWLQN [29], which is a large-scale LASSO solver written in C++. The empirical results of SI-LASSO in this case are obtained by modifying the OWLQN to incorporate the SI.

Fig. 1 shows the predicted and empirical MSEs of LASSO and SI-LASSO. When  $\gamma_s^2 = \infty$ , SI-LASSO reduces to LASSO, we get the same curve as Fig. 9 in [7]. In each case, the predicted MSE is quite accurate in both LASSO and SI-LASSO. When  $\gamma_s^2 = 4$ , the minimal MSE of SI-LASSO is 20% less than LASSO. When  $\gamma_s^2 = 1$ , the SI is more accurate, and the minimal MSE is 60% less.

We now compared the performances of LASSO and SI-LASSO for n = 2000. In addition to SI-OWLQN, we also implement the SI-AMP in (10) by Matlab. We generate in each case 20 random realizations, with parameters  $\delta \in \{0.10, 0.25, 0.50\}, \gamma_s^2 = 1, \sigma^2 = 1$  and  $\rho \in \{\frac{1}{2}\rho(\delta), \frac{19}{20}\rho(\delta), 1.9\}$ . The results are summarized in Table 1, where eMSE and fMSE denote the empirical MSE and theoretical MSE respectively. The optimal  $\lambda$  and  $\tau$  are also listed. Note that  $\rho = 1.9$  is far above the phase transition of AMP, and it's very close



Fig. 1. The predicted and empirical MSEs of LASSO and SI-LASSO with different  $\lambda$  and  $\gamma_s^2$ . The sample rate is  $\delta = 0.64$ .

to the maximum  $\rho$  for  $\delta = 0.5$  where the maximum  $\rho$  is 2.

Some observations can be drawn from Table 1. First, the MSE in SI-AMP is much lower than that in AMP. Secondly, the fMSE and eMSE of SI-AMP match very well, even when the operating point is far above the phase transition boundary of AMP. For example, for  $\delta = 0.100$ , the fMSE of SI-AMP with  $\rho = 1.9$  is still very similar to the eMSE. For AMP, this  $\rho$  is much higher than its phase transition boundary. Its MSE is thus unbounded and represented by "UB" in the table. In addition, the empirical MSE of SI-OWLQN is very similar to that of SI-AMP. However, SI-OWLQN is much slower. For example, in a computer with Intel Core i7 3.07GHz CPU and 6.00 GB memory, our Matlab implementation of SI-AMP is already about 10 times faster than the C++ implementation of SI-OWLQN. This is because OWLQN needs to calculate the gradient in each iteration.

# 6. CONCLUSIONS

This paper studies the side-information (SI)-aided CS problem, where an additional noisy version of the original signal is available for CS reconstruction which is prevalent in hybrid distributed sensor network. We formulate a SI-LASSO problem from the MAP rule, develop a SI-aided approximate message passing algorithm (SI-AMP), and study its state evolution and noise sensitivity. Compared to conventional LASSO algorithms, the better performance of SI-AMP has been verified both theoretically and experimentally.

## 7. REFERENCES

- E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, 2008.
- [2] J. A. Troop and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transaction on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [3] A. Beck and M. Teboulle, "A fast iterative shrinkagethresholding algorithm for linear inverse problems," *SIAM Journal in Imgace Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [4] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of map estimation via the replica method and applications to compressed sensing," *IEEE Transaction on Information Theory*, vol. 58, no. 3, pp. 1902–1923, Mar. 2012.
- [5] R. Tibshirani, "Regression shrinkage and selection with the lasso," J. Royal. Statist. Soc., vol. B 58, pp. 267–288, 1996.
- [6] D. Donoho, A. Maleki, and A. Montanari, "The noisesensitivity phase transition in compressed sensing," *IEEE Transaction on Information Theory*, vol. 57, no. 10, pp. 6920– 6941, Oct. 2011.
- [7] A. Montanari, "Graphical models concepts in compressed sensing," *Arxiv.org*, 2011.
- [8] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transaction on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [9] J. P. Vila and P. Schniter, "Expectation-maximization gaussianmixture approximate message passing," *arxiv.org*, Jun. 2013.
- [10] L. Kang and C. Lu, "Distributed compressive video sensing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 1169–1172.
- [11] X. Xiu and J. Liang, "Projective rectification-based view interpolation for multiview video coding and free viewpoint generation," in *Proc. Picture Coding Symposium*, 2009, pp. 1–4.
- [12] D. Pang, X. Xiu, and J. Liang, "Multiview video coding using projective rectification-based view extrapolation and synthesis bias correction," in *Proc. IEEE International Conference on Multimedia and Expo*, 2009, pp. 5–8.
- [13] X. Xiu, D. Pang, and J. Liang, "Rectification-based view interpolation and extrapolation for multiview video coding," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 21, no. 6, pp. 693–707, Jun. 2011.
- [14] X. Wang and J. Liang, "View interpolation confidence aidedcompressed sensing of multiview images," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2013.

- [15] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Transaction on Signal Processing*, vol. 58, no. 9, pp. 4595– 4607, Sep. 2010.
- [16] M. Trocan, T. Maugey, J. E. Fowler, and B. Pesquet-Popescu, "Disparity-compensation compressed-sensing reconstruction for multiview images," in *IEEE International Conference on Multimedia and Expo*, 2010, pp. 1225–1228.
- [17] P. Beigi, X. Xiu, and J. Liang, "Compressive sensing based multiview image coding with belief propagation," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, 2010, pp. 430–433.
- [18] R. Baron, S. Sarvoham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Proc.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.
- [19] A. S. Charles, M. S. Asif, J. Romberg, and C. J. Rozell, "Sparsity penalties in dynamic system estimation," in *Conference on Information Science and Systems*, 2011.
- [20] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Transactions on Signal Processing*, Nov. 2013.
- [21] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Transactions on Signal Processing*, Jan. 2013.
- [22] 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.
- [23] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse, bandlimited signals," *IEEE Trans. Inform. Theory*, pp. 520– 544, Jan. 2010.
- [24] D. Baron, M. F. Duarte, and M. B. Wakin, "Distributed compressive sensing," *arxiv.org*, p. 0901.3403, 2009.
- [25] U. Kamilovr, M. Unser, A. Flectche, and S. Rangan, "Approximate message passing with consistent parameter estimation and applications to sparse learning," *arxiv.org*, p. 1207.3859, 2012.
- [26] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," *arxiv.org, preprint*, 2013.
- [27] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biomefrika*, vol. 81, pp. 425–455, 1994.
- [28] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," Sept. 2013.
- [29] G. Andrew and J. Gao, "Scalable training of ℓ<sub>1</sub>-regularized log-linear models," in *Proc. of International Conference on Machine Learning*, 2007, pp. 33–40.