Approximate Message Passing for Recovery of Sparse Signals with Markov-Random-Field Support Structure

Subhojit Som SUBHOJIT@GATECH.EDU
School of Electrical and Computer Engineering,
Georgia Institute of Technology, Atlanta, GA 30332 USA

Philip Schniter SCHNITER@ECE.OSU.EDU
Department of Electrical and Computer Engineering,
The Ohio State University, Columbus, OH 43210 USA

The main objective in sparse reconstruction or compressive sensing is to estimate a signal $\boldsymbol{x} \in \mathbb{R}^N$ from M noisy linear observations $\boldsymbol{y} \in \mathbb{R}^M$,

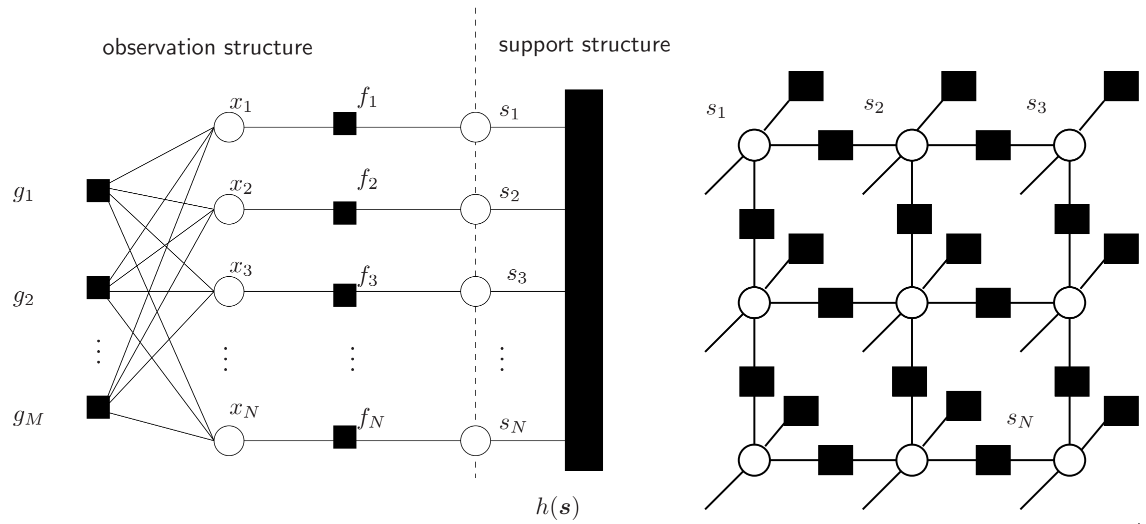
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}, \tag{1}$$

In (1), $\boldsymbol{x} \in \mathbb{R}^N$ has only K non-zero coefficients, $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is a known measurement matrix, and $\boldsymbol{e} \in \mathbb{R}^M$ is additive noise, often modeled as white and Gaussian, i.e., $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma_e^2\boldsymbol{I})$. In many problems of interest, the linear measurement system (1) is underdetermined i.e., $M < N$. Even so, when $K \leq M$ and the columns of the measurement matrix $\boldsymbol{A}$ are incoherent, it is possible to reliably recover $\boldsymbol{x}$ from these small number of observations $\boldsymbol{y}$ (see, e.g., the references in (Baraniuk et al., 2010)).

Three major classes of algorithms have been proposed for compressive sensing with provable performance guarantees: convex optimization, greedy search and iterative thresholding (Baraniuk et al., 2010). The iterative thresholding approach requires only two matrix-vector multiplications per iteration and a few vector operations; hence it is very fast. Moreover, these matrix multiplies can often be implemented using a fast transform like fast Fourier transform (FFT). A version of iterative thresholding, known as *approximate message passing* (AMP), has overcome (Donoho et al., 2009) the performance deficiencies (Maleki & Donoho, 2010) of earlier algorithms and achieves the same *sparsity-undersampling trade-off* (Donoho & Tanner, 2009) that characterizes the much more expensive convex-optimization algorithms. The AMP algorithm can be understood as a simplified version of Bayesian *message passing* (MP) on a dense loopy factor graph that exploits the central limit theorem to approximate messages as Gaussian. Remarkably, in the large system limit (i.e., $M, N \to \infty$ with $M/N \to \delta$ and $\delta \in (0, 1]$), the analysis in (Bayati & Montanari, 2011) suggests that AMP inference generates exact posteriors

The MP formulation not only gives very efficient compressive sensing algorithms, but also gives a framework to incorporate prior knowledge of signal structure beyond simple sparsity. One such possibility is structure in signal support. For example, natural images are not only sparse in the wavelet domain, but also exhibit *persistence across scales* (Crouse et al., 1998), which makes certain support patterns much more likely. Likewise, in radar images, spatial pixel supports show *clustering* perpendicular to the look direction (Jakowatz, Jr. et al., 1996). Exploitation of signal structure beyond simple sparsity helps to reduce the number of observations needed for reliable reconstruction (Baraniuk et al., 2010; Schniter, 2010).

In this work we address the reconstruction of sparse signals with support structure showing spatial clustering. We adopt the *turbo reconstruction* approach (Schniter, 2010) which exploits the support structure of sparse signals in a probabilistic framework. It can be interpreted as an extension of the AMP principle from (Donoho et al., 2010) to the expanded factor graph depicted in Fig. 1(a). The MP on this extended factor graph is done in an iterative fashion. Beliefs on sparsity pattern (i.e., support elements) are exchanged between two soft inference blocks – one exploiting the linear observation model and the other exploiting the hidden Markov structure of the support. At every iteration the belief propagation within the observation block is done using the AMP framework. Support structures in the form of Markov chain (Schniter, 2010) and Markov tree (Som et al., 2010) have been studied earlier. Since the graphical models for these priors do not have any loops, a single pass of *forward-backward* algorithm computes exact belief on them. In this work



(a) Factor graph for the structured sparse signal reconstruction problem. The function nodes are shown by filled rectangles and variable nodes are shown by empty circles.

(b) Factor graph representation of $h(\boldsymbol{s}) \triangleq p(\boldsymbol{s})$ modeled as an Ising MRF.

Figure 1. Factor graph representation.

we consider a Markov random fields (MRF) (Li, 2009) prior which models spatial clustering and whose graphical model has loops. Hence computing exact belief is intractable. However we use loopy belief propagation on its factor graph and demonstrate that we can get significant performance gain in terms of *phase transition* behavior by exploiting the support structure. We also propose a method to learn the model parameters from the data. After every turo iteration, model parameters are learnt and updated. The signal and noise variances are learnt using maximum likelihood method. But maximizing the likelihood function for the Ising model is impractical. Instead, we maximize the pseudo-likelihood (Besag, 1977; 1986) function which approximates the likelihood function as a product of conditional probability terms.

Now we give the signal model in detail. We model the support elements of the signal coefficients using hidden binary indicators $\{s\}_{n=1}^{N}$, where $s_n \in \{-1, 1\}$. Here, $s_n = 1$ indicates $x_n \neq 0$ w.p. 1 while $s_n = -1$ indicates $x_n = 0$ w.p. 1. For the binary vector $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]^T \in \{-1, 1\}^N$, we impose an Ising prior which models interaction between coefficients of the vector $\boldsymbol{s}$ and also controls the sparsity of the signal. The Ising prior is given by

$$p(\boldsymbol{s}) = \frac{1}{Z} \exp\left(\sum_n s_n \left(\frac{1}{2} \sum_{s_m \in S_n^c} \beta_{mn} s_m - \alpha_n \right) \right), \quad (2)$$

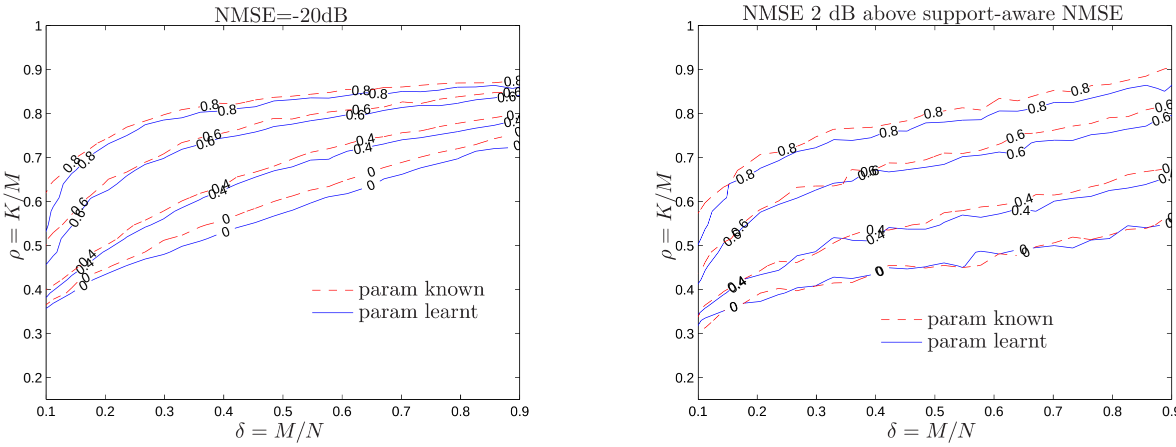
where S_n^c is the set of all neighboring variables s_m of the node s_n. Here β_{mn}, α_n are parameters of the distribution and Z is the normalization factor (function of β_{mn}, α_n) known as partition function. This probability model is depicted by the factor graph presented in Fig. 1(b) for 2D planar grid with 4-connected neighborhood structure. There are two types of function nodes: nodes connecting two neighboring variable nodes characterized by the interaction parameter β and function nodes connected to a single variable node characterized by the bias parameter α. Higher magnitude of β means stronger spatial coherence and higher value of α enforces sparser signal activity.

Conditioned on $\boldsymbol{s}$, the signal coefficients are independent and follow the Bernoulli-Gaussian prior of the form

$$p_n(x_n|s_n) = \delta(s_n - 1)\mathcal{N}(x_n; 0, \sigma_n^2) + \delta(s_n + 1)\delta(x_n), \quad (3)$$

where $\delta(.)$ denotes the Dirac delta and $\mathcal{N}(.)$ is a Gaussian pdf; σ_n^2 is the variance of the coefficient x_n given $s_n = +1$. The factor graph for this signal model is shown in Fig. 1(a).

We now present preliminary results of a numerical experiment studying the performance of the algorithm for various values of undersampling ($\delta = M/N$) and sparsity ($\rho = K/M$), where K is the number of active coefficients in the signal $\boldsymbol{x}$. Fig. 2 shows a number of empirically calculated phase transition curves on the (δ, ρ) plane for numerical simulations with 30dB SNR. Each curve partitions the sparsity-undersampling plane into "success" and "failure" re-



(a) For any point (δ, ρ) above the phase transition curve the average NMSE is more than -20dB and for any point below it the average NMSE is less than -20dB. The dotted lines correspond to the case when the true parameters of the model were all known. The solid lines are for the case when they were learnt from the data.

(b) For any point (δ, ρ) above the phase transition curve the average degradation of NMSE over support-aware NMSE is more than 2dB and for any point below it the degradation is less than 2dB. The dotted lines correspond to the case when the true parameters of the model were all known. The solid lines are for the case when they were learnt from the data.

Figure 2. Empirical phase transition curves for robust turbo reconstruction algorithm for different values of β which are shown as labels on the curves. The

gions, defined as follows. In Fig. 2(a), for any (δ, ρ) that lies below the curve, the average normalized reconstruction MSE (NMSE) from 200 trials was less than -20dB and for points above the curve, it was more than -20dB. Similarly in Fig. 2(b), the average NMSE degradation over support-aware NMSE was less than 2dB below the curve and more than 2 dB above the curve. Each curve in Fig. 2 is labelled with the value of the Ising model's β that was used to generate the signal. We assumed α and β to be uniform throughout the 2D-planar grid with 4-connected neighborhood structure. A higher value of β means more clustering, i.e., more signal structure. Note that, when $\beta = 0$, the s_n's are independent, i.e., the signal is simply sparse. We can see from Fig. 2 that, as β increases, "successful" sparse reconstruction is achieved over a larger portion of the sparsity-undersampling plane. We also note that at higher β regime, the gain achieved by exploiting the MRF structure of the support is significantly more than the performance drop due to not knowing the true model parameters shown by the difference between the solid and dotted lines.

Acknowledgments

This work was supported in part by NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090.

References

Baraniuk, R. G., Cevher, V., Duarte, M. F., and Hegde, C. Model-based compressive sensing. *IEEE Trans. Inform. Theory*, 56(4):1982–2001, Apr. 2010.

Bayati, M. and Montanari, A. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Information Theory*, 2011.

Besag, J. Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrica*, 64(3):616–618, 1977.

Besag, J. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48(3): 259–302, 1986.

Crouse, Matthew S., Nowak, Robert D., and Baraniuk, Richard G. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. on Signal Processing*, 46(4):886–902, 1998.

Donoho, D. L. and Tanner, J. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Phil. Trans. Royal Soc. A*, 367 (1906):4273–4293, 2009.

Donoho, D. L., Maleki, A., and Montanari, A. Message passing algorithms for compressed sensing.

Proc. National Academy of Sciences, 106(45):18914–18919, Nov. 2009.

Donoho, D. L., Maleki, A., and Montanari, A. Message passing algorithms for compressed sensing: I. Motivation and construction. In *Proc. Inform. Theory Workshop*, Cairo, Egypt, Jan. 2010.

Jakowatz, Jr., C. V., Wahl, D. E., Eichel, P. H., Ghiglia, D. C., and Thompson, P. A. *Spotlightmode Synthetic Aperture Radar: A Signal Processing Approach*. Springer, New York, 1996.

Li, S. Z. *Markov Random Field Modeling in Image Analysis*. Springer, London, 3rd edition, 2009.

Maleki, A. and Donoho, D. L. Optimally tuned iterative reconstruction algorithms for compressed sensing. *IEEE J. Sel. Topics Signal Process.*, 4(2):330–341, Apr. 2010.

Schniter, P. Turbo reconstruction of structured sparse signals. In *Proc. Conf. Inform. Science & Syst.*, Princeton, NJ, Mar. 2010.

Som, S., Potter, L. C., and Schniter, P. Compressive imaging using approximate message passing and a Markov-tree prior. In *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 2010.