Compressive Phase Retrieval via Generalized Approximate Message Passing

Philip Schniter

Joint work with Sundeep Rangan.

Supported in part by NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090.

Allerton – Oct. 3, 2012
Phase Retrieval

- **Goal:** Recover signal $x_0 \in \mathbb{C}^n$ from $m$ magnitude-only measurements
  
  $$y = |Ax_0 + w|,$$
  
  where $A \in \mathbb{C}^{m \times n}$ is a known linear transform and $w \in \mathbb{C}^m$ is noise.

- **Motivation:** In many applications, it feasible to measure the intensity, but not the phase, of the Fourier transform of the signal-of-interest:
  - X-ray crystallography,
  - transmission electron microscopy,
  - coherent diffractive imaging,
  - astronomical imaging, etc.

- **Feasibility:** To make the solution to $y = |Ax|$ unique (up to a global phase) w.p.1, $m=3n-2$ i.i.d Gaussian measurements are necessary [Finkelson'04] and $m=4n-2$ are sufficient [Balan/Casazza/Edidin'06].
Phase Retrieval: Classical Approaches

Most classical approaches are **iterative** in nature. For example,

- Alternate between...
  - projecting $A\hat{x}$ onto the magnitude constraint $y$, yielding $\hat{z}$,
  - projecting $A^+\hat{z}$ onto an apriori known support set, yielding $\hat{x}$.

However, due to the non-convexity of the first projection, it is easy for such algorithms to get trapped in **local minima**.
Phase Retrieval: Convex Approaches

Recently, some convex relaxations have been proposed.

- Noting that $y_i^2 = |a_i^H x|^2 = \text{tr}(a_i a_i^H X)$ for $X = xx^H$, pose as “min $X \succeq 0 \ \text{rank}(X)$ s.t. $\text{tr}(a_i a_i^H X) = y_i^2$ for $i = 1...m$.” (NP hard!)

Relax to “min $\text{tr}(X)$ s.t. $\text{tr}(a_i a_i^H X) = y_i^2$ for $i = 1...m$,” (convex!) known as PhaseLift [Candes/Eldar/Strohmer/Voroninski’11].

- Another semidefinite program (with similar performance) known as PhaseCut was proposed in [Waldspurger/D’Aspremont/Mallat’12].

It was recently shown [Candes/Li’12] that

- with very high probability, PhaseLift perfectly recovers an arbitrary $x$ from $m \geq c_0 n$ noiseless measurements, where $c_0$ is a constant,
- and PhaseLift can be made robust to noise.
Recall that $m \geq 3n - 2$ magnitude measurements are needed for $y = |Ax|$ to have a unique solution for $x \in \mathbb{C}^n$.

Sometimes we can only afford $m < 3n - 2$ magnitude measurements, in which case the problem becomes one of compressive phase retrieval.

For successful compressive phase retrieval (CPR), one needs to leverage additional structure in $x$, such as sparsity.
Compressive Phase Retrieval: Prior Work

- Assuming knowledge of $\|x_0\|_1$, [Moravec/Romberg/Baraniuk'07]
  - appended this constraint onto the classical RAAR algorithm, and
  - used RIP-based arguments to establish that $m \gtrsim k^2 \log(n/k^2)$ magnitude measurements suffice for recovery.

However, the algorithm was prone to local minima and slow convergence. Also, knowledge of $\|x_0\|_1$ is rarely available in practice.

- Taking a convex approach, [Ohlsson/Yang/Dong/Sastry'12] proposed the following generalization of PhaseLift, which they call CPRL:
  $$\min_{X \succeq 0} \text{tr}(X) + \lambda \|X\|_1 + \mu \sum_{i=1}^m \left| \text{tr}(a_i a_i^H X) - y_i^2 \right|^2 \quad \text{for } i = 1 \ldots m,$$
  and performed both RIP and mutual coherence analyses. Seems promising...
Bring out the GAMP

Zed: Bring out the Gimp.

Maynard: Gimp’s sleeping.

Zed: Well, I guess you’re gonna have to go wake him up now, won’t you?

—Pulp Fiction, 1994.

We propose a new approach to CPR based on generalized approximate message passing (GAMP).

Numerical results show

- excellent phase transitions,
- excellent NMSE & robustness to noise,
- excellent runtime,

with direct application to compressive image retrieval.
The evolution of GAMP:

- The original AMP [Donoho/Maleki/Montanari’09] solves the LASSO problem \( \min_x \|y - Ax\|_2^2 + \lambda \|x\|_1 \) popular in compressive sensing, i.e., MAP estimation under i.i.d Laplacian signal and AWGN.
- The Bayesian AMP [Donoho/Maleki/Montanari’10] extended the above to generic i.i.d signal priors and MMSE estimation.
- The generalized AMP [Rangan’10] extended the above to generic i.i.d likelihood models of the form \( p_{Y|Z}(y_i|a_i^H x) \).

In the end, GAMP produces a sophisticated iterative thresholding alg, whose complexity is dominated by one application of \( A \) and \( A^H \) per iteration with relatively few (e.g., tens) iterations. Very fast!

- Rigorous large-system analyses (under i.i.d Gaussian \( A \)) have established that (G)AMP follows a state-evolution trajectory with optimal properties [Bayati/Montanari’10], [Rangan’10].
GAMP Heuristics (Sum-Product)

1. Message from $y_i$ node to $x_j$ node:
   \[
   p_{i\rightarrow j}(x_j) \propto \int p_{Y|Z}(y_i; \sum_r a_{ir} x_r, \psi) \prod_{r \neq j} p_{i\leftarrow r}(x_r) \approx \mathcal{N} \text{ via CLT}
   \]
   \[
   \approx \int_{z_i} p_{Y|Z}(y_i; z_i, \psi) \mathcal{N}(z_i; \hat{z}_i(x_j), \nu_i^z(x_j)) \approx \mathcal{N}
   \]
   To compute $\hat{z}_i(x_j), \nu_i^z(x_j)$, the means and variances of $\{p_{i\leftarrow r}\}_{r \neq j}$ suffice, thus Gaussian message passing!

Remaining problem: we have $2mn$ messages to compute (too many!).

2. Exploiting similarity among the messages $\{p_{i\leftarrow j}\}_{i=1}^m$, GAMP employs a Taylor-series approximation of their difference, whose error vanishes as $m \to \infty$ for dense $A$ (and similar for $\{p_{i\rightarrow j}\}_{j=1}^n$ as $n \to \infty$).

Finally, need to compute only $O(m+n)$ messages!
To apply GAMP, we need an appropriate likelihood function $p_{Y|Z}(y_i|z_i)$, where r.v. $Y$ represents the noisy magnitude measurements $y_i$ and r.v. $Z$ represents the corresponding noiseless transform outputs $z_i \triangleq a_i^H x$.

For this, we assume the statistical model

$$y_i = e^{j\theta_i}(z_i + w_i) \quad \text{with} \quad \theta_i \in \mathcal{U}[0, 2\pi) \quad \text{and} \quad w_i \sim \mathcal{CN}(0, \nu^w),$$

from which we margin out $\theta_i$ and $w_i$ to obtain

$$p_{Y|Z}(y_i|z_i) = \frac{1}{\pi \nu^w} e^{-\frac{(|y_i| - |z_i|)^2}{\nu^w}} I_0(\rho) e^{-\rho} \quad \text{for} \quad \rho \triangleq \frac{2|y_i| |z_i|}{\nu^w},$$

where $I_0(\cdot)$ is the 0th-order modified Bessel function of the first kind.

See paper for other algorithmic details.
Numerical Results

For our numerical results we generated

- the signal $x_0$ as $k$-sparse Bernoulli-circular-Gaussian,
- the matrix as $A = \Phi F$ where $\Phi \in \mathbb{C}^{m \times n}$ is i.i.d circular Gaussian and $F$ is the $n \times n$ DFT matrix,
- the (pre-magnitude) noise $w$ as circular white Gaussian,

and we monitored the phase-corrected normalized reconstruction MSE

$$\text{NMSE} \triangleq \min_{\theta} \frac{\|\hat{x} - e^{i\theta} x_0\|^2}{\|x_0\|^2}.$$
Phase transition

PR-GAMP’s empirical success rate, averaged over 500 realizations, was

where \( \text{success} \triangleq \{ \text{NMSE} < 10^{-4} \} \).
Comparison to phase-oracle

Comparing the 50%-success contours of PR- and phase-oracle GAMP:

we see that PR-GAMP requires about $4\times$ the number of measurements.
Noise Robustness of PR-GAMP

The median NMSE, measured over 2000 realizations:

![Graph showing noise robustness of PR-GAMP]

shows that PR-GAMP loses about 3 dB at medium-to-high SNR.
Compressive Image Recovery

65536 image pixels, 32768 measurements, 30dB SNR:

original image

PR-GAMP (-29.7dB NMSE)

PR-GAMP runtime: only 11.1 sec.
Comparison to CPRL [Ohlsson/Yang/Dong/Sastry’12]

Empirical success rate (and runtime) on two toy problems:

|         | \((m, n) = (20, 32)\) | \((m, n) = (30, 48)\) | \((m, n) = (40, 64)\) |
|---------|------------------------|------------------------|------------------------|
| \(k = 1\): | CPRL 0.96 (4.9 sec)   | CPRL 0.97 (51 sec)    | CPRL 0.99 (291 sec)   |
|         | PR-GAMP 0.83 (0.4 sec) | PR-GAMP 0.94 (0.3 sec) | PR-GAMP 0.99 (0.3 sec) |
| \(k = 2\): | CPRL 0.55 (5.8 sec)   | CPRL 0.55 (58 sec)    | CPRL 0.58 (316 sec)   |
|         | PR-GAMP 0.72 (0.4 sec) | PR-GAMP 0.92 (0.3 sec) | PR-GAMP 1.0 (0.3 sec)  |

Notice:

- CPRL works great with sparsity \(k = 1\), but poorly when \(k \geq 2\). GAMP instead suffers when problem dimensions are small.
- CPRL’s runtime grows very quickly with problem dimensions! GAMP’s runtime is negligible for these toy problems.
Conclusions

- **(Compressive) phase retrieval** is a longstanding problem that is experiencing a rebirth through compressive sensing and convex relaxation.

- We proposed a new approach to CPR based on **generalized approximate message passing (GAMP)**.

- Empirical results show an **excellent phase transition** ($4 \times$ meas of phase-oracle), **excellent noise robustness** ($\sim 3$ dB worse than phase-oracle), and **excellent runtime** (many orders of magnitude faster than convex relaxation).

- As a practical demonstration, we accurately recovered a **64k-pixel image** from **32k measurements** in only **11 seconds**.