Performance Analysis of Approximate Message Passing for Distributed Compressed Sensing

Gabor Hannak[1], Alessandro Perelli[2], Gerald Matz[1], Mike E. Davies[2], Norbert Goertz[1].

[*1]Institute of Telecommunications, Vienna University of Technology (e-mail: firstname.lastname@nt.tuwien.ac.at).

[2]Institute of Digital Communications, Edinburgh University, (e-mail: firstname.lastname@ed.ac.uk).

***Abstract*— Bayesian approximate message passing (BAMP) is an efficient method in compressed sensing that is nearly optimal in the minimum mean squared error (MMSE) sense. Multiple measurement vector (MMV)-BAMP performs joint recovery of multiple vectors with identical support and accounts for correlations in the signal of interest and in the noise. In this paper, we show how to reduce the complexity of vector BAMP via a simple joint decorrelation (diagonalization) transform of the signal and noise vectors, which also facilitates the subsequent performance analysis. We prove that the corresponding state evolution (SE) is equivariant with respect to the joint decorrelation transform and preserves diagonality of the residual noise covariance for the Bernoulli-Gauss (BG) prior. We use these results to analyze the dynamics and the mean squared error (MSE) performance of BAMP via the replica method, and thereby understand the impact of signal correlation and number of jointly sparse signals.**

Algorithm 1 BAMP for MMV/DCS

input: $\mathbf{y}(b)$, $\mathbf{A}(b)$, $\boldsymbol{\Sigma}_{\vec{\mathbf{x}}}$, ε, $t_{\max}$, ε_{tol}

$t = 0$, $\hat{\vec{\mathbf{x}}}_n^t = \mathbf{0}_{B\times 1}$, $\vec{\mathbf{r}}_m^t = \vec{\mathbf{y}}_m, \forall m, n$

do

$\quad t \leftarrow t + 1$

$\quad \mathbf{u}^{t-1}(b) = \hat{\mathbf{x}}^{t-1}(b) + \mathbf{A}(b)^T \mathbf{r}^{t-1}(b), \forall b$

$\quad \boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^{t-1} = \begin{cases} \boldsymbol{\Sigma}_{\vec{r}_m}^{t-1} & \text{for MMV} \\ D\big(\boldsymbol{\Sigma}_{\vec{r}_m}^{t-1}\big) & \text{for DCS} \end{cases}$

$\quad \hat{\vec{\mathbf{x}}}_n^t = F(\vec{\mathbf{u}}_n^{t-1}; \boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^{t-1}), \forall n$

$\quad \vec{\mathbf{r}}_m^t = \vec{\mathbf{y}}_m - \big(\mathbf{A}(1)\hat{\mathbf{x}}^t(1), \ldots, \mathbf{A}(B)\hat{\mathbf{x}}^t(B)\big)_m + \frac{1}{M}\sum_{n=1}^N F'(\vec{\mathbf{u}}_n^{t-1}; \boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^{t-1})\vec{\mathbf{r}}_m^t, \forall m$

while $\sum_{b=1}^B \left\|\hat{\mathbf{x}}^t(b) - \hat{\mathbf{x}}^{t-1}(b)\right\|_2^2 > \varepsilon_{\text{tol}} \sum_{b=1}^B \left\|\hat{\mathbf{x}}^{t-1}(b)\right\|_2^2$ **and** $t < t_{\max}$

return $\hat{\mathbf{x}}(b) = \hat{\mathbf{x}}^t(b), \forall b$

1 Introduction

Compressed sensing (CS) is a signal processing technique aiming at recovering a high-dimensional sparse vector from a (noisy) system of linear equations [2]. Joint sparsity refers to multiple vectors having the same support set [3, 4]. There are two prominent CS scenarios in the context of joint sparsity: the multiple measurement vector (MMV) (identical measurement matrices) and the distributed compressed sensing (DCS) problem (mutually independent measurement matrices). Joint sparsity arises in a number of real-world scenarios, e.g., when multiple sensors or antennas observe the same signal corrupted by different channels and noise.

Several methods for jointly sparse recovery have been proposed in the literature. In this paper we investigate the vector BAMP [5] because (i) it directly incorporates the vector signal pdf and is nearly MMSE optimal, (ii) it allows for theoretical performance analysis called SE. In [6], the replica method is used to calculate the MMSE of the CS measurement, given the BG signal prior and uncorrelated isotropic unitary signal and uncorrelated isotropic Gaussian noise distribution, i.e., with a single noise parameter.

1.1 Contributions

We provide an analytical performance prediction for the BAMP algorithm with a BG signal prior with arbitrary signal and noise correlation by (i) incorporating a linear joint decorrelation of the measurements, (ii) showing the equivariance of Bayesian approximate message passing (BAMP) w.r.t. invertible linear transformations, (iii) extending the replica analysis from [6] to arbitrary diagonal noise covariance matrices.

2 BAMP with Vector Denoiser

We consider the measurement model

$$\mathbf{y}(b) = \mathbf{A}(b)\mathbf{x}(b) + \mathbf{w}(b)\,, \tag{1}$$

with $\mathbf{y}(b) \in \mathbb{R}^M$, $\mathbf{x}(b) \in \mathbb{R}^N$, $\mathbf{w}(b) \in \mathbb{R}^M$, and $\mathbf{A}(b) \in \mathbb{R}^{M\times N}$, for $b = 1, \ldots, B$. We denote the measurement rate by $R = M/N$. We assume that the measurement matrices $\mathbf{A}(b)$ are realizations of Gaussian or Rademacher random matrices with normalized columns. We define the length-B column vectors

$$\vec{\mathbf{x}}_n = (x_n(1), \ldots, x_n(B))^T,$$

(similar notation will be used throughout the paper). We focus on signals with multivariate BG probability density function (pdf), i.e.,

$$f_{\vec{\mathbf{x}}_n}(\vec{\mathbf{x}}_n) = f_{\vec{\mathbf{x}}}(\vec{\mathbf{x}}_n) = (1-\epsilon)\,\delta(\vec{\mathbf{x}}_n) + \epsilon\,\mathcal{N}(\vec{\mathbf{x}}_n; \mathbf{0}, \boldsymbol{\Sigma}_{\vec{\mathbf{x}}}), \tag{2}$$

independent and identically distributed (i.i.d.) over n; here, $\boldsymbol{\Sigma}_{\vec{\mathbf{x}}}$ is the covariance matrix of $\vec{\mathbf{x}}_n$ given that it is non-zero vector. The additive noise in (1) is assumed to be i.i.d. Gaussian over m with zero mean and covariance $\boldsymbol{\Sigma}_{\vec{\mathbf{w}}}$,

$$\vec{\mathbf{w}}_m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\vec{\mathbf{w}}})\,. \tag{3}$$

The BAMP algorithm for joint sparse recovery of $\mathbf{x}(b)$, $b = 1, \ldots, B$ is described in detail in, e.g., [**?**]. The state evolution is defined as

$$\boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^{t+1} = \begin{cases} \boldsymbol{\Sigma}_{\vec{\mathbf{w}}} + \frac{1}{R}\,\mathrm{E}_{\vec{\mathbf{x}},\vec{\mathbf{v}}}\left\{\langle \mathbf{e}(\vec{\mathbf{x}}, \vec{\mathbf{v}})\rangle\right\} & \text{for MMV,} \\ D\big(\boldsymbol{\Sigma}_{\vec{\mathbf{w}}} + \frac{1}{R}\,\mathrm{E}_{\vec{\mathbf{x}},\vec{\mathbf{v}}}\left\{\langle \mathbf{e}(\vec{\mathbf{x}}, \vec{\mathbf{v}})\rangle\right\}\big) & \text{for DCS,} \end{cases} \tag{4}$$

*The original of this paper has been published in the IEEE Journal of Selected Topics in Signal Processing Special Issue on 'Information-Theoretic Methods in Data Acquisition, Analysis, and Processing' [1]

where $\mathbf{e}(\vec{\mathbf{x}}, \vec{\mathbf{v}}) = F(\vec{\mathbf{x}} + \vec{\mathbf{v}}; \boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^t) - \vec{\mathbf{x}}$ is the error achieved by the MMSE estimator of BAMP ($F(\vec{\mathbf{u}}; \boldsymbol{\Sigma}_{\vec{\mathbf{v}}}) = \mathrm{E}_{\vec{\mathbf{x}}}\{\vec{\mathbf{x}} \mid \vec{\mathbf{x}} + \vec{\mathbf{v}} = \vec{\mathbf{u}}; \boldsymbol{\Sigma}_{\vec{\mathbf{v}}}\}$), $\langle\cdot\rangle$ is the outer product of a vector with itself, and $D(\cdot)$ is the orthogonal projection which zeros off-diagonal elements. The state in the MMV scenario is in general $B(B+1)/2$ dimensional. From (4), the MSE prediction directly follows as

$$\widehat{\mathrm{MSE}}^t(b) = R\left(\boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^t - \boldsymbol{\Sigma}_{\vec{\mathbf{w}}}\right)_{b,b}.$$

3 Results

3.1 Joint Diagonalization for MMV

Under the assumption that the covariance matrices $\boldsymbol{\Sigma}_{\vec{\mathbf{x}}}$ and $\boldsymbol{\Sigma}_{\vec{\mathbf{w}}}$ are full rank and using the fact that covariance matrices are symmetric and positive definite, there exists a nonsingular (but generally non-orthogonal) matrix $\mathbf{T}$ that simultaneously diagonalizes the covariance matrices of the signal $\vec{\mathbf{x}}$ and the noise $\vec{\mathbf{w}}$. In the transformed model

$$\tilde{\vec{\mathbf{y}}}_m = \mathbf{T}\vec{\mathbf{y}}_m, \quad \tilde{\vec{\mathbf{x}}}_n = \mathbf{T}\vec{\mathbf{x}}_n, \quad \tilde{\vec{\mathbf{w}}}_m = \mathbf{T}\vec{\mathbf{w}}_m,$$

$$\boldsymbol{\Sigma}_{\tilde{\vec{\mathbf{x}}}} = \mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{x}}}\mathbf{T}^T = \mathbf{I}_{B\times B},$$

$$\boldsymbol{\Sigma}_{\tilde{\vec{\mathbf{w}}}} = \mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{w}}}\mathbf{T}^T = \boldsymbol{\Lambda}^{-1} = \epsilon\,\mathrm{diag}\left(\frac{1}{\mathrm{SNR}(1)}, \dots, \frac{1}{\mathrm{SNR}(B)}\right).$$

This means that under mild conditions every MMV problem has an equivalent (DCS) CS problem but with diagonal signal and noise covariance matrix and possibly rescaled SNRs. Furthermore, one of the covariance matrices can be chosen as the identity and then the other carries the B (inverse) SNR values on its diagonal.

3.2 Equivariance of BAMP for MMV

Theorem 1. *The BAMP algorithm for MMV and its SE are equivariant w.r.t. invertible linear transformations. Denote one BAMP iteration by* $(\hat{\vec{\mathbf{x}}}_n^{t+1}, \vec{\mathbf{r}}_m^{t+1}, \boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^{t+1}) = V(\vec{\mathbf{y}}_m, \hat{\vec{\mathbf{x}}}_n^t, \vec{\mathbf{r}}_m^t, \boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^t)$. *For any nonsingular* $\mathbf{T}$*, we have for all* m *and* n

$$V(\mathbf{T}\vec{\mathbf{y}}_m, \mathbf{T}\hat{\vec{\mathbf{x}}}_n^t, \mathbf{T}\vec{\mathbf{r}}_m^t, \mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^t) = (\mathbf{T}\hat{\vec{\mathbf{x}}}_n^{t+1}, \mathbf{T}\vec{\mathbf{r}}_m^{t+1}, \mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^{t+1}\mathbf{T}^T).$$

Furthermore, the SE equation (4) *translates to the transformed domain as*

$$\mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{v}}}^{t+1}\mathbf{T}^T = \mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{w}}}\mathbf{T}^T + \frac{1}{R}\,\mathrm{E}_{\vec{\mathbf{x}},\vec{\mathbf{v}}}\left\{\langle F(\mathbf{T}(\vec{\mathbf{x}}+\vec{\mathbf{v}}); \mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{v}}}{}^t\mathbf{T}^T) - \mathbf{T}\vec{\mathbf{x}}\rangle\right\}. \quad (5)$$

Assume that BAMP converges to $\hat{\vec{\mathbf{x}}}_n$ with inputs $\vec{\mathbf{y}}_n$, $\boldsymbol{\Sigma}_{\vec{\mathbf{x}}}$, and $\boldsymbol{\Sigma}_{\vec{\mathbf{w}}}$; then, Theorem 1 implies that BAMP with inputs $\mathbf{T}\vec{\mathbf{y}}_n$, $\mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{x}}}\mathbf{T}$, and $\mathbf{T}\boldsymbol{\Sigma}_{\vec{\mathbf{w}}}\mathbf{T}$ converges to the solution $\mathbf{T}\hat{\vec{\mathbf{x}}}_n$.

3.3 Bernoulli-Gauss Prior

For the BG prior, after applying the transformation $\mathbf{T}$, the equivalent measurement model becomes

$$\tilde{\mathbf{y}}(b) = \mathbf{A}(b)\tilde{\mathbf{x}}(b) + \tilde{\mathbf{w}}(b), \quad \forall b \quad (6)$$

with signal and noise pdfs

$$f_{\tilde{\vec{\mathbf{x}}}}(\tilde{\vec{\mathbf{x}}}_n) = (1-\epsilon)\,\delta(\tilde{\vec{\mathbf{x}}}_n) + \epsilon\,\mathcal{N}(\tilde{\vec{\mathbf{x}}}_n; \mathbf{0}, \mathbf{I}), \quad (7)$$

$$f_{\tilde{\vec{\mathbf{w}}}}(\tilde{\vec{\mathbf{w}}}_m) = \mathcal{N}(\tilde{\vec{\mathbf{w}}}_m; \mathbf{0}, \boldsymbol{\Lambda}^{-1}). \quad (8)$$

That is, we retain a BG prior in the transformed domain, only with uncorrelated components. This is a distinctive feature of the BG prior and in general doesn't hold for other types of distributions.

In [**?**] we demonstrate that for the decorrelated model (6) with BG prior (7)–(8), the BAMP iterations under the MMV model preserve the diagonal structure of $\boldsymbol{\Sigma}_{\tilde{\vec{\mathbf{v}}}}^t$. Therefore, for CS measurements with multivariate BG signal prior, the decorrelation transformation has to be done only once before recovery.

3.4 Replica Analysis

In [6], the replica method was used to determine the MSE performance of BAMP for the measurement (1) and the BG prior (2), assuming $\boldsymbol{\Sigma}_{\vec{\mathbf{x}}} = \mathbf{I}$ and isotropic uncorrelated noise, i.e., $\boldsymbol{\Sigma}_{\vec{\mathbf{w}}} = \sigma_{\mathsf{w}}^2\mathbf{I}$. In this special case MMV and DCS are equivalent. Due to the joint diagonalization approach from Section 3.1, it suffices to extend the replica analysis to the case with $\boldsymbol{\Sigma}_{\vec{\mathbf{x}}} = \mathbf{I}$ and $\boldsymbol{\Sigma}_{\vec{\mathbf{w}}} = \mathrm{diag}(\sigma_{\mathsf{w}}^2(1), \dots, \sigma_{\mathsf{w}}^2(B))$ to cover the analysis of arbitrary $\boldsymbol{\Sigma}_{\vec{\mathbf{x}}}$ and $\boldsymbol{\Sigma}_{\vec{\mathbf{w}}}$. In particular, the replica method is capable of predicting the fixed points of BAMP in the asymptotic regime ($N, M \to \infty$, $R = M/N = \mathrm{const.}$), as a function of the set of B MSEs. Assuming $\boldsymbol{\Sigma}_{\vec{\mathbf{x}}} = \mathbf{I}$ and $\boldsymbol{\Sigma}_{\vec{\mathbf{w}}} = \mathrm{diag}\left(\sigma_{\mathsf{w}}^2(1), \dots, \sigma_{\mathsf{w}}^2(B)\right)$, we compute in [**?**] the free energy $\mathcal{F}(\vec{\mathbf{E}})$ as a function of the MSE vector $\vec{\mathbf{E}} = (E(1), \dots, E(b))^T$ with $E(b) = \mathrm{MSE}(b)$. The stationary points of $\mathcal{F}(\vec{\mathbf{E}})$ correspond to fixed points of BAMP in the asymptotic regime. Thus, we can determine the component-wise MSEs of BAMP by evaluating the free energy and finding the largest components of $\vec{\mathbf{E}}$ that correspond to a local maximum of $\mathcal{F}(\vec{\mathbf{E}})$. In the free energy function, local maxima correspond to stable fixed points and local minima to unstable fixed points, whereas the global maximum of $\mathcal{F}(E)$ corresponds to the MMSE. BAMP typically achieves the largest MSE associated with a local maximum.

The dynamics of the free energy function and that of the SE show a close match. In particular, the stationary points of the former match with the fixed points of the latter, and the gradient of the free energy shows close similarity with the SE arrows in the B-dimensional state space. This enables us to gain new insights into the behaviour of BAMP. In particular, simulation results suggest several fundamental observations and questions:

- When K of the B signal vectors fully correlate, BAMP essentially collapses into a $B-K+1$-dimensional CS problem (proven).
- For large but non-asymptotic B we observe that the phase transition does not occur and the MSE decreases smoothly with R. Further research is necessary.
- The stationary points of the multidimensional free energy function appear to lie on a globally attracting 1-D submanifold. Can the B-dimensional SE dynamics be compressed back into a one-dimensional evolution in some way?
- In which subregion of parameters (R, ϵ, $\sigma_{\mathsf{w}}^2(1), \dots, \sigma_{\mathsf{w}}^2(B)$) is BAMP MMSE optimal?

References

[1] G. Hannak, A. Perelli, N. Goertz, G. Matz, and M. E. Davies, "Performance analysis of approximate message passing for distributed compressed sensing," in IEEE Journal of Selected Topics in Signal Processing, 2017.

[2] E. J. Candes, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” Communications on Pure and Applied Mathematics, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” IEEE Transactions on Signal Processing, vol. 53, pp. 2477–2488, Jul. 2005.

[4] M. F. Duarte, S. Sarvotham, M. B. Wakin, D. Baron, and R. G. Baraniuk, “Joint sparsity models for distributed compressed sensing,” in Proceedings of the Workshop on Signal Processing with Adaptative Sparse Structured Representations, IEEE, 2005

[5] D. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing: I. motivation and construction,” in 2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo), pp. 1–5, Jan. 2010.

[6] J. Zhu, D. Baron, and F. Krzakala, “Performance limits for noisy multi-measurement vector problems,” IEEE Transactions on Signal Processing, vol. 65, no. 9, pp. 2444–2454, 2017.