BAYESIAN FRAMEWORK AND MESSAGE PASSING FOR JOINT SUPPORT AND SIGNAL RECOVERY OF APPROXIMATELY SPARSE SIGNALS

Shubha Shedthikere and A. Chockalingam

Department of ECE, Indian Institute of Science, Bangalore, INDIA

ABSTRACT

In this paper, we develop a low-complexity message passing algorithm for joint support and signal recovery of approximately sparse signals. The problem of recovery of strictly sparse signals from noisy measurements can be viewed as a problem of recovery of approximately sparse signals from noiseless measurements, making the approach applicable to strictly sparse signal recovery from noisy measurements. The support recovery embedded in the approach makes it suitable for recovery of signals with same sparsity profiles, as in the problem of multiple measurement vectors (MMV). Simulation results show that the proposed algorithm, termed as JSSR-MP (joint support and signal recovery via message passing) algorithm, achieves performance comparable to that of sparse Bayesian learning (M-SBL) algorithm in the literature, at one order less complexity compared to the M-SBL algorithm.

Keywords: Sparse signal recovery, approximately sparse signals, support recovery, Bayesian framework, message passing..

1. INTRODUCTION

An approximately sparse signal $\mathbf{x}$ will have few components with large magnitudes and many components with very small but non-zero magnitudes [1],[2],[3]. Support of an approximately sparse signal is defined as the positions of the large coefficients in $\mathbf{x}$. Finding the support of sparse signals is of interest in several diverse applications, including medical imaging, cognitive radio, etc. In several applications, finding sparse solutions (i.e., solutions where only a very small number of entries are non-zero) are of interest [4]. In the context of approximately sparse signals, sparse solutions refer to those solutions which have very few large coefficients and many small non-zero coefficients.

In this paper, we are interested in computing sparse solutions to the linear inverse problem: find $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{y}$, where $\mathbf{y} \in \mathbb{R}^M$ of M measurements is obtained from an approximately sparse signal $\mathbf{x} \in \mathbb{R}^N$ using a measurement matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$, $M < N$. The solution can be formulated as

$$minimize \ \sum_{i=1}^{N} \mathcal{I}_{\{|x_i| > \ \epsilon\}} \ subject \ to \ \|\mathbf{y} - \mathbf{Ax}\|_2^2 \leq \beta, \quad (1)$$

where $\mathcal{I}$ denotes the indicator function and $E(\mathbf{x}) = \sum_{i=1}^{N} \mathcal{I}_{\{|x_i|> \ \epsilon\}}$ is the diversity measure used.[1] The above problem can be cast in a Bayesian framework with appropriate sparsity promoting priors. In addition to signal recovery in this setting, we are interested in support recovery as well. To achieve this, we propose a Bayesian framework and an associated iterative message passing algorithm to jointly recover the signal and support.

[1]Instead of using the criterion $|x_i| > \epsilon$, other appropriate criterion could be used to distinguish the large coefficients from the small ones.

This work in part was supported by Indo-French Centre for the Promotion of Advanced Research (IFCPAR) Project No. 4000-IT-1.

Iterative message passing on graphical models is an efficient tool to solve inference problems. In [2], a message passing algorithm that uses mixture Gaussian as the sparsity promoting priors is presented for recovery of approximately sparse signals measured using sparse matrices (e.g., LDPC-like measurement matrices). In [5], another message passing algorithm, referred to as the approximate message passing (AMP) algorithm that uses Laplacian priors and approximates the messages by Gaussian densities, has been proposed for the recovery of sparse signals for a variety of measurement matrices (e.g., random Gaussian/Bernoulli measurement matrices). In this paper, we propose a message passing algorithm that jointly recovers support and signal; we refer to this algorithm as *joint support and signal recovery via message passing (JSSR-MP) algorithm*. The joint support and signal recovery in our algorithm allows it to be applicable to recovery of signals with same sparsity profile using multiple measurement vectors (MMV). Our work here differs from that in [2], in the following aspect. The algorithm in [2] sends either the samples of the pdf as messages or the parameters of the mixture Gaussian as messages. Whereas, we approximate the mixture by a single Gaussian distribution and pass the parameters of this Gaussian distribution which results in lower complexity even for a dense measurement matrix. Obtaining sparse solutions to linear inverse problems in MMV settings has been studied in [4],[6]. However, these are not based on message passing. The proposed JSSR-MP algorithm, on the other hand, is based on message passing and it is shown to achieve better performance than AMP algorithm in [5] at same complexity order, and performance comparable to that of the sparse Bayesian learning (M-SBL) algorithm in [6] at one order less complexity.

2. PROBLEM FORMULATION

Approximate Sparse Signal Model: Let the approximately sparse signal consist of K 'large' coefficients and $N-K$ 'small' coefficients, where $K << N$ and $\alpha = K/N$ is the sparsity rate. Let $\mathbf{s} = \left[s_1 \, s_2 \, \cdots \, s_K\right]^T \in \left[N\right]^K$ such that $\left\{s_1, s_2, \cdots, s_K\right\}$ is uniformly distributed over all size-K subsets of $\left[N\right]$. Define a state vector $\left[Q_1 \, Q_2 \, \cdots \, Q_N\right] \in \left\{0,1\right\}^N$ which indicates whether a particular signal coefficient is a large coefficient or a small coefficient: $Q_i = 1$, if $i \in \left\{s_1, s_2, \cdots, s_K\right\}$ and $Q_i = 0$, if $i \notin \left\{s_1, s_2, \cdots, s_K\right\}$. The signal of interest is defined as

$$X_i = X_{Li} \ \text{ if } Q_i = 1, \quad \text{and } X_i = X_{Si} \ \text{ if } Q_i = 0, \quad (2)$$

where X_{Li} is drawn from a distribution $p_{X_L}(\cdot; \mu_L, \sigma_L^2)$ and X_{Si} is drawn from $p_{X_S}(\cdot; \mu_S, \sigma_S^2)$, $\sigma_L > \sigma_S$. X_i's, $i = 1, \cdots, N$ are drawn independent of each other. The 2-state mixture distribution characterizing the approximately sparse signal is then given by

$$p_{X_i}(x_i) = p(Q_i = 1) p_{X_L}(x_i; \mu_L, \sigma_L^2) + p(Q_i = 0) p_{X_S}(x_i; \mu_S, \sigma_S^2).$$

Support of the above approx. sparse signal is $S = \left\{s_1, s_2, \cdots, s_K\right\}$. Note that strictly sparse signal is a special case of the above model for $\mu_S = \sigma_S = 0$.

System Models: A noiseless system model is given by $\mathbf{Y} = \mathbf{AX}$, where $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M < N$, is the measurement matrix, $\mathbf{Y} \in \mathbb{R}^{M \times L}$, $L < M$, is the collection of L measurement vectors (multiple measurement vectors), and $\mathbf{X} \in \mathbb{R}^{N \times L}$ is the collection of L vectors of unknown coefficients to be estimated. The L vectors of $\mathbf{X}$ of have a common sparsity profile (i.e., the same support). A more practical noisy system model is given by $\mathbf{Y} = \mathbf{AX} + \mathbf{N}$, where $\mathbf{N} \in \mathbb{R}^{M \times L}$ is the noise vector whose entries are i.i.d Gaussian r.v's with zero mean and variance σ_n^2. The goal is to recover signal and support from the noiseless/noisy measurements.

We note that noisy measurements of a strictly sparse signal can be modeled approximately as noiseless measurements of an approximately sparse signal. Let $\mathbf{x}$ be a strictly sparse signal and $\mathbf{x}' = \mathbf{x} + \mathbf{e}$ be an approximately sparse signal. A noiseless measurement of $\mathbf{x}'$ of the form $\mathbf{y} = \mathbf{Ax}'$ can be viewed as equivalent to the noisy measurement $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$, if $\mathbf{n}' = \mathbf{Ae}$ has the same statistical characteristics as $\mathbf{n}$. The following lemma gives the the statistical characteristics of $\mathbf{e}$ for $\mathbf{n}$ and $\mathbf{n}'$ to have the same distribution. We state the lemma for $L = 1$ (single measurement vector (SMV)). It can be easily extended to $L > 1$ (MMV).

Lemma 1 *If* $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M < N$, *with i.i.d entries from* $\mathcal{N}(0, \sigma_a^2)$, $\mathbf{n} \in \mathbb{R}^M$ *with i.i.d entries from* $\mathcal{N}(0, \sigma_n^2)$ *and* $\mathbf{e}$ *whose entries are i.i.d and distributed from* $\mathcal{N}(0, \sigma_e^2)$, *then* $\mathbf{n}' = \mathbf{Ae}$ *has the same distribution as that of* $\mathbf{n}$, *where* $\sigma_e^2 = \frac{\sigma_n^2}{N\sigma_a^2}$.

Once a measurement is made, matrix $\mathbf{A}$ is not random anymore. Thus, $\mathbf{n}'$ is not i.i.d but has $\sigma_e^2 \mathbf{I} \mathbf{A}\mathbf{A}^T$ as the covariance matrix. However, for reasonable problem dimensions, $\mathbf{A}\mathbf{A}^T$ will be close to identity. Therefore, this approximate equivalence in the distribution of $\mathbf{n}$ and $\mathbf{n}'$ will be used to characterize the priors in the recovery of strictly sparse signals from noisy measurements. This leads to the following statistical characterization of $\mathbf{x}'$:

$$p_{X_i'}(x_i') = \{p(Q_i = 1) p_{X_L}(x_i'; \mu_L, \sigma_L^2) + p(Q_i = 0)\delta(x_i')\} * \mathcal{N}(x_i'; 0, \sigma_e^2), \quad (3)$$

where $*$ is convolution operator and $\delta(.)$ is Dirac delta function.

3. BAYESIAN FRAMEWORK

Bayesian Inference Problem: Recovery of an approximately sparse signal can be stated as the optimization problem (1), which can be viewed as a regression problem. This, in turn, can be viewed as a Bayesian inference problem with the solution being equivalent to the maximum a posteriori probability (MAP) estimate of $\mathbf{x}$ as

$$\widehat{\mathbf{x}} = \arg\max_{\mathbf{x}} \underbrace{\exp\{-\|\mathbf{y} - \mathbf{Ax}\|_2^2\}}_{likelihood} \underbrace{\exp\{-\lambda E(\mathbf{x})\}}_{sparsity\ promoting\ prior}. \quad (4)$$

where λ is regularization parameter, and $E(\mathbf{x})$ is diversity measure.

Sparse Bayesian Prior: A popular choice of sparsity promoting prior is a Gaussian distribution, parametrized by (inverse) variance controlling hyperparameters [7],[6], given by

$$p(\mathbf{x}|\gamma_1, \cdots, \gamma_N) = \prod_{i=1}^{N} \left[(2\pi)^{-1/2} {\gamma_i}^{1/2} \exp\left(\frac{-\gamma_i x_i^2}{2} \right) \right], \quad (5)$$

where $\gamma_1, \cdots, \gamma_N$ are the N hyperparameters independently controlling the variance of the signal components. Here, we choose the state Q_i of the approximately sparse signal as the hyperparameter. $p(x_i|Q_i)$ is Gaussian and is given as $p(x_i|Q_i = 1) = \mathcal{N}(x_i; 0, \sigma_L^2)$ and $p(x_i|Q_i = 0) = \mathcal{N}(x_i; 0, \sigma_S^2)$. Thus, $p(x_i)$ is given as

$$p(x_i) = p(Q_i = 1)\mathcal{N}(x_i; 0, \sigma_L^2) + p(Q_i = 0)\mathcal{N}(x_i; 0, \sigma_S^2), \quad (6)$$

giving rise to a 2-state mixture Gaussian prior. The Q_i's are independent Bernoulli r.v's. To ensure that the prior $p(x_i)$ adequately represents the statistical characteristics of an approximately sparse signal with K large coefficients, we choose $p(Q_i = 1) = \alpha$ (i.e., sparsity rate K/N) if α is known, or use an estimate of α.

Signal and Support Recovery: The MAP estimate of $\mathbf{x}$ can be obtained by marginalizing out the hyperparameters from the complete posterior distribution $p(\mathbf{x}, \mathbf{q}|\mathbf{y})$, as

$$\widehat{\mathbf{x}}_{MAP} = \arg\max_{\mathbf{x}} \sum_{\mathbf{q} \in \{0,1\}^N} p(\mathbf{x}|\mathbf{y}, \mathbf{q}) p(\mathbf{q}|\mathbf{y}), \quad (7)$$

where $\mathbf{q} = [q_1\ q_2 \cdots q_N]^T \in \{0,1\}^N$ denotes the sparsity profile of $\mathbf{x}$, and the posterior distribution of the hyperparameters $p(\mathbf{q}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{q})p(\mathbf{q})$ is obtained by marginalizing out $\mathbf{x}$ as

$$p(\mathbf{q}|\mathbf{y}) \propto \left(\int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}|\mathbf{q}) d\mathbf{x} \right) p(\mathbf{q}). \quad (8)$$

The support is recovered from the posterior distribution of Q_i's as

$$\widehat{S} = \{i : p(Q_i = 1|\mathbf{y}) \geq p(Q_i = 0|\mathbf{y})\}. \quad (9)$$

We see that in the above Bayesian framework, support recovery is an integral part of signal recovery. In MMV settings [4], a more accurate recovery of support can be made compared to that in SMV settings because of the availability of multiple measurements of signal vectors with the same sparsity profile, which can lead to improved signal recovery.

4. PROPOSED JSSR-MP ALGORITHM

Notation of Indices: The letters u, v, w denote indices in $[M] \equiv \{1, \cdots, M\}$, letters i, j, k denote indices in $[N] \equiv \{1, \cdots, N\}$, and letters l, m, n denote indices in $[L] \equiv \{1, \cdots, L\}$.

The factor graph representing the joint distribution of the signal, hyperparameters, and observed measurements for our problem in MMV setting is shown in Fig. 1. We have two types of variable nodes; *hyperparameter variable nodes (HVN)* representing q_i, $i = 1, \cdots N$, and *signal variable nodes (SVN)* representing signal components x_{il}, $i = 1, \cdots N, l = 1, \cdots, L$. We have two types of constraint nodes; one gives the relationship between signal components and observed measurements (*measurement factor nodes (MFN)*), and the other connects HVNs & SVNs and determines the conditional distribution of signal given the hyperparameter; call these nodes, w_{il}'s, as *prior-mixing factor nodes (PFN)*.

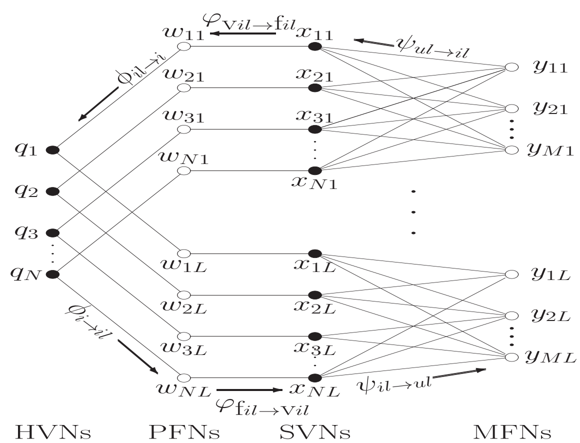

Fig. 1. Factor graph for joint support and signal recovery.

Message Passing Algorithm: Belief propagation (BP) is known to give the exact marginal when the factor graph is a tree. The factor graph in our problem (Fig. 1) is loopy. We apply BP on this factor graph treating it as a loop-free graph, and compute approximate marginal posterior distributions of the signal components (x_{il}'s) and hyperparameters (q_i's), conditioned on the observed measurements

Message	From	To	Nature of message	Parameters
$\psi_{ul\to il}$	MFN	SVN	Gaussian Density	$\psi^t_{ul\to il} : \mathcal{N}(\cdot; \mu^t_{ul\to il}, \upsilon^t_{ul\to il})$, where
				$\mu^t_{ul\to il} = \frac{1}{a_{ui}}\{y_{ul} - \sum_{j\neq i} a_{uj}\mu^{t-1}_{jl\to ul}\},\ \upsilon^t_{ul\to il} = \frac{1}{a^2_{ui}} \sum_{j\neq i} a^2_{uj}\upsilon^{t-1}_{jl\to ul}$
ψ_{il}	SVN	MFN	Gaussian Density	$\psi^t_{il} : \mathcal{N}(\cdot; \mu^t_{il}, \upsilon^t_{il})$, where
			(Broadcast Message)	$\upsilon^t_{il} = \left(\frac{1}{(\tilde{\sigma}^t_{il})^2} + \frac{1}{\upsilon^t_{\mathrm{v}il\to \mathrm{f}il}}\right)^{-1},\ \mu^t_{il} = \left(\frac{\tilde{\mu}^t_{il}}{(\tilde{\sigma}^t_{il})^2} + \frac{\mu^t_{\mathrm{v}il\to \mathrm{f}il}}{\upsilon^t_{\mathrm{v}il\to \mathrm{f}il}}\right)\upsilon^t_{il}$
$\psi_{il\to ul}$	SVN	MFN	Gaussian Density	$\psi^{t-1}_{il\to ul} : \mathcal{N}(\cdot; \mu^{t-1}_{il\to ul}, \upsilon^{t-1}_{il\to ul})$, where
			(Extracted Message)	$\mu^{t-1}_{il\to ul} = \left(\frac{\mu^{t-1}_{il}}{\upsilon^{t-1}_{il}} - \frac{\mu^{t-1}_{ul\to il}}{\upsilon^{t-1}_{ul\to il}}\right)\upsilon^{t-1}_{il\to ul},\ \upsilon^{t-1}_{il\to ul} = \left(\frac{1}{\upsilon^{t-1}_{il}} - \frac{1}{\upsilon^{t-1}_{ul\to il}}\right)^{-1}$
$\varphi_{\mathrm{v}il\to \mathrm{f}il}$	SVN	PFN	Gaussian Density	$\varphi^t_{\mathrm{v}il\to \mathrm{f}il} : \mathcal{N}(\cdot; \mu^t_{\mathrm{v}il\to \mathrm{f}il}, \upsilon^t_{\mathrm{v}il\to \mathrm{f}il})$, where
				$\upsilon^t_{\mathrm{v}il\to \mathrm{f}il} = \left(\sum_{u=1}^{M} \frac{1}{\upsilon^t_{ul\to il}}\right)^{-1},\ \mu^t_{\mathrm{v}il\to \mathrm{f}il} = \left(\sum_{u=1}^{M} \frac{\mu^t_{ul\to il}}{\upsilon^t_{ul\to il}}\right)\upsilon^t_{\mathrm{v}il\to \mathrm{f}il}$
$\varphi_{\mathrm{f}il\to \mathrm{v}il}$	PFN	SVN	Gaussian Mixture	$\varphi^t_{\mathrm{f}il\to \mathrm{v}il} = \zeta^t_i\,\mathcal{N}(\cdot; 0, \sigma^2_L) + (1-\zeta^t_i)\mathcal{N}(\cdot; 0, \sigma^2_S)$
$\phi_{il\to i}$	PFN	HVN	LLR	Λ^t_{il}
$\phi_{i\to il}$	HVN	PFN	ζ_i	$\zeta^t_i, \quad \zeta^0_i = \widehat{\alpha}$, where $\widehat{\alpha} = \frac{1}{(\sigma^2_L - \sigma^2_S)}\left(\frac{1}{MLN\sigma^2_a}\lVert\mathbf{Y}\rVert^2_F - \sigma^2_S\right)$

Table 1. Messages and their parameters. $\mathcal{N}(\cdot; a, b)$ denotes Gaussian distribution with mean a and variance b.

($\mathbf{Y}$). This results in computational efficiency as well as good performance. Treating the graph as loop-free, the marginal posterior distributions of x_{il}'s and q_i's are given by

$$p(x_{il}|\mathbf{Y}) \propto \prod_{u=1}^{M} p(y_{ul}|x_{il}) \cdot [\zeta_i\,\mathcal{N}(x_{il}; 0, \sigma^2_L) + (1-\zeta_i)\mathcal{N}(x_{il}; 0, \sigma^2_S)], \tag{10}$$

where $\zeta_i = p(Q_i = 1|\mathbf{Y})$, and

$$p(q_i|\mathbf{Y}) \propto \prod_{l=1}^{L} \int_{x_{il}} \prod_{u=1}^{M} p(y_{ul}|x_{il}) p(x_{il}|q_i)\, dx_{il}\, p(q_i). \tag{11}$$

The various messages and their nature/parameters are summarized in Table I. The computations carried out and messages passed in each iteration of the algorithm are described below.

Message Passing: In each iteration, HVNs compute the approximate posterior distribution $p(Q_i|\mathbf{Y})$ and pass $\phi_{i\to il} := (\zeta_i)$, where $\zeta_i = p(Q_i = 1|\mathbf{Y})$ to the PFNs. At the PFNs, the constraint function $p(x_{il}|Q_i)$ is marginalized w.r.t q_i to obtain $p(x_{il})$, which is a mixture of two Gaussians with ζ_i governing the ratio in which the two distributions are mixed. The PFNs pass the parameters of this mixture to the SVNs, i.e., $\varphi_{\mathrm{f}il\to \mathrm{v}il} := (\zeta_i, \sigma^2_L, \sigma^2_S)$. At the SVNs, the Gaussian mixture is approximated by a single Gaussian density $\mathcal{N}(.; 0, (\tilde{\sigma}_{il})^2)$, where $(\tilde{\sigma}_{il})^2 = \zeta_i\sigma^2_L + (1-\zeta_i)\sigma^2_S$. The messages to MFNs from SVNs (i.e., $\psi_{il\to ul}$'s) are computed by taking the product of messages coming to SVNs from MFNs ($\psi_{ul\to il}$) and this single Gaussian density. The approximation of the mixture to single Gaussian at the SVNs facilitates the use Gaussian BP [8] and the associated computationally efficient broadcast messaging strategy. Accordingly, the SVNs broadcast the messages ψ_{il}.

At the MFNs, relevant messages are extracted from these broadcast messages (which are Gaussian densities), which are then multiplied with the constraint function (defined by $\delta_{\{y_{ul} = \mathbf{a}_{u\cdot}\mathbf{x}_{\cdot l}\}}$, which denotes a Dirac distribution on the hyperplane $y_{ul} = \mathbf{a}_{u\cdot}\mathbf{x}_{\cdot l}$), and marginalized to obtain the new messages $\psi_{ul\to il}$ to be sent to SVNs. The SVNs take the product of these messages $\psi_{ul\to il}$ coming from MFNs, and send the parameters of the resultant Gaussian density to the PFNs via $\varphi_{\mathrm{v}il\to \mathrm{f}il}$ messages. The PFNs take the product of the $\varphi_{\mathrm{v}il\to \mathrm{f}il}$ messages coming from SVNs and the constraint function $p(x_{il}|Q_i)$, and marginalize with respect to x_{il} to obtain $\int_{x_{il}} \varphi_{\mathrm{v}il\to \mathrm{f}il}\, p(x_{il}|Q_i)\, dx_{il}$, and compute the LLRs

$$\Lambda_{il} = \log\left\{\frac{\int_{x_{il}} \varphi_{\mathrm{v}il\to \mathrm{f}il} p(x_{il}|Q_i = 1) dx_{il}}{\int_{x_{il}} \varphi_{\mathrm{v}il\to \mathrm{f}il} p(x_{il}|Q_i = 0) dx_{il}}\right\}. \tag{12}$$

PFNs send these LLRs as $\phi_{il\to i}$ messages to HVNs. At the ith HVN, approximate posterior distribution $p(Q_i|\mathbf{Y})$ is computed by taking the sum of all the incoming LLRs along with the LLR of the prior density $\Lambda_i = \sum_{l=1}^{L} \Lambda_{il} + \Lambda^0_i$, where $\Lambda^0_i = \log\left(\frac{p(Q_i=1)}{p(Q_i=0)}\right)$ is the prior. Messages are exchanged for a certain number of iterations. The marginals of the hyperparameters at the end are used to recover the support $\widehat{S}$ (and also the sparsity profile $\widehat{\mathbf{q}}$) as per (9). The algorithm is then re-initialized with the recovered sparsity profile $\widehat{\mathbf{q}}$; i.e., the priors of the signal components are re-initialized as: $p(x_{il}) = \mathcal{N}(\cdot; 0, \sigma^2_L)$ if $\widehat{q}_i = 1$, and $p(x_{il}) = \mathcal{N}(\cdot; 0, \sigma^2_S)$ if $\widehat{q}_i = 0$). Messages are exchanged between SVNs and MFNs for several iterations, after which the marginal posterior distribution of signal components are computed, which are used to recover $\widehat{\mathbf{X}}$.

To improve the convergence behavior of the algorithm, we have applied damping [9] to the messages, where, in each iteration, a message is computed as a weighted average of the old message (of the previous iteration) and the newly evaluated message (of the current iteration). $\beta_m, \beta_q, \beta_s$ are the damping factors for messages ψ, ϕ and signal $\widehat{x}_{il}$'s, respectively. For recovery of strictly sparse signals from noisy measurements, σ_L and σ_S values to be used in the recovery algorithm have to be updated according to *Lemma 1* using σ_n, and, in addition, in the final step all those signal components whose $\widehat{q}_i$ turn out to be zero are set to zero.

Complexity: The proposed JSSR-MP algorithm has a complexity of order $\mathcal{O}(MNL)$. Since $L << M$ typically, the complexity order is $\mathcal{O}(MN)$. This complexity is one order less in M compared to that of M-SBL algorithm in [6], whose complexity is $\mathcal{O}(M^2N + M^2L)$. Thus, in the proposed algorithm, the approximation of the Gaussian mixture by a single Gaussian density at the SVNs, the consequent use of Gaussian BP, and the associated broadcast messaging strategy has led to a computationally efficient solution whose complexity is linear in M and N.

5. SIMULATION RESULTS

Experiment 1: In this experiment, we compare the mean square error (MSE) performance of the proposed JSSR-MP algorithm with those of the AMP algorithm in [5] and the CoSaMP algorithm in [11] for the case of noiseless measurements of *approximately sparse signals* with $L = 1$ (i.e., SMV). The MSE results as a function of M are shown in Fig. 2 for $N = 500$, $\sigma_L = 10$, $\sigma_S = 1$, $\sigma_n = 0$, and $K = 50, 100$. The damping factors used are $\beta_m = 0.5$, $\beta_q = 0.3$, $\beta_s = 0.3$. In addition to the MSE plots for JSSR-MP, AMP and

CoSaMP algorithms, we also show the plot for a genie-aided scheme which is nothing but the JSSR-MP algorithm with perfect knowledge of support. In the actual JSSR-MP, however, the support is estimated and hence the estimated support can be imperfect. The genie-aided scheme, therefore, is an indicator of the best performance possible with JSSR-MP. The following observations can be made from Fig. 2. As expected, as the sparsity rate (K/N) is increased, the MSE degrades. The MSE improves as M is increased, which is also expected. The JSSR-MP algorithm is found to significantly outperform CoSaMP and to perform almost the same as or better than the AMP (note that AMP does not jointly recover the support and the signal). The order of complexity in CoSaMP, AMP and JSSR-MP algorithms are the same. Comparing with the performance of the genie-aided scheme, it is seen that significant improvement in performance is possible if the quality of support recovery is improved. This is indeed achieved by JSSR-MP in MMV $(L > 1)$ as illustrated in the Fig. 2 for $L = 5$, which is quite close to the performance achieved by the genie-aided scheme.

Experiment 2: In this experiment, we recover *strictly sparse signals* from noisy measurements. MSE as well as percentage success (PS) of support recovery results are obtained for JSSR-MP algorithm and M-SBL algorithm in [6]. Figures 3 and 4 show the MSE performance of signal recovery and PS performance of support recovery, respectively, for $N = 500$, $K = 160$, SNR = 20 dB, $\sigma_L = 10$, and $\sigma_S = 0$. MSE plots as a function of M for $L = 5, 10, 25$ are shown in Fig. 3. PS plots as a function of L for $M = 200, 250$ are shown in Fig. 4. σ_n is assumed to be known for recovery in both JSSR-MP and M-SBL algorithms. From Figs. 3 and 4, we observe that the JSSR-MP performance is comparable to that of M-SBL. This is interesting given that this comparable performance is achieved by JSSR-MP at one order less complexity than M-SBL; $\mathcal{O}(MN)$ complexity of JSSR-MP versus $\mathcal{O}(M^2N)$ complexity of M-SBL.

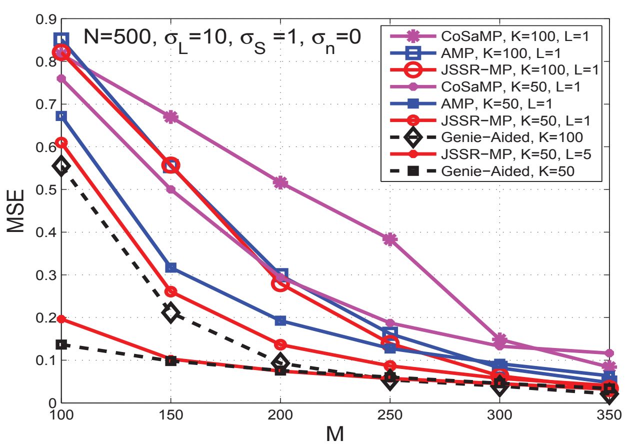

Fig. 2. MSE versus M performance of signal recovery of approximately sparse signals with noiseless measurements for SMV $(L = 1)$ and MMV $(L = 5)$

6. CONCLUSIONS

We developed a low-complexity message passing algorithm which jointly recovers the support and signal of approximately sparse signals with multiple measurement vectors. The algorithm achieved good performance at low complexities due to the approximation of the Gaussian mixture prior by a single Gaussian and message damping. By exploiting the approximate equivalence between noiseless measurements of approximately sparse signals and noisy measurements of strictly sparse signals, the proposed JSSR-MP algorithm is applicable to a wider class of signals and measurement models.

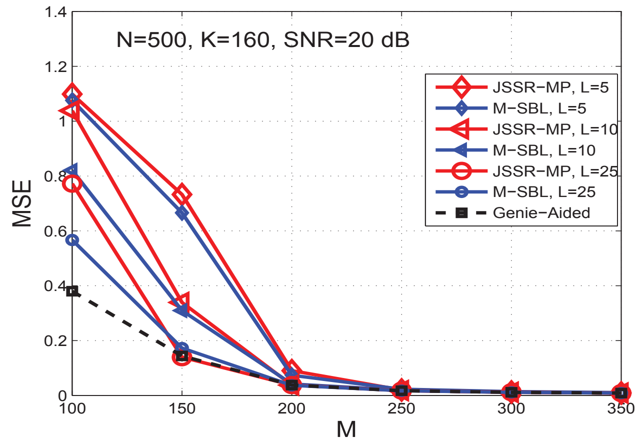

Fig. 3. MSE versus M performance of signal recovery of strictly sparse signals with noisy measurements for MMV $(L = 5, 10, 25)$.

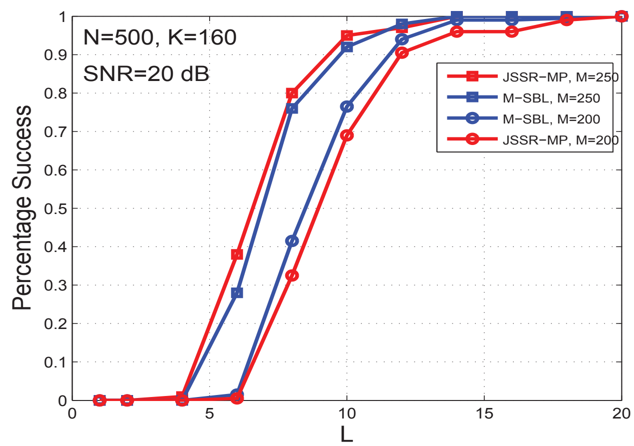

Fig. 4. % success versus L performance of support recovery of strictly sparse signals with noisy measurements for MMV.

7. REFERENCES

[1] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing of approximately sparse signals," *Proc. IEEE ISIT 2008*, Toronto, Jul. 2008.

[2] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressing sensing via belief propagation," *IEEE Trans. Sig. Proc.*, Jan. 2010.

[3] E. Candes, "Compressive sampling," *Proc. Intl. Congress of Mathematics*, vol. 3, pp. 1433-1452, Madrid, 2006.

[4] S. F. Cotter and B. D. Rao, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Sig. Proc.*, vol. 53, no. 7, pp. 2477-2488, Jul. 2005.

[5] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *PNAS*, Sept. 2009.

[6] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Sig. Proc.*, vol. 55, no. 7, pp. 3704-3716, Jul. 2007.

[7] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Jl. of Mach. Learn. Res.*, vol. 1, pp. 211-244, Sept. 2001.

[8] O. Shental, D. Bickson, P. H. Siegel, J. K. Wolf, and D. Dolev, "Gaussian belief propagation solver for systems of linear equations," *Proc. IEEE ISIT'2008*, Toronto, Jul. 2008.

[9] M. Pretti, "A message passing algorithm with damping," *Jl. Stat. Mech.: Theory and Practice*, Nov. 2005.

[10] D. Needell, J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Jl. of Appl. Comput. Harmon. Anal*, vol. 26, pp. 301321, 2009.

[11] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing,* vol. 56, no. 6, pp. 2346-2356, June 2008.