THE AIR FORCE RESEARCH LABORATORY
LEAD | DISCOVER | DEVELOP | DELIVER

Compressive Sensing under Matrix Uncertainties: An Approximate Message Passing Approach

Asilomar 2011

Jason T. Parker (AFRL/RYAP)
Philip Schniter (OSU)
Volkan Cevher (EPFL)

THE OHIO STATE UNIVERSITY

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Problem Statement

- Traditional Compressive Sensing (CS) addresses underdetermined linear regression

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}$$

$$\boldsymbol{y}, \boldsymbol{w} \in \mathbb{C}^M; \quad \boldsymbol{A} \in \mathbb{C}^{M \times N}; \quad \boldsymbol{x} \in \mathbb{C}^N; \quad M < N$$

- More generally, consider an unknown matrix perturbation $\boldsymbol{E} \in \mathbb{C}^{M \times N}$

$$\boldsymbol{y} = \underbrace{\left(\hat{\boldsymbol{A}} + \boldsymbol{E}\right)}_{\text{unknown } \boldsymbol{A}} \boldsymbol{x} + \boldsymbol{w}$$

- We characterize $\boldsymbol{A} = \hat{\boldsymbol{A}} + \boldsymbol{E}$ with entry-wise means and variances given by

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

$$\nu^A_{mn} = \mathrm{var}\{A_{mn}\}$$

Previous Work

- Notice that

$$\begin{aligned} \boldsymbol{y} &= \left(\hat{\boldsymbol{A}} + \boldsymbol{E}\right)\boldsymbol{x} + \boldsymbol{w} \\ &= \hat{\boldsymbol{A}}\boldsymbol{x} + \underbrace{(\boldsymbol{E}\boldsymbol{x} + \boldsymbol{w})}_{\text{signal dependent noise}} \end{aligned}$$

- Standard CS performance analysis for bounded $\boldsymbol{E}$ [1; 2]
- LASSO → Sparsity-Cognizant Total Least Squares [3]

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

- Dantzig Selector → Matrix Uncertain Selector [4]

$$\hat{\boldsymbol{x}}_{\text{MU-Selector}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{ subject to } \|\hat{\boldsymbol{A}}^H\left(\boldsymbol{y} - \hat{\boldsymbol{A}}\boldsymbol{x}\right)\|_\infty \leq \lambda\|\boldsymbol{x}\|_1 + \epsilon$$

Generalized Approximate Message Passing

- Approximate Message Passing (AMP) [5] is derived from (approximate) belief propagation

$$\hat{\boldsymbol{x}}^{k+1} = \eta_s\left(\hat{\boldsymbol{x}}^k + \boldsymbol{A}^H\boldsymbol{z}^k, \beta_k\theta_k\right),$$
$$\boldsymbol{z}^k = \boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}^k + b^k\boldsymbol{z}^{k-1}$$

- η_s as soft-thresholding $\longrightarrow$ near minimax performance (robust)
- η_s distribution specific $\longrightarrow$ approximate MMSE inference
- Generalized AMP (GAMP) [6; 7]
    - MMSE or MAP estimates of $\boldsymbol{x} \in \mathbb{C}^N$, $p(\boldsymbol{x}) = \prod_n p_X(x_n)$
    - Arbitrary separable output channel from *noiseless* measurements $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{C}^M$ to $\boldsymbol{y}$, $p(\boldsymbol{y} \mid \boldsymbol{z}) = \prod_m p_{Y|Z}(y_m \mid z_m)$
    - Handles variable $|A_{mn}|$
    - Provides approximate posteriors

Matrix Uncertain GAMP

- Recall noise-free measurements are $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$.
- For large N, the Central Limit Theorem motivates treating $z_m \,|\, x_n$ as Gaussian
- Using the zero mean quantities $\tilde{A}_{mn} \triangleq A_{mn} - \hat{A}_{mn}$ and $\tilde{x}_{mn} \triangleq x_{mn} - \hat{x}_{mn}$, we can write

$$z_m = (\hat{A}_{mn} + \tilde{A}_{mn})x_n + \sum_{r \neq n} (\hat{A}_{mr}\hat{x}_r + \hat{A}_{mr}\tilde{x}_r + \tilde{A}_{mr}\hat{x}_r + \tilde{A}_{mr}\tilde{x}_r)$$

- From which we can conclude

$$\mathrm{E}\{z_m \,|\, x_n\} = \hat{A}_{mn} x_m + \sum_{r \neq n} \hat{A}_{mr}\hat{x}_{mr}$$

$$\mathrm{var}\{z_m \,|\, x_n\} = \nu^A_{mn}|x_n|^2 + \sum_{r \neq n} \hat{A}^2_{mr}\nu^x_{mr} + \nu^A_{mr}|\hat{x}_{mr}|^2 + \nu^A_{mr}\nu^x_{mr}$$

- Terms in red modify the original GAMP variance calculation

MU-GAMP Algorithm Summary

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

$$\begin{aligned}
\forall m : \hat{z}_m(t) &= \textstyle\sum_{n=1}^{N} \hat{A}_{mn}\hat{x}_n(t) && \text{(R1)}\\
\forall m : \nu^z_m(t) &= \textstyle\sum_{n=1}^{N} |\hat{A}_{mn}|^2 \nu^x_n(t) && \text{(R2a)}\\
\forall m : \nu^p_m(t) &= \nu^z_m(t) + \textstyle\sum_{n=1}^{N} \nu^A_{mn} \left(\nu^x_n + |\hat{x}_n(t)|^2\right) && \text{(R2b)}\\
\forall m : \hat{p}_m(t) &= \hat{z}_m(t) - \nu^z_m(t)\, \hat{u}_m(t-1) && \text{(R3)}\\
\forall m : \hat{u}_m(t) &= g_{\text{out}}(y_m, \hat{p}_m(t), \nu^p_m(t)) && \text{(R4)}\\
\forall m : \nu^u_m(t) &= -g'_{\text{out}}(y_m, \hat{p}_m(t), \nu^p_m(t)) && \text{(R5)}\\
\forall n : \nu^r_n(t) &= \left(\textstyle\sum_{n=1}^{N} |\hat{A}_{mn}|^2 \nu^u_m(t)\right)^{-1} && \text{(R6)}\\
\forall n : \hat{r}_n(t) &= \hat{x}_n(t) + \nu^r_n(t) \textstyle\sum_{m=1}^{M} \hat{A}^*_{mn}\hat{u}_m(t) && \text{(R7)}\\
\forall n : \nu^x_n(t{+}1) &= \nu^r_n(t) g'_{\text{in}}(\hat{r}_n(t), \nu^r_j(t)) && \text{(R8)}\\
\forall n : \hat{x}_n(t{+}1) &= g_{\text{in}}(\hat{r}_n(t), \nu^r_n(t)) && \text{(R9)}
\end{aligned}$$

end

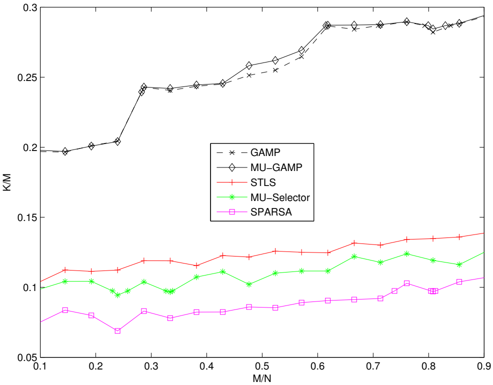
Independent Identically Distributed Matrix Errors

- Consider i.i.d. matrix errors with $\nu^A_{mn} = \nu^A$.
- For additive noise, a CLT argument suggests that, for large N, we can well approximate

$$p(\boldsymbol{y} \,|\, \boldsymbol{x}) \sim \mathcal{N}(\hat{\boldsymbol{A}}\boldsymbol{x}, \nu^A \|\boldsymbol{x}\|_2^2 + \nu^w)$$

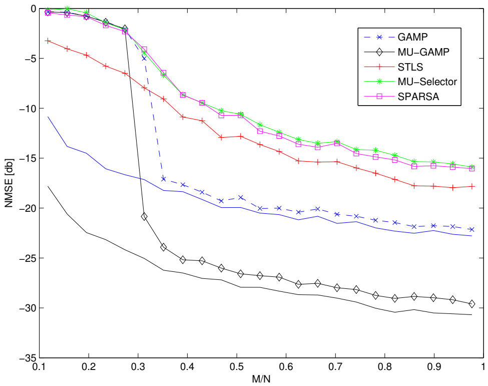
- Law of large numbers $\rightarrow$ $\|\boldsymbol{x}\|_2^2 \approx$ constant for large N
- Conclusion: i.i.d. matrix uncertainty can be addressed by tuning standard algorithms for large N

Phase Transition - i.i.d. Matrix Errors



- $N = 256$; $\hat{\boldsymbol{A}}$ is i.i.d. $\mathcal{N}(0,1)$; $\nu^A = 0.05$
- $\boldsymbol{x} \sim$ Bernoulli-Radamacher (± 1 non-zero entries)
- Gaussian additive noise at 20 dB SNR. Effective SNR is about 12 dB
- LASSO (using SPARSA), STLS, and MU-Selector parameters use genie-aided tuning
- GAMP uses genie-aided computation of effective noise variance
- Curves show $-15dB$ NMSE contours based on median from 100 trials

NMSE vs M/N for Sparse Matrix Errors



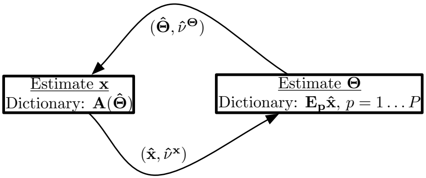
- Same setup, except that the entries of $\boldsymbol{E}$ are now Bernoulli-Radamacher with 99% zeroes and $\nu^A = 5$ for the non-zeroes.
- MU-GAMP is given the true entries ν^A_{mn} while GAMP is given only the true effective noise variance.
- The solid lines are linear estimates given the true support of $\boldsymbol{x}$ using $\hat{\boldsymbol{A}}$ (blue) and $\hat{\boldsymbol{A}} + \boldsymbol{E}$ (black)
- Naive versions of STLS and MU-Selector are used with genie-aided tuning. The parametric STLS or "compensated" MU-Selector would likely show improved performance.

Parametric MU-GAMP

Parametric Model

- Consider $\boldsymbol{\Theta} \in \mathbb{C}^P$ an unknown parameter vector
$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{\Theta})\boldsymbol{x} + \boldsymbol{w},$$

- We employ a first order Taylor series expansion, similar to parametric STLS [3]
$$\boldsymbol{y} \approx \left(\boldsymbol{A}(\hat{\boldsymbol{\Theta}}) + \sum_{p=1}^{P} (\Theta_p - \hat{\Theta}_p) \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}}) \right) \boldsymbol{x} + \boldsymbol{w}$$
$$\boldsymbol{E}_p(\hat{\boldsymbol{\Theta}}) \triangleq \frac{\partial \boldsymbol{A}(\boldsymbol{\alpha})}{\partial \alpha_p}\Big|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\Theta}}}$$



Parametric MU-GAMP: Compute $\hat{x}$

Data Model

$$\boldsymbol{y} \approx \left(\boldsymbol{A}(\hat{\boldsymbol{\Theta}}) + \sum_{p=1}^{P} (\Theta_p - \hat{\Theta}_p) \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}}) \right) \boldsymbol{x} + \boldsymbol{w}$$

$$\boldsymbol{E}_p(\hat{\boldsymbol{\Theta}}) \triangleq \frac{\partial \boldsymbol{A}(\boldsymbol{\alpha})}{\partial \alpha_p} \Big|_{\boldsymbol{\alpha} = \hat{\boldsymbol{\Theta}}}$$

- First, assume we have an estimate of the parameter as $(\hat{\boldsymbol{\Theta}}, \boldsymbol{\nu}^{\boldsymbol{\Theta}})$
- We can immediately write

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{w}$$

$$\boldsymbol{C} \triangleq \boldsymbol{A}(\hat{\boldsymbol{\Theta}}) + \sum_{p=1}^{P} (\Theta_p - \hat{\Theta}_p) \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}})$$

$$\hat{\boldsymbol{c}} \triangleq \mathrm{E}\{\boldsymbol{C}\} = \boldsymbol{A}(\hat{\boldsymbol{\Theta}})$$

$$\boldsymbol{\nu}^{c} \triangleq \mathrm{var}\{\boldsymbol{C}\} = \sum_{p=1}^{P} \nu_p^{\Theta} |\boldsymbol{E}_p|^2,$$

  where squares on matrix terms are understood to be element-wise squared magnitudes. In addition, the mean and variance of the matrix are interpreted element-wise.
- We can use MU-GAMP to compute an estimate $(\hat{\boldsymbol{x}}, \boldsymbol{\nu}^{\boldsymbol{x}})$ from this model.

Parametric MU-GAMP: Compute $\hat{\boldsymbol{\Theta}}$

Alternate form

$$\left(\boldsymbol{A}(\hat{\boldsymbol{\Theta}}) + \sum_{p=1}^{P} (\Theta_p - \hat{\Theta}_p)\boldsymbol{E}_p(\hat{\boldsymbol{\Theta}})\right)\boldsymbol{x} =$$

$$\left(\sum_{p=1}^{P} \Theta_p \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}})\right)\boldsymbol{x} + \underbrace{\left(\boldsymbol{A}(\hat{\boldsymbol{\Theta}}) - \sum_{p=1}^{P} \hat{\Theta}_p \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}})\right)\hat{\boldsymbol{x}}}_{\text{known constant}} + \underbrace{\left(\boldsymbol{A}(\hat{\boldsymbol{\Theta}}) - \sum_{p=1}^{P} \hat{\Theta}_p \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}})\right)\tilde{\boldsymbol{x}}}_{\text{zero-mean}}$$

- We can leverage this expression to obtain a linear model for $\boldsymbol{\Theta}$ with a known dictionary $\boldsymbol{B}$.

  $\boldsymbol{u} = \boldsymbol{B}\boldsymbol{\Theta} + \boldsymbol{n}$ We can estimate $\boldsymbol{\Theta}$ from this model using MU-GAMP!

$$\boldsymbol{u} \triangleq \boldsymbol{y} - \left(\boldsymbol{A}(\hat{\boldsymbol{\Theta}}) - \sum_{p=1}^{P} \hat{\Theta}_p \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}})\right)\hat{\boldsymbol{x}}; \qquad \boldsymbol{n} \triangleq \boldsymbol{w} + \left(\boldsymbol{A}(\hat{\boldsymbol{\Theta}}) - \sum_{p=1}^{P} \hat{\Theta}_p \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}})\right)\tilde{\boldsymbol{x}}$$

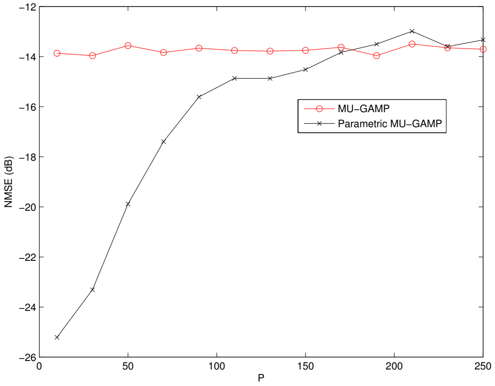
$$\mathrm{E}\{\boldsymbol{n}\} = \boldsymbol{0}; \qquad \mathrm{var}\{\boldsymbol{n}\} = \boldsymbol{\nu}^w + \left|\boldsymbol{A}(\hat{\boldsymbol{\Theta}}) - \sum_{p=1}^{P} \hat{\Theta}_p \boldsymbol{E}_p(\hat{\boldsymbol{\Theta}})\right|^2 \boldsymbol{\nu}^{\boldsymbol{x}}$$

$$\boldsymbol{B} \triangleq \left[\begin{array}{cccc} \boldsymbol{E}_1(\hat{\boldsymbol{\Theta}})\boldsymbol{x} & \boldsymbol{E}_2(\hat{\boldsymbol{\Theta}})\boldsymbol{x} & \ldots & \boldsymbol{E}_P(\hat{\boldsymbol{\Theta}})\boldsymbol{x} \end{array}\right]$$

$$\hat{\boldsymbol{b}} \triangleq \mathrm{E}\{\boldsymbol{B}\} = \left[\begin{array}{cccc} \boldsymbol{E}_1(\hat{\boldsymbol{\Theta}})\hat{\boldsymbol{x}} & \boldsymbol{E}_2(\hat{\boldsymbol{\Theta}})\hat{\boldsymbol{x}} & \ldots & \boldsymbol{E}_P(\hat{\boldsymbol{\Theta}})\hat{\boldsymbol{x}} \end{array}\right]$$

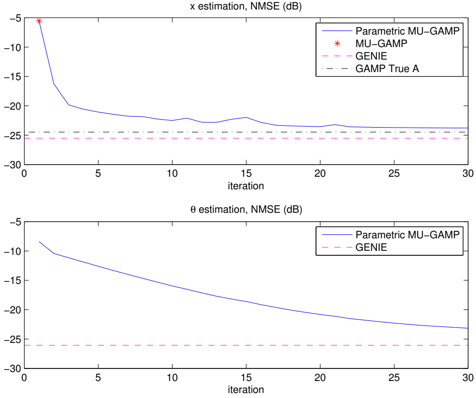
$$\boldsymbol{\nu}^b \triangleq \mathrm{var}\{\boldsymbol{B}\} = \left[\begin{array}{cccc} |\boldsymbol{E}_1(\hat{\boldsymbol{\Theta}})|^2\boldsymbol{\nu}^{\boldsymbol{x}} & |\boldsymbol{E}_2(\hat{\boldsymbol{\Theta}})|^2\boldsymbol{\nu}^{\boldsymbol{x}} & \ldots & |\boldsymbol{E}_P(\hat{\boldsymbol{\Theta}})|^2\boldsymbol{\nu}^{\boldsymbol{x}} \end{array}\right]$$

Example 1: NMSE vs Parameter Dimension



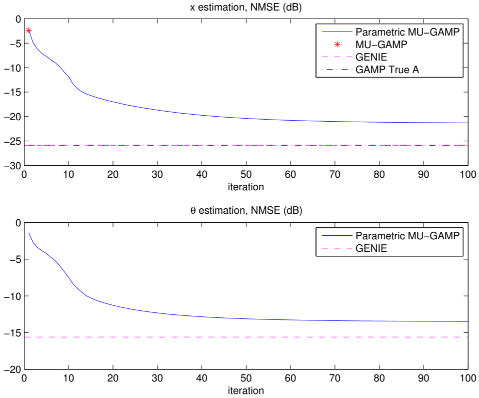
- In this toy example ($N = 256$, $M = 103$),
$$\boldsymbol{A} = \boldsymbol{A}_0 + \sum_{p=1}^{P} \theta_p \boldsymbol{E}_p$$
- $\boldsymbol{A}_0$ and all the $\boldsymbol{E}_p$ have entries that are drawn i.i.d. Gaussian.
- As we vary P, the entries of the $\boldsymbol{E}_p$ matrices are scaled to keep $\mathrm{E}\{\nu^A_{mn}\} = \nu^A; \forall m, n$.
- MU-GAMP is given the true element-wise variances, but is unaware of the underlying $\boldsymbol{\Theta}$ structure.
- Parametric MU-GAMP leverages this underlying structure to improve performance for small P

Example 2: Joint Calibration and Recovery



- $N = 256$; $M = 103$; $K = 20$; $P = 10$
- Each $\boldsymbol{E}_p$ contains ones for M/P of the rows, and zeros elsewhere. This model is a surrogate for channel calibration errors in a measurement system.
- The unknown $\boldsymbol{x}$ is Bernoulli-Gaussian, while $\boldsymbol{\Theta}$ is Gaussian.
- In this example, all signals are complex-valued.
- The "GENIE" result for $\boldsymbol{x}$ assumes perfect knowledge of $\boldsymbol{\Theta}$ and vice-versa. The "GENIE" also knows the signal support when applicable.

Example 3: Blind Deconvolution



- We consider here the model $\boldsymbol{Y} = \boldsymbol{\Psi A}(\boldsymbol{\Theta})\boldsymbol{X}$
- $\boldsymbol{A}(\boldsymbol{\Theta}) \in \mathbb{C}^{N \times N}$ is circulant, and $\boldsymbol{\Theta} \in \mathbb{C}^{N}$ represents perturbations to the first column.
- $\boldsymbol{Y} \in \mathbb{C}^{M \times S}$; $\boldsymbol{X} \in \mathbb{C}^{N \times S}$
- $\boldsymbol{\Psi} \in \mathbb{C}^{M \times N}$ is a mixing matrix.
- $N = 256$; $M = 103$; $K = 20$; $S = 8$
- Each $\boldsymbol{E}_p$ represents the change to the system response for a given coefficient of the impulse response. This model is a surrogate for learning a system impulse response from S snapshots, where each signal is K sparse.
- The unknown $\boldsymbol{x}$ is complex Bernoulli-Gaussian, while $\boldsymbol{\Theta}$ is complex Gaussian.
- “GENIE” estimators same as previous example

Conclusions and Future Work

- We have developed a Matrix Uncertain version of GAMP
  - Adaptively adjusts noise power for i.i.d. matrix errors
  - Incorporate element-wise variances for independent, non-identical errors
  - Iterative approach for parametric matrix uncertainty
- Future Work
  - MU-GAMP for spectral estimation
  - Parametric MU-GAMP for dictionary learning and matrix completion
  - Extension of rigorous GAMP performance analysis to MU-GAMP case.
  - Incorporation of MU-GAMP into EM tuning approach to learn hyper-parameters from data

Bibliography

[1] Y. Chi, A. Pezeshki, L. Scharf, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," in *Proc. IEEE Int Acoustics Speech and Signal Processing (ICASSP) Conf*, pp. 3930–3933, 2010.

[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] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *Signal Processing, IEEE Transactions on*, vol. 59, pp. 2002 –2016, May 2011.

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

[5] 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. 18914–18919, 2009.

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

[7] P. Schniter, "Turbo Reconstruction of Structured Sparse Signals," *ICASSP*, 2010.

Questions?