

Approximate Message Passing for Bilinear Models

Philip Schniter

Department of Electrical and
Computer Engineering
The Ohio State University
Email: schniter@ece.osu.edu

Volkan Cevher

Laboratory for Information and Inference Systems
Idiap Research Institute
Ecole Polytechnique Federale de Lausanne
Email: volkan.cevher@epfl.ch

I. EXTENDED SUMMARY

Problem: We consider the following bilinear model in the unknowns $\mathbf{X} \in \mathbb{R}^{N \times L}$ and $\Phi \in \mathbb{R}^{M \times N}$, which has applications in dictionary learning, matrix completion, collaborative filtering, compressive system calibration, compressive sensing with dictionary uncertainty, and Bayesian experimental design:

$$\mathbf{Y} = \mathcal{P}(\Phi \mathbf{X}) + \mathbf{W}. \quad (1)$$

In (1), \mathbf{Y} are known observations, $\mathcal{P}(\cdot)$ accomplishes element-wise selection or linear projection, and \mathbf{W} models additive perturbation. Please see [1] for further details.

Approach: We take a Bayesian approach to the inference problems (in particular, posterior estimation) that revolve around the bilinear model (1). In particular, we leverage the approximate message passing (AMP) framework of [2], [3] and extend it to the bilinear domain. Compared to Bayesian approaches that rely on Gibbs sampling methods or variational inference, the AMP framework allows us to fully exploit the *blessings-of-dimensionality* (e.g., the asymptotic normality and concentration-of-measures) to achieve salient advantages in computation and estimation accuracy. Our “turbo AMP” framework also allows us to characterize the impact of our message scheduling using extrinsic information transfer (EXIT) charts, originally developed to predict the convergence of turbo decoding.

Example Application: For concreteness, we describe the application of the bilinear model (1) to the *compressive system calibration problem*. Based on the theoretical premise of compressive sensing, a great deal of research has revolved around the design of sampling systems, such as Analog-to-Information receivers and Xampling. The sampling matrices in these systems are pre-designed with certain desired theoretical properties to guarantee recovery along with the constraints of hardware implementations. However, when implementing the mathematical “sampling” operation—here defined by the matrix Φ —in real hardware, one often introduces what are effectively perturbations on Φ that create an undesired gap between theoretical and practical system performance. As a means of closing this gap, we are interested in jointly learning the true matrix Φ while simultaneously recovering the signal \mathbf{X} .

Suppose, then, that our compressive sensing system produces a sequence of vector observations \mathbf{y}_l ($l = 1, \dots, L$; collectively referred as \mathbf{Y}) that correspond to a sequence of unknown sparse signals \mathbf{x}_l (collectively, \mathbf{X}). We assume that the signal coefficients $\{x_{jl}\}$ are drawn i.i.d from a (known) compressible prior $x_{jl} \sim p_X(\cdot)$, and we model the entries of the true (unknown) sampling matrix Φ as i.i.d Gaussian with variance μ^w and known mean $\bar{\Phi}$. For ease of description, we assume that μ^w is known, that the signals \mathbf{x}_l are canonically sparse, and that the projection operator $\mathcal{P}(\cdot)$ is identity. This calibration problem yields the factor graph in Fig. 1, to which

we apply *bilinear AMP* in order to generate (approximate) posterior marginals on the elements of Φ and \mathbf{X} .

This calibration problem can be interpreted as an instance of dictionary learning, whereby one seeks a sparsifying dictionary for some training data. In this setting, it is known that ℓ_1 -norm minimization can locally identify the correct dictionary (i.e., Φ) given $L = \mathcal{O}(N^3 K)$ training samples, where K is the “sparsity” of \mathbf{x}_l [4]. We note, however, that the computational complexity of this approach is extremely demanding for large scale problems.

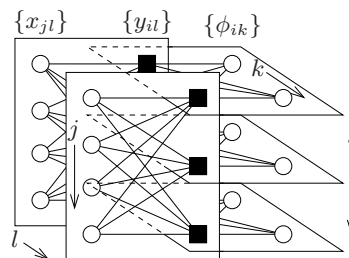


Fig. 1. An illustration of the factor graph for our message passing solution.

Preliminary Results: Figure 2 shows example results for the application of bilinear AMP to the calibration problem. The non-convexity of the problem is quite apparent from the plots. Here, to generate the signals, we used an i.i.d Bernoulli-Gaussian prior that generated zero-mean unit-variance active coefficients with probability K/N . The nominal sampling matrix $\bar{\Phi}$ was generated i.i.d Gaussian with zero mean and $1/M$ -variance, and true Φ was generated by perturbing $\bar{\Phi}$ with an additive noise of the same distribution.

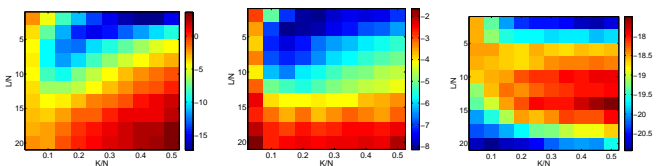


Fig. 2. Recovery errors in dB: the dictionary, the signals, and the data.

REFERENCES

- [1] L. Carin, R. Baraniuk, V. Cevher, D. Dunson, M. Jordan, G. Sapiro, and M. Wakin, “A Bayesian Approach to Learning Low-Dimensional Signal Models from Incomplete Measurements,” *IEEE SP Mag*, 2010.
- [2] D. L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing,” *Proc. National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
- [3] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” *arXiv:1010.5141*, Oct. 2010.
- [4] Q. Geng, H. Wang, and J. Wright, “On the Local Correctness of L^1 Minimization for Dictionary Learning,” *arXiv:1101.5672*, 2011.