Compressive Sensing under Matrix Uncertainties: An Approximate Message Passing Approach

Jason T. Parker
Sensors Directorate
US Air Force Research Laboratory
Wright Patterson AFB, OH 45433
Email: jason.parker@wpafb.af.mil

Volkan Cevher
Laboratory for Information and Inference Systems
Ecole Polytechnique Federale de Lausanne
Lausanne, Switzerland
Email: volkan.cevher@epfl.ch

Philip Schniter
Dept. of ECE
The Ohio State University
Columbus, OH 43210
Email: schniter@ece.osu.edu

***Abstract*—In this work, we consider a general form of noisy compressive sensing (CS) when there is uncertainty in the measurement matrix as well as in the measurements. Matrix uncertainty is motivated by practical cases in which there are imperfections or unknown calibration parameters in the signal acquisition hardware. While previous work has focused on analyzing and extending classical CS algorithms like the LASSO and Dantzig selector for this problem setting, we propose a new algorithm whose goal is either minimization of mean-squared error or maximization of posterior probability in the presence of these uncertainties. In particular, we extend the Approximate Message Passing (AMP) approach originally proposed by Donoho, Maleki, and Montanari, and recently generalized by Rangan, to the case of probabilistic uncertainties in the elements of the measurement matrix. Empirically, we show that our approach performs near oracle bounds. We then show that our matrix-uncertain AMP can be applied in an alternating fashion to learn both the unknown measurement matrix and signal vector. We also present a simple analysis showing that, for suitably large systems, it suffices to treat uniform matrix uncertainty as additive white Gaussian noise.**

I. Introduction

In compressive sensing (CS), the goal is to reconstruct an N-dimensional signal $\boldsymbol{x}$ from $M < N$ linear measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}$, where $\boldsymbol{w}$ is additive noise. In the noiseless case, it is by now well known that, when the signal is exactly K-sparse and the measurement matrix $\boldsymbol{A}$ satisfies certain properties (e.g., restricted isometry, null space, or spark), it is possible to exactly reconstruct the signal from $M = \mathcal{O}(K \log N/K)$ measurements using polynomial-complexity algorithms (e.g., greedy or convex-optimization based). Moreover, these methods can accurately reconstruct the signal in the noisy case, even when the signal is compressible rather than exactly sparse (e.g., [1]).

These results are, however, predicated on knowing the measurement matrix $\boldsymbol{A}$ perfectly. In practical applications of CS, it is reasonable to expect uncertainty in the linear measurement matrix $\boldsymbol{A}$ due to, e.g., imperfections in the signal acquisition hardware, model mismatch, parameter discretization, and other factors.

Several authors have analyzed the impact of measurement-matrix uncertainty on existing CS algorithms, e.g., Herman and Strohmer [2], Herman and Needell [3], and Chi, Pezeshki, Scharf, and Calderbank [4]. Herman et al. analyze the effect of additive perturbations on the Basis Pursuit and CoSaMP algorithms, respectively, whereas Chi et al. analyze the effect, on Basis Pursuit, of a multiplicative basis mismatch matrix that takes the form of the identity matrix plus a perturbation. In [2]–[4], the authors study the worst-case effects on established algorithms, but stop short of proposing new algorithms.

We are aware of only a few algorithms that explicitly address measurement-matrix uncertainty, all of which consider the additive uncertainty model $\boldsymbol{A} = \hat{\boldsymbol{A}} + \boldsymbol{E}$, where $\hat{\boldsymbol{A}}$ is known and $\boldsymbol{E}$ is an unknown perturbation, yielding the observations

$$\boldsymbol{y} = (\hat{\boldsymbol{A}} + \boldsymbol{E})\boldsymbol{x} + \boldsymbol{w}. \tag{1}$$

In [5], Zhu et al. develop the Sparsity-cognizant Total Least Squares (S-TLS) approach, which extends the classical TLS approach (widely applied in the context of ℓ_2 regularization) to ℓ_1 regularization, yielding

$$\{\hat{\boldsymbol{x}}_{\text{S-TLS}}, \hat{\boldsymbol{E}}_{\text{S-TLS}}\} = \arg\min_{\boldsymbol{x},\boldsymbol{E}} \|(\hat{\boldsymbol{A}} + \boldsymbol{E})\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda_E \|\boldsymbol{E}\|_F^2 + \lambda \|\boldsymbol{x}\|_1. \tag{2}$$

In [6], Rosenbaum and Tsybakov propose the MU-Selector, a modified version of the Dantzig selector [7], which reads

$$\{\hat{\boldsymbol{x}}_{\text{MU-Selector}}\} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ s. t. } \|\hat{\boldsymbol{A}}^H(\boldsymbol{y} - \hat{\boldsymbol{A}}\boldsymbol{x})\|_\infty \le \lambda \|\boldsymbol{x}\|_1 + \epsilon. \tag{3}$$

The above criteria assume relatively little about the structure of the perturbations $\boldsymbol{w}$ and $\boldsymbol{E}$, and thus obtain algorithms with wide applicability, but—as we shall see—limited performance. In [5], Zhu et al. also proposed a Weighted S-TLS (WS-TLS) that can exploit structure in the matrix uncertainty $\boldsymbol{E}$ and perform significantly better than S-TLS.

In this paper, we address sparse-signal recovery under matrix uncertainty in a Bayesian framework with informative priors. In particular, we extend the Approximate Message Passing (AMP) algorithm recently proposed by Donoho,

Jason Parker is supported by an AFOSR lab task under the direction of Dr. Arje Nachman.

Volkan Cevher is supported by the European Commission under Grant MIRG-268398, SNF 200021_132548, and DARPA KECom program #11-DARPA-1055. He also acknowledges his Faculty Fellow position in the Dept. of Electrical Engineering at Rice University.

Philip Schniter is supported in part by NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090.

Maleki, and Montanari [8]—and in particular the Generalized AMP (GAMP) proposed by Rangan [9]—to the case of probabilistic uncertainty on the elements of the measurement matrix $\boldsymbol{A}$. Initially, we treat the entries of $\boldsymbol{A}$ as independent random variables that are known only in mean and variance, which can both vary across the entries. The resulting Matrix-Uncertain GAMP (MU-GAMP) provides a computationally efficient way to obtain nearly minimum-mean-squared-error (MMSE) estimates of the unknown signal $\boldsymbol{x}$ in the presence of uncertainties in both the linear matrix transformation $\boldsymbol{A}$ as well as the observations of the transformed outputs $\boldsymbol{Ax}$.

We then turn our attention to parametric matrices of the form $\boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{A}_0 + \sum_{p=1}^{P} \theta_p \boldsymbol{A}_p$, where $\{\boldsymbol{A}_p\}$ are known and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_P]^{\mathsf{T}}$ unknown. We then propose a scheme that alternates between the estimation of $\boldsymbol{\theta}$ and the estimation of $\boldsymbol{x}$. Conveniently, both estimation steps can be performed using the already developed MU-GAMP framework. A salient feature of this approach is that we alternate soft estimates as opposed to point estimates.

Throughout the paper, we use boldface capital letters to denote matrices and boldface small letters to denote vectors, $\boldsymbol{I}$ and $\boldsymbol{0}$ to denote the identity matrix and zero matrices, $(\cdot)^{\mathsf{T}}$ transpose, and $(\cdot)^*$ conjugate. For x_j a realization of random variable X_j, we use $\mathrm{E}_{X_j}\{x_j\}$ to denote mean, $\mathrm{var}_{X_j}\{x_j\}$ variance, $p_{X_j}(x_j)$ the pdf, and $p_{X_j|D_j}(x_j \,|\, d_j)$ the pdf conditioned on $D_j = d_j$, and sometimes we omit the subscript when there is no danger of confusion. To denote the Gaussian pdf with mean $\hat{x}$ and variance ν^x, we use $\mathcal{N}(x; \hat{x}, \nu^x)$.

II. A Large-System Blessing?

Before getting into the details of MU-GAMP, we make a curious observation: As the problem dimensions grow large, the effect of *uniform* matrix uncertainty is identical to additive white Gaussian noise (AWGN) on the observations. The following proposition makes our claim precise.

Proposition 2.1: Consider an M-dimensional observation of the form in (1), written equivalently as

$$\boldsymbol{y} = \hat{\boldsymbol{A}}\boldsymbol{x} + \boldsymbol{e} + \boldsymbol{w} \quad \text{for} \quad \boldsymbol{e} \triangleq \boldsymbol{E}\boldsymbol{x}. \tag{4}$$

Suppose that N-dimensional $\boldsymbol{x}$ is K-sparse, and that the matrix uncertainty $\boldsymbol{E}$ is *uniform*, i.e., $\{E_{mn}\}$ are i.i.d zero-mean random variables with variance $\nu^E = c^E/M$ for bounded positive c^E (but otherwise arbitrary distribution). In the large-system limit (i.e., $M, N, K \to \infty$ with fixed $\delta \triangleq M/N$ and $\rho \triangleq K/M$), the additive perturbation $\boldsymbol{e}$ becomes i.i.d zero-mean Gaussian with variance $\nu^e = c^E \delta^{-1} \|\boldsymbol{x}\|_2^2 / N$.

Proof: Since the rows of $\boldsymbol{E}$ are statistically independent, the elements $\{e_m\}$ of $\boldsymbol{e}$ are independent as well. Moreover, $e_m = \sum_{k=1}^{K} E_{m,n(k)} x_{n(k)}$, where $n(k)$ indexes the k^{th} non-zero element of $\boldsymbol{x}$. Thus, in the large-system limit (i.e., $K \to \infty$), the central limit theorem implies that e_m is zero-mean Gaussian with variance $\nu^e \triangleq \nu^E \|\boldsymbol{x}\|_2^2 = c^E \delta^{-1} \|\boldsymbol{x}\|_2^2 / N$. ■

The implication of Proposition 2.1 is that, for problems of *uniform* matrix uncertainty and *suitably large* dimension, there is no need to design new algorithms that handle matrix uncertainty; those designed to handle AWGN (e.g., LASSO [10], GAMP, etc.) suffice, so long as they are properly tuned to handle the additional AWGN power ν^e.

Now, whether or not the large-system behavior predicted by Proposition 2.1 manifests at a given *finite* (M, N, K) depends on the distribution of i.i.d $\{E_{mn}\}$ and the sparsity K. If $\{E_{mn}\}$ are far from Gaussian (e.g., sparse) and K is relatively small, the distribution of $\{e_m\}$ can be far from Gaussian. On the other hand, if $\{E_{mn}\}$ is Gaussian, then e_m will also be Gaussian, for any K.

Although, to our knowledge, Proposition 2.1 is novel, the empirical results in previous works support its claim; see, e.g., the negligible difference between optimally-tuned versions of S-TLS and LASSO under i.i.d Gaussian $\boldsymbol{E}$ in [5, Fig. 3]. In Section III-C, we will provide further empirical evidence for our claim.

III. Matrix-Uncertain GAMP

A. Background on GAMP

In the Bayesian approach to compressed sensing, it is typically presumed that the signal $\boldsymbol{x}$ is drawn from a known separable pdf $p(\boldsymbol{x}) = \prod_n p_X(x_n)$, where $p_X(.)$ promotes sparsity or compressibility. Similarly, the noise $\boldsymbol{w}$ is drawn from a known separable pdf $p(\boldsymbol{w}) = \prod_m p_W(w_m)$. Given the observations $\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{w}$, one would ideally like to compute the full joint posterior $p(\boldsymbol{x} \,|\, \boldsymbol{y})$. This is, however, not tractable for the pdfs and problem dimensions typical in compressed sensing. Thus, one often settles for approximate MAP or MMSE estimates.

The original AMP algorithm [8] assumes Laplacian $p_X(.)$ and Gaussian $p_W(.)$, and seeks the MAP solution using an approximation of loopy belief propagation. The approximation, which becomes tight in the large-system limit, is based on the CLT and Taylor-series expansions, and relies on the elements of $\boldsymbol{A}$ to be known realizations of an independent zero-mean $1/M$-variance random variable.

Rangan proposed a Generalized AMP (GAMP) [9] that 1) handles either MAP or MMSE, 2) allows arbitrary A_{mn}, 3) allows an arbitrary signal distribution $p_X(.)$, and 4) allows an arbitrary separable pdf $p(\boldsymbol{y} \,|\, \boldsymbol{z}) = \prod_m p_{Y|Z}(y_m \,|\, z_m)$ relating the observations $\boldsymbol{y}$ to the linearly transformed outputs $\boldsymbol{z} \triangleq \boldsymbol{Ax}$. This observation-uncertainty model subsumes the case of additive noise $\boldsymbol{w}$ with arbitrary distribution $p_W(.)$ via $p_{Y|Z}(y_m \,|\, z_m) = p_W(y_m - z_m)$, but also handles nonlinear output transformations like that used in logistic regression.

B. Matrix-Uncertain GAMP

We now propose a Matrix-Uncertain GAMP (MU-GAMP) that extends GAMP [9] to the case of uncertainty in the measurement matrix $\boldsymbol{A}$. Unlike GAMP, which treats $\{A_{mn}\}$ as fixed and known, MU-GAMP treats $\{A_{mn}\}$ as independent random variables with known mean and variance,

$$\hat{A}_{mn} = \mathrm{E}\{A_{mn}\} \tag{5}$$

$$\nu^A_{mn} = \mathrm{var}\{A_{mn}\}, \tag{6}$$

reducing to GAMP in the case that $\nu^A_{mn} = 0$. Note that, with $\boldsymbol{E} \triangleq \boldsymbol{A} - \hat{\boldsymbol{A}}$, we recover exactly the perturbation model $\boldsymbol{A} =$

definitions:

$$p_{Z|Y}(z|y;\hat{z},\nu^z) = \frac{p_{Y|Z}(y|z)\,\mathcal{N}(z;\hat{z},\nu^z)}{\int_{z'} p_{Y|Z}(y|z')\,\mathcal{N}(z';\hat{z},\nu^z)} \quad \text{(D1)}$$

$$g_{\text{out}}(y,\hat{z},\nu^z) = \tfrac{1}{\nu^z}\big(\mathrm{E}_{Z|Y}\{z|y;\hat{z},\nu^z\} - \hat{z}\big) \quad \text{(D2)}$$

$$g'_{\text{out}}(y,\hat{z},\nu^z) = \tfrac{1}{\nu^z}\Big(\frac{\mathrm{var}_{Z|Y}\{z|y;\hat{z},\nu^z\}}{\nu^z} - 1\Big) \quad \text{(D3)}$$

$$p_{X|\boldsymbol{Y}}(x|\boldsymbol{y};\hat{r},\nu^r) = \frac{p_X(x)\,\mathcal{N}(x;\hat{r},\nu^r)}{\int_{x'} p_X(x')\,\mathcal{N}(x';\hat{r},\nu^r)} \quad \text{(D4)}$$

$$g_{\text{in}}(\hat{r},\nu^r) = \textstyle\int_x x\, p_{X|\boldsymbol{Y}}(x|\boldsymbol{y};\hat{r},\nu^r) \quad \text{(D5)}$$

$$g'_{\text{in}}(\hat{r},\nu^r) = \tfrac{1}{\nu^r}\textstyle\int_x |x-g_{\text{in}}(\hat{r},\nu^r)|^2\, p_{X|\boldsymbol{Y}}(x|\boldsymbol{y};\hat{r},\nu^r) \quad \text{(D6)}$$

initialize:

$$\forall n: \hat{x}_n(1) = \textstyle\int_x x\, p_X(x) \quad \text{(I1)}$$

$$\forall n: \nu^x_n(1) = \textstyle\int_x |x-\hat{x}_n(1)|^2 p_X(x) \quad \text{(I2)}$$

$$\forall m: \hat{u}_m(0) = 0 \quad \text{(I3)}$$

for $t = 1,2,3,\ldots$

$$\forall m: \hat{z}_m(t) = \textstyle\sum_{n=1}^N \hat{A}_{mn}\hat{x}_n(t) \quad \text{(R1)}$$

$$\forall m: \nu^z_m(t) = \textstyle\sum_{n=1}^N |\hat{A}_{mn}|^2\nu^x_n(t) \quad \text{(R2a)}$$

$$\forall m: \nu^p_m(t) = \nu^z_m(t) + \textstyle\sum_{n=1}^N \nu^A_{mn}\left(\nu^x_n + |\hat{x}_n(t)|^2\right) \quad \textbf{(R2b)}$$

$$\forall m: \hat{p}_m(t) = \hat{z}_m(t) - \nu^z_m(t)\,\hat{u}_m(t-1) \quad \text{(R3)}$$

$$\forall m: \hat{u}_m(t) = g_{\text{out}}(y_m,\hat{p}_m(t),\nu^p_m(t)) \quad \text{(R4)}$$

$$\forall m: \nu^u_m(t) = -g'_{\text{out}}(y_m,\hat{p}_m(t),\nu^p_m(t)) \quad \text{(R5)}$$

$$\forall n: \nu^r_n(t) = \big(\textstyle\sum_{n=1}^N |\hat{A}_{mn}|^2\nu^u_m(t)\big)^{-1} \quad \text{(R6)}$$

$$\forall n: \hat{r}_n(t) = \hat{x}_n(t) + \nu^r_n(t)\textstyle\sum_{m=1}^M \hat{A}^*_{mn}\hat{u}_m(t) \quad \text{(R7)}$$

$$\forall n: \nu^x_n(t+1) = \nu^r_n(t)\,g'_{\text{in}}(\hat{r}_n(t),\nu^r_j(t)) \quad \text{(R8)}$$

$$\forall n: \hat{x}_n(t+1) = g_{\text{in}}(\hat{r}_n(t),\nu^r_n(t)) \quad \text{(R9)}$$

end

TABLE I
THE MU-GAMP ALGORITHM

$\hat{\boldsymbol{A}} + \boldsymbol{E}$ used in (1), but now with the implicit assumption that E_{mn} has zero mean and variance ν^A_{mn}.

Due to lack of space, we are unable to provide a derivation of MU-GAMP here, but we note that the approximations on which it is based (and the notation we use to summarize it) are the same as those used for standard GAMP. The resulting algorithm is given in Table I,[1] where the only difference from the original GAMP is the additional step (R2b). With this step, MU-GAMP requires an additional matrix multiply, although the cost of this multiplication may be reduced when ν^A_{mn} is structured. For example, when $\nu^A_{mn} = \nu^A_m \;\forall n$, the matrix multiplication in (R2b) reduces to a sum.

C. Empirical Study

We now study empirical performance under *uniform* and *non-uniform* matrix uncertainty. In both cases, we plot Normalized Mean Squared Error (NMSE) versus M/N at $N = 256$ and $K/M = 0.2$, where the relatively small problem size was used due to the implementation complexity of the MU-Selector. The K non-zero entries of the signal $\boldsymbol{x}$ were drawn ± 1 with equal probability, the (known) matrix means $\{\hat{A}_{mn}\}$ were i.i.d $\mathcal{N}(0, 1/M)$, and the noise $\boldsymbol{w}$ was i.i.d $\mathcal{N}(0,\nu^w)$.

To illustrate the effect of *uniform* matrix uncertainty, we drew the matrix errors $\{E_{mn}\}$ i.i.d $\mathcal{N}(0,\nu^E)$, noting that in this case $\boldsymbol{e} = \boldsymbol{E}\boldsymbol{x}$ is truly i.i.d Gaussian (for any given $\boldsymbol{x}$). Moreover, we set $\nu^E = \nu^w$ such that the effective SNR $\mathrm{E}\{\|\hat{\boldsymbol{A}}\boldsymbol{x}\|_2^2\}/\mathrm{E}\{\|\boldsymbol{e}+\boldsymbol{w}\|_2^2\} = 20$ dB. Under this setup, we ran MU-GAMP under the true (uniform) matrix error variance $\nu^A_{mn} = \nu^E$, the true noise statistics, the true signal variance and sparsity rate, but a (mismatched) Bernoulli-Gaussian signal pdf. We also ran the original GAMP under the same signal prior and the compensated AWGN variance $\nu^e + \nu^w$, for $\nu^e \triangleq \mathrm{var}\{e_m\} = K\nu^E$. We then ran S-TLS, the MU-Selector, and LASSO (via SpaRSA [11]), each debiased and with "genie-aided" tuning: for each realization, each algorithm was run under several values of its tuning parameter, and the tuning yielding minimal NMSE was selected.

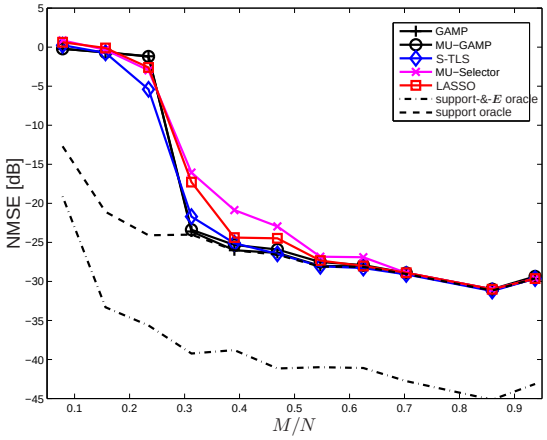

Fig. 1. 10-trial median NMSE under uniform matrix error variance ν^E.

Figure 1 shows the resulting NMSE performance of each algorithm, as well as that of two oracle estimators: support-aware LMMSE, and support-and-$\boldsymbol{E}$-aware LMMSE. We note that, under a Bernoulli-Gaussian signal pdf, the NMSEs of GAMP and MU-GAMP are lower bounded by these respective oracles. The figure shows that GAMP and MU-GAMP yield essentially identical NMSE, and that for $M/N > 0.3$, this NMSE essentially coincides with the support-oracle bound. Meanwhile, the debiased and genie-tuned incarnations of S-TLS, the MU-Selector, and LASSO show performance that is only slightly worse than GAMP and MU-GAMP for $M/N > 0.3$. The fact that the matrix-uncertain algorithms (i.e., MU-GAMP, S-TLS, MU-Selector) and the standard algorithms (i.e., GAMP, LASSO) perform near-identically under *uniform* matrix uncertainty confirms the claim of Proposition 2.1.

Next, we examine the effect of *non-uniform* matrix uncertainty. For this, we used the same setup as in the previous experiment, except that we used *non-uniform* variances $\{\nu^E_{mn}\}$ such that $\nu^E_{mn} = 0$ for 99% of the entries, while $\nu^E_{mn} = C^E$ for the remaining 1% of the entries, where C^E was chosen to make the cumulative error ν^e identical to the previous experiment. MU-GAMP was then run under the true (now non-uniform) $\nu^A_{mn} = \nu^E_{mn}$, while GAMP was run under the compensated AWGN variance $\nu^e + \nu^w$, as before. We also implemented the Weighted S-TLS (WS-TLS) from [5], which was given knowledge of the non-uniform $\{\nu^E_{mn}\}$.

Figure 2 shows the resulting NMSE. In the figure, we see that the algorithms assuming uniform matrix uncertainty ν^E (i.e., S-TLS and the MU-Selector) perform essentially the same in this experiment as they did in the previous experiment, which is due to the fact that ν^e was calibrated across experiments. Furthermore, these algorithms do essentially no better than those designed for AWGN (i.e., LASSO and GAMP),

[1] A MATLAB implementation of GAMP, including the MU extension, is available at `http://sourceforge.net/projects/gampmatlab/`.

which makes sense in light of Proposition 2.1. However, the algorithms exploiting non-uniform uncertainty $\{\nu^E_{mn}\}$ (i.e., WS-TLS and MU-GAMP) do significantly better. In fact, MU-GAMP performs quite close to the support-and-$\boldsymbol{E}$-aware oracle bound for $M/N > 0.3$.

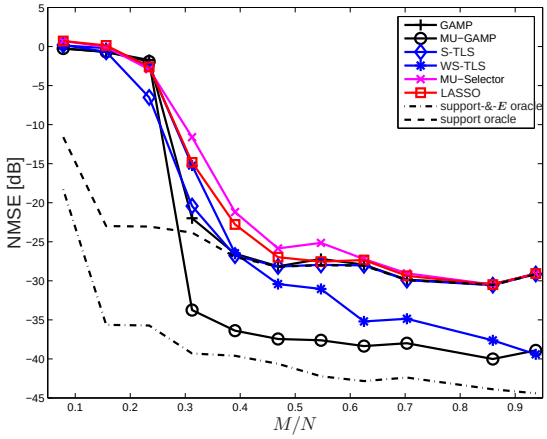

Fig. 2. 10-trial median NMSE under non-uniform error variance $\{\nu^E_{mn}\}$.

IV. Alternating MU-GAMP

The performance of any reasonable compressive-sensing algorithm will improve as matrix uncertainty diminishes, and one way to reduce uncertainty is to explicitly estimate the unknown matrix $\boldsymbol{A}$. In fact, this is the goal of Dictionary Learning [12], where a large number of measurement vectors $\{\boldsymbol{y}_t\}_{t=1}^T$ are assumed to be available. Since we are interested in estimating $\boldsymbol{A}$ from one (or very few) measurement vectors, we consider structured forms of $\boldsymbol{A}$ that depend on only a few parameters $\boldsymbol{\theta} \in \mathbb{C}^P$. In particular, we consider affine linear[2] models of the form (noting similarities to [5])

$$\boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{A}_0 + \textstyle\sum_{p=1}^P \theta_p \boldsymbol{A}_p \tag{7}$$

with known $\{\boldsymbol{A}_p\}_{p=0}^P$ and unknown $\boldsymbol{\theta}$. Several examples of this structure are discussed in the sequel. Moreover, (7) handles the case of *un*structured $\boldsymbol{A}$ via $P = MN$, $\boldsymbol{A}_0 = \boldsymbol{0}$, and $\{\boldsymbol{A}_p\}_{p=1}^P$ each containing a single distinct non-zero entry.

A. Alternating MU-GAMP

We now propose a scheme to jointly estimate $\{\boldsymbol{x}, \boldsymbol{\theta}\}$ based on the previously developed MU-GAMP. The proposed scheme is an iterative one that alternates between the estimation of $\boldsymbol{x}$ and $\boldsymbol{\theta}$. Say the mean and variance of θ_p are given by $\hat{\theta}_p$ and ν^θ_p, respectively. Then it holds that

$$\hat{A}_{mn} \triangleq \mathrm{E}\{A_{mn}(\boldsymbol{\theta})\} = A_{0,mn} + \textstyle\sum_{p=1}^P \hat{\theta}_p A_{p,mn} \tag{8}$$

$$\nu^A_{mn} \triangleq \mathrm{var}\{A_{mn}(\boldsymbol{\theta})\} = \textstyle\sum_{p=1}^P \nu^\theta_p |A_{p,mn}|^2, \tag{9}$$

where $A_{p,mn}$ denotes the m^{th} row and n^{th} column of $\boldsymbol{A}_p$. Thus, given the soft parameter estimates $(\hat{\boldsymbol{\theta}}, \boldsymbol{\nu}^\theta)$, one can directly compute the matrix uncertainty statistics $\{\hat{A}_{mn}\}$ and $\{\nu^A_{mn}\}$, and—with them—run MU-GAMP to estimate the signal vector $\boldsymbol{x}$, which will produces the marginal posterior mean and variance vectors $(\hat{\boldsymbol{x}}, \boldsymbol{\nu}^x)$.

Then, given the soft signal estimates $(\hat{\boldsymbol{x}}, \boldsymbol{\nu}^x)$, we can update the parameter means and variances $(\hat{\boldsymbol{\theta}}, \boldsymbol{\nu}^\theta)$, also using MU-GAMP. To see how, we first notice that the linear outputs $\boldsymbol{z}$ in the GAMP observation model $p(\boldsymbol{y} \,|\, \boldsymbol{z})$ take the form

$$\boldsymbol{z} = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x} = \boldsymbol{A}_0\boldsymbol{x} + \textstyle\sum_{p=1}^P \boldsymbol{A}_p \boldsymbol{x}\, \theta_p = \boldsymbol{B}(\boldsymbol{x})\boldsymbol{\theta} \tag{10}$$

for $\boldsymbol{\theta} \triangleq [\theta_0, \theta_1, \ldots, \theta_P]^\mathsf{T}$, $\theta_0 \triangleq 1$, and the (uncertain) matrix

$$\boldsymbol{B}(\boldsymbol{x}) \triangleq \left[\, \boldsymbol{A}_0\boldsymbol{x} \,\middle|\, \boldsymbol{A}_1\boldsymbol{x} \,\middle|\, \cdots \,\middle|\, \boldsymbol{A}_P\boldsymbol{x} \,\right]. \tag{11}$$

Given $(\hat{\boldsymbol{x}}, \boldsymbol{\nu}^x)$, the mean and variance of B_{mp} are simply

$$\hat{B}_{mp} \triangleq \mathrm{E}\{B_{mp}(\boldsymbol{x})\} = \textstyle\sum_{n=1}^N A_{p,mn}\hat{x}_n \tag{12}$$

$$\nu^B_{mp} \triangleq \mathrm{var}\{B_{mp}(\boldsymbol{x})\} = \textstyle\sum_{n=1}^N |A_{p,mn}|^2 \nu^x_n, \tag{13}$$

which, together with an appropriate prior pdf on $\{\theta_p\}$, are the ingredients needed to estimate $\boldsymbol{\theta}$ with MU-GAMP, yielding updated soft outputs $(\hat{\boldsymbol{\theta}}, \boldsymbol{\nu}^\theta)$. For example, if $\{\theta_p\}_{p=1}^P$ were known to be sparse, then a sparsifying prior would be appropriate. For θ_0, a prior with all mass at 1 would suffice to handle the constraint $\theta_0 = 1$.

Alternating between these two MU-GAMP steps, we can obtain successively refined estimates of $(\hat{\boldsymbol{x}}, \boldsymbol{\nu}^x)$ and $(\hat{\boldsymbol{\theta}}, \boldsymbol{\nu}^\theta)$. Each MU-GAMP step itself involves several iterations, but relatively few would be needed if they were "warm started" at the values of the previous estimates. Note that, unlike typical iterative schemes for dictionary learning [12], which alternate between point estimates, ours alternate between *soft* estimates, i.e., mean/variance pairs.

B. Empirical Study

Below, we present the results of three empirical experiments that investigate MU-GAMP and alternating MU-GAMP (A-MU-GAMP) under parametric matrix uncertainty. In all cases, we used $M = 103$, $N = 256$, i.i.d Gaussian $\boldsymbol{A}_0 \in \mathbb{C}^{M\times N}$ and $\boldsymbol{\theta} \in \mathbb{C}^P$, i.i.d Bernoulli-Gaussian $\boldsymbol{x} \in \mathbb{C}^N$ with $K = 20$, and complex AWGN with SNR=20 dB. MU-GAMP was given the apriori matrix statistics $\{\hat{A}_{mn}, \nu^A_{mn}\}$ from (8)-(9). A-MU-GAMP was initialized with the same statistics, but was able to drive down the variances $\{\nu^A_{mn}\}$ through its iterations.

First, we study the role of matrix-uncertainty dimension P on the NMSE performance of MU-GAMP and A-MU-GAMP. For this example, we used i.i.d Gaussian $\{\boldsymbol{A}_p\}_{p=1}^P$. As P was varied, $\{\nu^\theta_p\}$ was normalized to fix the energy of the uncertainty term $\boldsymbol{E} = \sum_{p=1}^P \theta_p \boldsymbol{A}_p$. Fig. 3 shows the resulting NMSE-versus-P, where—as expected—MU-GAMP maintains a constant performance versus P, whereas A-MU-GAMP benefits when P is small (and thus $\boldsymbol{\theta}$ can be learned).

Next, we consider a *channel-calibration* example involving $P = 10$ parallel linear measurement "channels", each with an unknown offset. For this, we constructed each matrix $\{\boldsymbol{A}_p\}_{p=1}^P$ to have ones in $1/P$ of its rows and zeros elsewhere, so that θ_p modeled the additive error in the p^{th} channel. Figure 4

[2]The affine linear model (7) may arise from a first-order Taylor series approximation of a non-linear model $\boldsymbol{A}(\boldsymbol{\theta})$ around the point $\hat{\boldsymbol{\theta}}$, in which case $\boldsymbol{A}_0 = \boldsymbol{A}(\hat{\boldsymbol{\theta}})$ and $\boldsymbol{A}_p = \partial \boldsymbol{A}(\boldsymbol{\theta})/\partial\theta_p|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

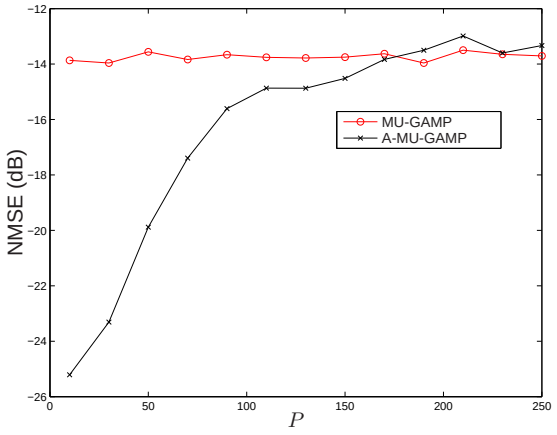

Fig. 3. 10-trial median NMSE for estimation of $\boldsymbol{x}$ versus the parametric matrix-uncertainty dimension P.

shows that, over its iterations, A-MU-GAMP approaches the performance of $\boldsymbol{\theta}$-aware GAMP when estimating $\boldsymbol{x}$, which comes within 2 dB of the support-and-$\boldsymbol{\theta}$-aware oracle MMSE. The star indicates the performance of MU-GAMP, which is about 20 dB worse. Meanwhile, when estimating $\boldsymbol{\theta}$, A-MU-GAMP approaches the performance of $\boldsymbol{x}$-aware GAMP.

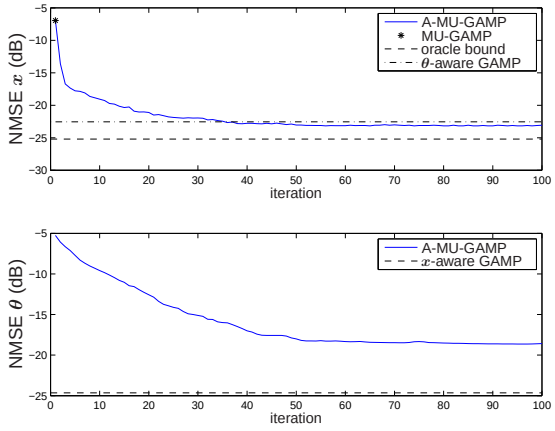

Fig. 4. 100-trial median NMSE of A-MU-GAMP when iteratively estimating $\boldsymbol{x}$ and $\boldsymbol{\theta}$ in the channel calibration example.

Finally, we consider a *compressive blind-deconvolution* example. Here, $\boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\,\mathcal{C}(\boldsymbol{\theta})$ where $\mathcal{C}(\boldsymbol{\theta})$ is circulant with first column $\boldsymbol{\theta} \in \mathbb{C}^N$ and $\boldsymbol{\Phi} = \begin{bmatrix}\boldsymbol{I}_M & \boldsymbol{0}\end{bmatrix}$. Due to the size of the uncertainty dimension, $P = N$, we used $T = 8$ measurement vectors $\{\boldsymbol{y}_t\}_{t=1}^T$, which is still much fewer than typical in dictionary learning. Figure 5 demonstrates that, once again, A-MU-GAMP is able to effectively learn both $\boldsymbol{x}$ and $\boldsymbol{\theta}$, doing ≈ 20 dB better than MU-GAMP.

V. Conclusions

In this paper, we propose a matrix-uncertainty (MU) extension of the GAMP algorithm, as well as an alternating A-MU-GAMP that aims to recover both the signal and the unknown (possibly parametric) measurement matrix. We also

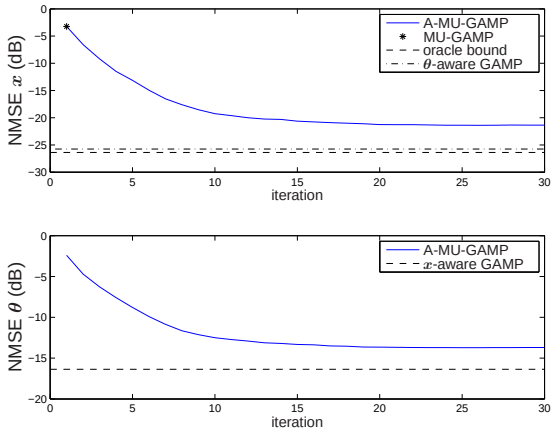

Fig. 5. 100-trial median NMSE of A-MU-GAMP when iteratively estimating $\boldsymbol{x}$ and $\boldsymbol{\theta}$ in the compressive blind deconvolution example.

provide theoretical and empirical evidence of the following surprising fact: as the dimensions grow large, the effect of *uniform* matrix uncertainty reduces to AWGN, and can thus be handled by matrix-certain algorithms. Our MU-GAMP approach can, however, exploit knowledge of *non*-uniform matrix uncertainty to do significantly better. Moreover, our A-MU-GAMP approach, which exploits soft information (as opposed to point estimates), achieves near-oracle performance. In future work, we plan to investigate the application of A-MU-GAMP to spectral estimation, dictionary learning, matrix completion, and robust principle components analysis (PCA).

References

[1] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus-Mathématique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[2] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, 2010.

[3] M. Herman and D. Needell, "Mixed operators in compressed sensing," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*. IEEE, 2010, pp. 1–6.

[4] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[5] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2002 –2016, May 2011.

[6] M. Rosenbaum and A. Tsybakov, "Sparse recovery under matrix uncertainty," *The Annals of Statistics*, vol. 38, no. 5, pp. 2620–2651, 2010.

[7] E. Candes and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[8] D. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[9] S. Rangan, "Generalized Approximate Message Passing for Estimation with Random Linear Mixing," *Arxiv preprint arXiv:1010.5141*, 2010.

[10] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[11] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2479 –2493, july 2009.

[12] I. Tošić and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.