An Empirical-Bayes Approach to Recovering Linearly Constrained Non-Negative Sparse Signals

Jeremy Vila and Philip Schniter

THE OHIO STATE UNIVERSITY

Supported in part by NSF grant CCF-1018368, by NSF grant 1218754, by NSF-I/UCRC grant IIP-0968910, and by DARPA/ONR grant N66001-10-1-4090.

Problem Statement

- **Goal**: Recover a $K$-sparse signal $x \in \mathbb{R}^N$ from $M \ll N$ noisy linear measurements

  $$y = Ax + w \in \mathbb{R}^M,$$

  where $A$ is a known sensing matrix and $w$ is noise.

- We focus on non-negative (NN) signals (i.e., $x_n \geq 0 \ \forall n$) that obey linear equality constraints $Bx = c$ such as the simplex

  $$\Delta^N_+ \triangleq \{x \in \mathbb{R}^N : x_n \geq 0 \ \forall n, \ 1^T x = 1\}.$$

- A common approach is to solve the $\ell_1$-regularized constrained optimization problem, i.e.,

  $$\hat{x} = \arg \min_{x \geq 0} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1 \ \text{s.t.} \ \ Bx = c. \quad (1)$$

  - **The Good**: Convex problem, with nearly minimax solutions.
  - **The Bad**: Performance is relatively poor for non-worst-case $x$.
  - **The Ugly**: Expensive to optimally tune the penalty $\lambda$. 

Contributions

- A novel algorithm that solves the optimization problem (1) efficiently.

- A novel approach for online tuning of the penalty parameter $\lambda$ in (1) and related problems (e.g., LASSO).

- A novel algorithm for approximate MMSE inference of $\mathbf{x}$ via powerful Gaussian-Mixture priors whose parameters are learned online.

- Robust extensions for outlier corrupted observations $\mathbf{y}$.

- Applied to real-world applications, such as sparse NN compressive imaging, portfolio optimization, and hyperspectral image inversion.
Proposed Approach

- We use the generalized approximate message passing (GAMP) algorithm with suitable choice of signal and noise priors.
- To enforce non-negativity, select signal priors with NN support.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Type</th>
<th>Signal Prior</th>
<th>EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN Least Squares AMP (NNLS-AMP)</td>
<td>MAP</td>
<td>NN Uniform</td>
<td>$\lambda = 0$</td>
</tr>
<tr>
<td>EM-NN LASSO-AMP (EM-NNL-AMP)</td>
<td>MAP</td>
<td>Exponential</td>
<td>Yes</td>
</tr>
<tr>
<td>EM-NN Gauss Mix-AMP (EM-NNGM-AMP)</td>
<td>MMSE</td>
<td>NNGM</td>
<td>Yes</td>
</tr>
</tbody>
</table>

- We use expectation-maximization (EM) for online parameter tuning.

![GAMP Algorithm Diagram](image-url)
Enforcing Linear Equality Constraints

- Linear equality constraints are enforced via the augmented model
  \[
  \begin{bmatrix}
  y \\
  c
  \end{bmatrix}
  =
  \begin{bmatrix}
  A \\
  B
  \end{bmatrix}
  x
  +
  \begin{bmatrix}
  w \\
  0
  \end{bmatrix}.
  \]

- Defining \( \bar{y} \triangleq [y_c] \), \( \bar{A} \triangleq [A_B] \), and \( \bar{z} \triangleq \bar{A}x \), the likelihood function corresponding to the augmented measurement model is
  \[
  p_{Y_m|Z_m}(y_m|z_m) = \begin{cases} 
  p_{Y|Z}(y_m|z_m) & m = 1, \ldots, M \\
  \delta(y_m - z_m) & m = M + 1, \ldots, M + P.
  \end{cases}
  \]

- For robustness to outliers in \( y \) we use additive white Laplacian noise (AWLN).
GAMP Summary

- GAMP is a computationally efficient approach to MAP or approximate MMSE inference of \( x \in \mathbb{R}^N \) that exploits:
  - known separable signal prior \( p_x(x) = \prod_n p_x(x_n) \),
  - known separable \( p_{y|z}(y|z) = \prod_m p_{y|z}(y_m|z_m) \), where \( z_m \triangleq a_m^T x \) are the noiseless transform outputs,
  - sufficiently large, dense, and random \( A \).

- GAMP is derived from Taylor-series and central-limit-theorem based approximations of loopy belief propagation.

- GAMP admits rigorous analysis in the large-system limit (i.e., \( M, N \to \infty \) for fixed ratio \( M/N \)) for i.i.d zero-mean sub-Gaussian \( A \).
The EM algorithm is guaranteed to converge to a **local maximum** of the likelihood $p(y; q)$ over the set of prior parameters $q$ by iteratively maximizing the lower bound

$$Q_{\hat{p}}(y; q) \triangleq E_{\hat{p}}\{\ln p_{x,y}(x, y; q)\} + H(\hat{p}) \leq \ln p(y; q).$$

**E step** For iteration $i$, set

$$\hat{p}^i = p_{x|y}(\cdot|y; q^i).$$

**M step** Maximizing $Q_{\hat{p}^i}(y; q)$ is difficult since exact $p_{x|y}(\cdot|y; q^i)$ is unavailable, so we instead use GAMP’s approximated posteriors.

Parameter updates for the 3 signal and 2 noise distributions can be computed in **closed-form** using quantities already supplied by GAMP!
Empirical Phase Transitions: Non-Negative Signals

- Fixed $N = 500$, constructed grid on $\frac{M}{N}$-versus-$\frac{K}{M}$ plane, set $\mathbf{A} \sim \mathcal{N}(0, M^{-1})$, and drew the $K$-sparse elements of $\mathbf{x} \sim \text{Dir}(a)$.

- Empirical (noiseless) PTCs for each algorithm segment the the regions where it can perfectly recover the signal (below) and where it cannot (above).

- EM-NNGM-AMP had the best PTCs, which greatly improved upon the theoretical $\ell_1$ recovery PTC!

- EM-NNL-AMP matches the theoretical $\ell_1$ recovery PTC, despite not knowing $\lambda$!
Sparse Non-negative Compressive Imaging

- **Goal**: Recover satellite image from compressive measurements at SNR = 60 dB.

- Leverage **fast structurally random** \( A = \Phi \Psi S \), where \( \Phi \) is a random \( M \times N \) row selector, \( \Psi \) is the \( N \times N \) Hadamard transform, and \( S \) is diagonal with random \( \pm 1 \)s.

- EM-NNGM-AMP had the best phase transition and recovery NMSE over all \( \frac{M}{N} \), and the second-fastest runtime.

- EM-NNL-AMP matched NMSE of TFOCS with genie-supplied \( \lambda \) (showing EM tuning works), but had much faster runtime, in fact the fastest of all techniques.
Portfolio Optimization (Fama and Fench 49 dataset)

- **Goal:** Divide capital among the $N$ securities to maximize returns and minimize risk.

- Using the previous 120 months of return data $A$, the average training returns $\mu$, and the expected return $\rho$, and the sparse Markowitz mean-variance framework solves

$$\hat{x} = \arg \min_{x \in \Delta_N^+} \| \rho 1 - Ax \|_2^2 \quad \text{s.t.} \quad \mu^T x = \rho.$$ 

- Average **Sharpe ratio** (reward vs. risk statistic) and constraint error $\mathcal{E}$ measured across the future returns.

- All algorithms outperform the naïve strategy, i.e., $x = \frac{1}{N}1$.

- EM-NNGM-AMP with AWLN likelihood gave highest Sharpe ratio.

- The matrix $A$ was highly correlated! Damping was critical for GAMP.

<table>
<thead>
<tr>
<th></th>
<th>SR</th>
<th>time (sec)</th>
<th>$\mathcal{E}$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>naïve</td>
<td>0.3135</td>
<td>-</td>
<td>$-\infty$</td>
</tr>
<tr>
<td>lsqmin</td>
<td>0.3725</td>
<td><strong>0.06</strong></td>
<td>-307.4</td>
</tr>
<tr>
<td>CV-TFOCS</td>
<td>0.3747</td>
<td>31.92</td>
<td>-56.9</td>
</tr>
<tr>
<td>NNLS-AMP</td>
<td>0.3724</td>
<td>0.74</td>
<td>-75.2</td>
</tr>
<tr>
<td>EM-NNL-AMP</td>
<td>0.3724</td>
<td>0.69</td>
<td>-75.4</td>
</tr>
<tr>
<td>EM-NNGM-AMP</td>
<td>0.3869</td>
<td>5.21</td>
<td>-59.8</td>
</tr>
<tr>
<td>1WGN-NNLS-AMP</td>
<td>0.3826</td>
<td>1.65</td>
<td>-61.0</td>
</tr>
<tr>
<td>1WGN-EM-NNL-AMP</td>
<td>0.3823</td>
<td>1.89</td>
<td>-55.0</td>
</tr>
<tr>
<td>1WGN-EM-NNGM-AMP</td>
<td><strong>0.4038</strong></td>
<td>6.12</td>
<td>-51.7</td>
</tr>
</tbody>
</table>
**Hyperspectral Image Inversion (SHARE 2012 dataset)**

**Goal**: Decompose a spectral dataset measured over $M$ spectral bands and $T = T_1 \times T_2$ pixels, into $N$ material endmember spectra and spatial abundances.

- **NNLS-AMP**, fully constrained least squares (FCLS), and GSSP performed “identically.”

- **EM-NNGM-AMP** recovers the “purest” abundances.

$N = 4$ material abundance maps.
Conclusions

- Proposed a suite of algorithms for recovering linearly constrained non-negative signals.
  - EM-NNL-AMP performs $\ell_1$ optimization while online tuning $\lambda$.
  - EM-NNGM-AMP performs MMSE inference via powerful GM prior & online learning.

- For robustness to outliers, developed Laplacian noise model with online parameter tuning.

- EM-NNGM-AMP yielded the best phase transitions and lowest MSE on all tested experiments, due to its powerful prior and ability to perform online parameter learning.

- Our proposed algorithms are much faster than competitors like TFOCS.
**Related Publications**


