Robust and Efficient Computational Methods for Phase Retrieval

Aditya Viswanathan
aditya@math.msu.edu

MICHIGAN STATE UNIVERSITY

ICERM Research Cluster
Computational Challenges in Sparse and Redundant Representations
$5^{\mathtt{th}}$ November 2014

Joint work with

Yang Wang

Mark Iwen

Research supported in part by National Science Foundation grant DMS 1043034.

Outline

The Phase Retrieval Problem

$$\text{find} \quad \mathbf{x} \in \mathbb{C}^d \quad \text{given} \quad |M\mathbf{x}| = \mathbf{b} \in \mathbb{R}^D,$$

where

- $\mathbf{b} \in \mathbb{R}^D$ are the magnitude or intensity measurements.
- $M \in \mathbb{C}^{D \times d}$ is a measurement matrix associated with these measurements.

Let $\mathcal{A} : \mathbb{R}^D \to \mathbb{C}^d$ denote the recovery method.

The phase retrieval problem involves designing measurement matrix and recovery method pairs.

Applications of Phase Retrieval

Important applications of Phase Retrieval

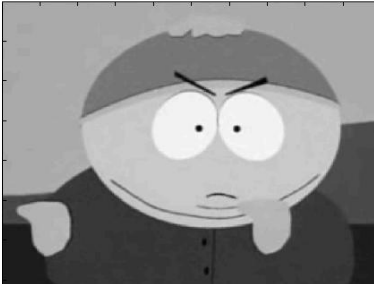
- X-ray crystallography
- Diffraction imaging
- Transmission Electron Microscopy (TEM)

In many molecular imaging applications, the detectors only capture intensity measurements.

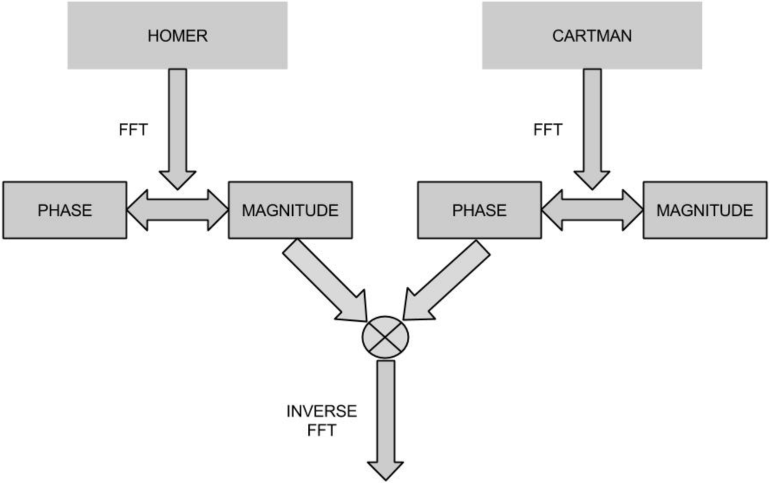
Indeed, the design of such detectors is often significantly simpler than those that capture phase information.

The Importance of Phase – An Illustration

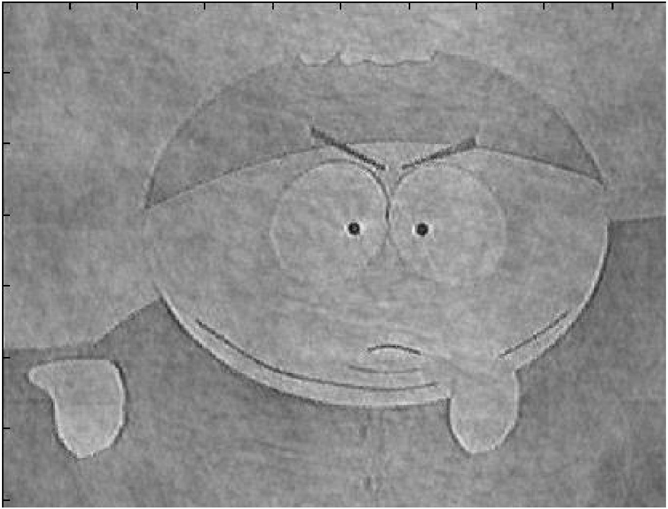
- The phase encapsulates vital information about a signal
- Key features of the signal are retained even if the magnitude is lost



The Importance of Phase – An Illustration



The Importance of Phase – An Illustration



Objectives

- Computational Efficiency – Can the recovery algorithm $\mathcal{A}$ be computed in $O(d \log d)$-time?
- Computational Robustness: The recovery algorithm, $\mathcal{A}$, should be robust to additive measurement errors (i.e., noise).
- Minimal Measurements: The number of linear measurements, D, should be minimized to the greatest extent possible.

Outline

Alternating Projection Methods

[Gerchberg and Saxton, 1972] and [Fienup, 1978]

- These methods operate by alternately projecting the current iterate of the signal estimate over two sets of constraints.
- One of the constraints is the magnitude of the measurements.
- The other constraint depends on the application – positivity, support constraints, . . .

Alternating Projection Methods

[Gerchberg and Saxton, 1972] and [Fienup, 1978]

Algorithm 1 Gerchberg–Saxton

Input: Measurements $\mathbf{b} = |M\mathbf{x}| \in \mathbb{R}^D$, Initial estimate $\mathbf{x}_0 \in \mathbb{C}^d$.

1: **for** $i = 0$ to $N - 1$ **do**
2: Compute $\mathbf{y} = M\mathbf{x}_i$
3: Set $\tilde{\mathbf{y}} = \mathbf{b} \angle \mathbf{y}$
4: Compute $\mathbf{x}_{i+1} = M^{\dagger}\tilde{\mathbf{y}}$
5: **end for**

- N is the number of iterations
- $M^{\dagger}$ is the Moore-Penrose pseudo-inverse

Alternating Projection Methods

[Gerchberg and Saxton, 1972] and [Fienup, 1978]

Issues

- Convergence is slow – the algorithm is likely to stagnate at stationary points
- Requires careful selection of and tuning of the parameters
- Mathematical aspects of the algorithm not well known. If there is proof of convergence, it is only for special cases.

Applications

- Can be used as a post-processing step to speed up more rigorous (but slow) computational approaches

PhaseLift [Candes et. al., 2012]

- Modify the problem to that of finding the rank-1 matrix $X = \mathbf{x}\mathbf{x}^*$
- Uses multiple random illuminations (or masks) as measurements.
- The resulting problem can be cast as a rank minimization optimization problem (`NP hard`)
- Instead, solve a convex relaxation – trace minimization problem (`SDP`)

PhaseLift [Candes et. al., 2012]

Notation

- Let $\mathbf{w}^m$ be a mask. The measurements may be written as

$$\begin{aligned}|\langle \mathbf{w}^m, \mathbf{x}\rangle|^2 = \mathsf{Tr}(\mathbf{x}^*\mathbf{w}^m(\mathbf{w}^m)^*\mathbf{x}) &= \mathsf{Tr}(\mathbf{w}^m(\mathbf{w}^m)^*\mathbf{x}\mathbf{x}^*)\\ &:= \mathsf{Tr}(W^m X).\end{aligned}$$

- Let $\mathcal{W}$ be the linear operator mapping positive semidefinite matrices into $\{\mathsf{Tr}(W^m X) : m = 0, \ldots, L\}$.
- The phase retrieval problem then becomes

$$\begin{aligned}\text{minimize} \quad & \mathsf{rank}(X)\\ \text{subject to} \quad & \begin{array}{l}\mathcal{W}(X) = b\\ X \succeq 0\end{array}\end{aligned}$$

PhaseLift [Candes et. al., 2012]

Notation

- Let $\mathbf{w}^m$ be a mask. The measurements may be written as

$$|\langle \mathbf{w}^m, \mathbf{x} \rangle|^2 = \mathsf{Tr}(\mathbf{x}^* \mathbf{w}^m (\mathbf{w}^m)^* \mathbf{x}) = \mathsf{Tr}(\mathbf{w}^m (\mathbf{w}^m)^* \mathbf{x}\mathbf{x}^*) \\ := \mathsf{Tr}(W^m X).$$

- Let $\mathcal{W}$ be the linear operator mapping positive semidefinite matrices into $\{\mathsf{Tr}(W^m X) : k = 0, \ldots, L\}$.
- The phase retrieval problem then becomes

$$\begin{array}{ll} \text{minimize} & \text{rank}(X) \\ \text{subject to} & \begin{array}{l} \mathcal{W}(X) = b \\ X \succeq 0 \end{array} \end{array}$$

- Unfortunately, this problem is NP hard!

PhaseLift [Candes et. al., 2012]

Notation

- Let $\mathbf{w}^m$ be a mask. The measurements may be written as

$$\begin{aligned} |\langle \mathbf{w}^m, \mathbf{x} \rangle|^2 = \mathsf{Tr}(\mathbf{x}^* \mathbf{w}^m (\mathbf{w}^m)^* \mathbf{x}) &= \mathsf{Tr}(\mathbf{w}^m (\mathbf{w}^m)^* \mathbf{x}\mathbf{x}^*) \\ &:= \mathsf{Tr}(W^m X). \end{aligned}$$

- Let $\mathcal{W}$ be the linear operator mapping positive semidefinite matrices into $\{\mathsf{Tr}(W^m X) : k = 0, \ldots, L\}$.
- Instead, use the convex relaxation

$$\begin{array}{ll} \text{minimize} & \text{trace}(X) \\ \text{subject to} & \begin{array}{l} \mathcal{W}(X) = b \\ X \succeq 0 \end{array} \end{array}$$

- Implemented using a semidefinite program (SDP).

PhaseLift [Candes et. al., 2012]

- Consider noisy measurements $\mathbf{b} = |M\mathbf{x}|^2 + \mathbf{n}, \quad \mathbf{b}, \mathbf{n} \in \mathbb{R}^D$.
- Let $\Phi_M : \mathbb{R}^D \to \mathbb{C}^d$ denote the PhaseLift recovery algorithm.

Theorem (Candes, Li 2014)

Let $M \in \mathbb{C}^{D\times d}$ have its D rows be independently drawn either uniformly at random from the sphere of radius $\sqrt{d}$ in $\mathbb{C}^d$, or else as complex normal random vectors from $\mathcal{N}(0, \mathcal{I}_d/2) + \mathbb{i}\mathcal{N}(0, \mathcal{I}_d/2)$. Then, $\exists$ universal constants $\tilde{B}, \tilde{C}, \tilde{A} \in \mathbb{R}^+$ such that the PhaseLift procedure $\Phi_M : \mathbb{R}^D \to \mathbb{C}^d$ satisfies

$$\min_{\theta\in[0,2\pi]} \left\| \Phi_M(\mathbf{b}) - \mathbb{e}^{\mathbb{i}\theta}\mathbf{x} \right\|_2 \leq \tilde{C} \cdot \min\left(\|\mathbf{x}\|_2, \frac{\|\mathbf{n}\|_1}{D\|\mathbf{x}\|_2} \right)$$

for all $\mathbf{x} \in \mathbb{C}^d$ with probability $1 - \mathcal{O}(\mathbb{e}^{-\tilde{B}D})$, provided that $D \geq \tilde{A}d$.

PhaseLift [Candes et. al., 2012]

Advantages

- Recovery guarantees for random measurements
- Optimization problems of the above type are well-understood
- Mature software for solving the resulting optimization problem.

Disadvantages

- SDP solvers are still slow!
- General-purpose solvers have complexity $\mathcal{O}\left(d^3\right)$; FFT-based measurements may be solved in $\mathcal{O}\left(d^2\right)$ time

Other Approaches

- Phase Retrieval with Polarization [Alexeev et. al. 2014]
  - Graph-theoretic frame-based approach
  - Requires $\mathcal{O}(d \log d)$ measurements
  - Error guarantee similar to PhaseLift

- Phase Recovery, MaxCut and Complex Semidefinite Programming [Waldspurger et. al. 2013]
  - Related to graph partitioning problems
  - Can be shown to be equivalent to PhaseLift under certain conditions
  - Requires solving a SDP

Outline

Overview of the Computational Framework

1. Use compactly supported masks and correlation measurements to obtain phase difference estimates.

$$|\texttt{corr}(\mathbf{w},\mathbf{x})|^2 \xrightarrow[\text{linear system}]{\text{solve}} x_j\overline{x}_{j+k}, \quad k = 0,\ldots,\delta$$

   - $\mathbf{w}$ is a mask or window function with $\delta + 1$ non-zero entries.
   - $x_j\overline{x}_{j+k}$ gives us the (scaled) difference in phase between entries x_j and x_{j+k}.

2. Solve an angular synchronization problem on the phase differences to obtain the unknown signal.

$$x_j\overline{x}_{j+k} \xrightarrow[\text{synchronization}]{\text{angular}} x_j$$

Constraints on $\mathbf{x}$: We require $\mathbf{x}$ to be non-sparse.
(The number of consecutive zeros in $\mathbf{x}$ should be less than δ)

Overview of the Computational Framework

1. Use compactly supported masks and correlation measurements to obtain phase difference estimates.

$$|\mathtt{corr}(\mathbf{w}, \mathbf{x})|^2 \xrightarrow[\text{linear system}]{\text{solve}} x_j \overline{x}_{j+k}, \quad k = 0, \ldots, \delta$$

   - $\mathbf{w}$ is a mask or window function with $\delta + 1$ non-zero entries.
   - $x_j \overline{x}_{j+k}$ gives us the (scaled) difference in phase between entries x_j and x_{j+k}.

2. Solve an angular synchronization problem on the phase differences to obtain the unknown signal.

$$x_j \overline{x}_{j+k} \xrightarrow[\text{synchronization}]{\text{angular}} x_j$$

Constraints on $\mathbf{x}$: We require $\mathbf{x}$ to be non-sparse.
(The number of consecutive zeros in $\mathbf{x}$ should be less than δ)

Correlations with Support-Limited Functions

- Let $\mathbf{x} = [x_0 \; x_1 \; \ldots \; x_{d-1}]^T \in \mathbb{C}^d$ be the unknown signal.

- Let $\mathbf{w} = [w_0 \; w_1 \; \ldots \; w_\delta \; 0 \; \ldots 0]^T$ denote a support-limited mask. It has $\delta + 1$ non-zero entries.

- We are given the (squared) correlation measurements

$$(b^m)^2 = |\mathtt{corr}(\mathbf{w}^m, \mathbf{x})|^2 , \qquad m = 0, \ldots, L$$

corresponding to $L + 1$ distinct masks.

Correlations with Support-Limited Functions

Explicitly writing out the correlation, we have

$$
\begin{aligned}
(b_k^m)^2 &= \left| \sum_{j=0}^{\delta} \overline{w}_j^m \cdot x_{k+j} \right|^2 \\
&= \sum_{i,j=0}^{\delta} w_i^m \overline{w}_j^m x_{k+j} \overline{x}_{k+i} \\
&=: \sum_{i,j=0}^{\delta} w_i \overline{w}_j Z_{k+j,i-j},
\end{aligned}
$$

where $Z_{n,l} := x_n \overline{x}_{n+l}, \quad -\delta \leq l \leq \delta.$

Solving for Phase Differences

Ordering $\{Z_{n,l}\}$ lexicographically, second index first, we obtain a linear system of equations for the phase differences.

Example: $\mathbf{x} \in \mathbb{R}^d, d = 4, \delta = 1$

$$
\begin{array}{c}
\begin{array}{cccc} \quad M_0' \quad & & \quad M_1' \quad & \end{array} \\
\underbrace{\begin{pmatrix}
\boxed{\begin{matrix} (w_0^0)^2 & 2w_0^0w_1^0 \\ (w_0^1)^2 & 2w_0^1w_1^1 \end{matrix}} & \boxed{\begin{matrix} (w_1^0)^2 & 0 \\ (w_1^1)^2 & 0 \end{matrix}} & \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \\
\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} (w_0^0)^2 & 2w_0^0w_1^0 \\ (w_0^1)^2 & 2w_0^1w_1^1 \end{matrix} & \begin{matrix} (w_1^0)^2 & 0 & 0 & 0 \\ (w_1^1)^2 & 0 & 0 & 0 \end{matrix} \\
\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} (w_0^0)^2 & 2w_0^0w_1^0 & (w_1^0)^2 & 0 \\ (w_0^1)^2 & 2w_0^1w_1^1 & (w_1^1)^2 & 0 \end{matrix} \\
\begin{matrix} (w_1^0)^2 & 0 \\ (w_1^1)^2 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 & 2w_0^0w_1^0 & (w_1^0)^2 \\ 0 & 0 & 2w_0^1w_1^1 & (w_1^1)^2 \end{matrix}
\end{pmatrix}}_{M'}
\underbrace{\begin{pmatrix} Z_{0,0} \\ Z_{0,1} \\ Z_{1,0} \\ Z_{1,1} \\ Z_{2,0} \\ Z_{2,1} \\ Z_{3,0} \\ Z_{3,1} \end{pmatrix}}_{\mathbf{y}}
=
\begin{pmatrix} b_0^0 \\ b_0^1 \\ b_1^0 \\ b_1^1 \\ b_2^0 \\ b_2^1 \\ b_3^0 \\ b_3^1 \end{pmatrix}
\end{array}
$$

The system matrix M' is *block circulant*!

Consequences of the Block Circulant Structure

- There exists a unitary decomposition of the system matrix
- The condition number of the system matrix is a function of the individual blocks. In particular, we have

$$\kappa(M') = \frac{\max\limits_{|t|=1} \ \sigma_1(t)}{\min\limits_{|t|=1} \ \sigma_\delta(t)},$$

  where $\sigma_1(t), \sigma_\delta(t)$ are the largest/smallest singular values of

$$J(t) = M_0' + tM_1' + \cdots + t^{\delta-1} M_\delta'$$

- The linear system for the phase differences can be solved efficiently using FFTs

Entries of the Measurement Matrix

Two strategies

- Random entries (Gaussian, Uniform, Bernoulli, . . .)
- Structured measurements

Random Measurements

Representative Condition Numbers

	$\delta = 3$	$\delta = 6$	$\delta = 9$
Critical sampling	34.89	124.05	465.32
Oversampling factor 2	3.73	11.73	13.30
Oversampling factor 3	3.29	6.08	8.60

- Critical sampling: $D = (2\delta + 1)d$, where D denotes the number of measurements.
- Oversampling: $D = \gamma \cdot (2\delta + 1)d$, where γ is the oversampling factor
- Typically, we use $\gamma = 1.5$.

Structured Measurements

Theorem (Iwen, V., Wang 2014)

Choose entries of the measurement mask $\mathbf{w}^m$ *as follows:*

$$w_i^\ell = \begin{cases} \frac{\mathbb{e}^{-i/a}}{\sqrt[4]{2\delta+1}} \cdot \mathbb{e}^{\frac{2\pi \mathbb{i} \cdot i \cdot \ell}{2\delta+1}}, & i \le \delta \\ 0, & i > \delta \end{cases} \qquad a \in \left[4, \frac{\delta - 1}{2}\right), \quad 0 \le \ell \le L.$$

Then, the resulting system matrix for the phase differences, M'*, has condition number*

$$\kappa(M') < \max\left\{144\mathbb{e}^2, \frac{9\mathbb{e}^2}{4} \cdot (\delta - 1)^2\right\}.$$

Note:

- w_i^ℓ are scaled entries of a DFT matrix.
- δ is typically chosen to be 6–12.
- No oversampling necessary!

Angular Synchronization

The Angular Synchronization Problem

Estimate d unknown angles $\theta_1, \theta_2, \ldots, \theta_d \in [0, 2\pi)$ from $\delta + 1$ noisy measurements of their differences $\theta_{ij} := \theta_i - \theta_j \mod 2\pi$.

Recall: $\angle Z_{i,j} = \angle x_i - \angle x_{i+j}$.

Angular Synchronization

The unknown phases (modulo a global phase offset) may be obtained by solving a simple greedy algorithm.

1. Set the largest magnitude component to have zero phase angle; i.e.,

$$\angle x_j = 0, \qquad j = \operatorname*{argmax}_i Z_{i,0}.$$

   Note: (by definition) $|x_i|^2 = Z_{i,0}, \qquad i = 0, \ldots, d-1.$

2. Use this entry to set the phase angles of the next δ entries; i.e.,

$$\angle x_k = \angle x_j - \angle Z_{j,k}, \qquad k = 1, \ldots, \delta.$$

3. Use the next largest magnitude component from these δ entries and repeat the process.

Recovering Arbitrary Vectors

- Recall: Due to compact support of our masks, only "flat" vectors can be recovered
- Arbitrary vectors can be "flattened" by multiplication with a random unitary matrix such as $W = PFB$, where
  - $P \in \{0,1\}^{d\times d}$ is a permutation matrix selected uniformly at random from the set of all $d \times d$ permutation matrices
  - F is the unitary $d \times d$ discrete Fourier transform matrix
  - $B \in \{-1,0,1\}^{d\times d}$ is a random diagonal matrix with i.i.d. symmetric Bernoulli entries on its diagonal

Runtime Analysis

- For flat vectors, total runtime complexity is

$$\mathcal{O}\left(d \cdot \delta^3 + \delta \cdot d \log d\right)$$

  - $\mathcal{O}(\delta \cdot d \log d)$ – FFT evaluations required to evaluate the unitary decomposition of system matrix M'
  - $\mathcal{O}(d \cdot \delta^3)$ – invert blocks in the matrix M'.

- For arbitrary vectors, using the flattening matrix W discussed previously, total runtime complexity is

$$\mathcal{O}\left(d \cdot \log^3(d) \cdot \log^3(\log d)\right)$$

Outline

Numerical Results

- Test signals (Real and Complex) – iid Gaussian, uniform random, sinusoidal signals
- Noise model

$$\tilde{\mathbf{b}} = \mathbf{b} + \tilde{\mathbf{n}}, \quad \tilde{\mathbf{n}} \sim U[0, a].$$

  Value of a determines SNR

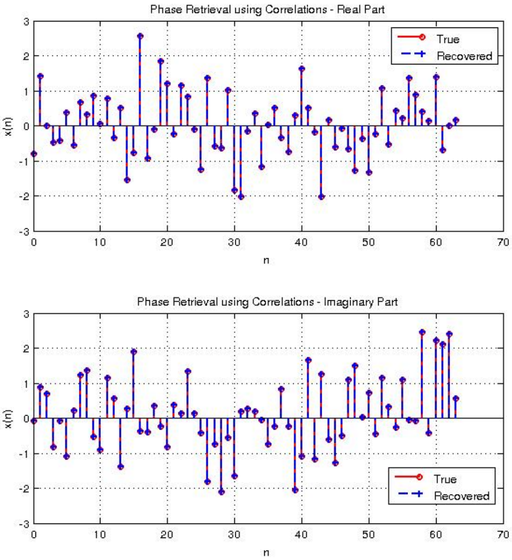
$$SNR = 10 \log_{10} \left(\frac{\text{noise power}}{\text{signal power}} \right) = 10 \log_{10} \left(\frac{a^2/3}{||b||^2/d} \right)$$

- Errors reported as SNR (dB)

$$\text{Error (dB)} = 10 \log_{10} \left(\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|^2}{\|\mathbf{x}\|^2} \right)$$

  ($\hat{\mathbf{x}}$ – recovered signal, $\mathbf{x}$ – true signal)

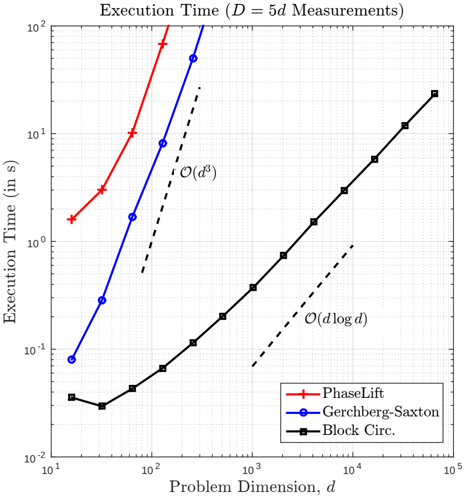
Noiseless Case



- iid Complex Gaussian signal
- $d = 64$
- $\delta = 1$
  ($3d$ measurements)
- No noise
- Reconstruction Error

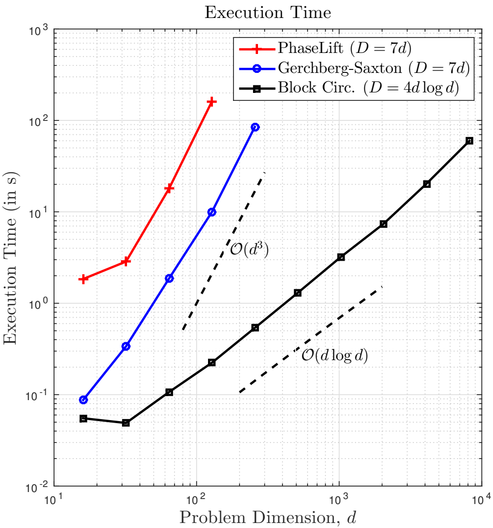
$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|^2}{\|\mathbf{x}\|^2} = 6.436 \times 10^{-15}$$

Efficiency



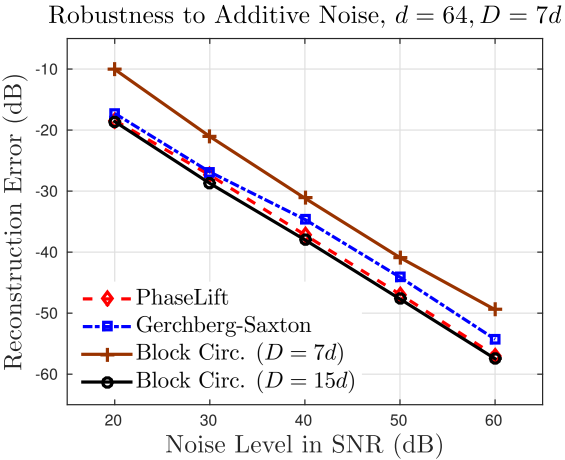
- iid Complex Gaussian signal
- High SNR applications
- $5d$ measurements
- $64k$ problem in ~ 20 s in Matlab!

Efficiency



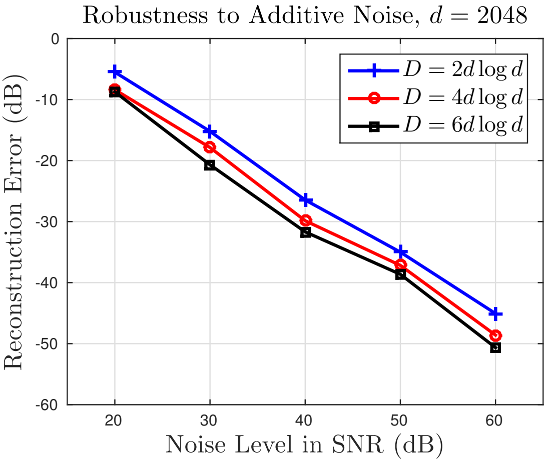
- iid Complex Gaussian signal
- Generic applications (wide range of SNRs)
- $4d \log d$ measurements

Robustness



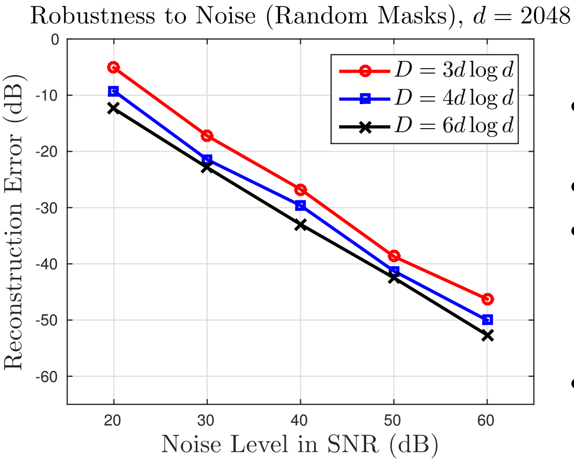
- iid complex Gaussian signal
- $d = 64$
- $7d$ measurements
- Deterministic (windowed Fourier-like) measurements

Robustness



- iid complex Gaussian signal
- $d = 2048$
- Not feasible with PhasLift or Alternating projection methods on a laptop in Matlab
- Deterministic (windowed Fourier-like) measurements

Robustness



- iid complex Gaussian signal
- $d = 2048$
- Not feasible with PhasLift or Alternating projection methods on a laptop in Matlab
- Random measurements

Discussion

(+)

- Well-conditioned deterministic measurement matrices with explicit condition number bounds
- Significantly faster (FFT time) than comparable SDP-based methods

(-)

- Requires $2\times$ to $4\times$ more measurements than equivalent SDP-based methods

Outline

The *Sparse* Phase Retrieval Problem

$$\text{find} \quad \mathbf{x} \in \mathbb{C}^d \quad \text{given} \quad |\mathcal{M}\mathbf{x}| = \mathbf{b} \in \mathbb{R}^D,$$

where

- $\mathbf{x}$ is k-sparse, with $k \ll d$.
- $\mathbf{b} \in \mathbb{R}^D$ are the magnitude or intensity measurements.
- $\mathcal{M} \in \mathbb{C}^{D \times d}$ is a measurement matrix associated with these measurements.

Let $\mathcal{A} : \mathbb{R}^D \to \mathbb{C}^d$ denote the recovery method.

The sparse phase retrieval problem involves designing measurement matrix and recovery method pairs.

Sparse Phase Retrieval – Objectives

- Computational Efficiency
- Computational Robustness: The recovery algorithm, $\mathcal{A}$, should be robust to additive measurement errors (i.e., noise).
- Minimal Measurements: The number of linear measurements, D, should be minimized to the greatest extent possible. In particular, can we have robust reconstruction for $D = O(k \log(d/k))$ measurements?

Recall – Alternating Projection Methods

[Gerchberg and Saxton, 1972] and [Fienup, 1978]

Algorithm 1 Gerchberg–Saxton

Input: Measurements $\mathbf{b} = |M\mathbf{x}| \in \mathbb{R}^D$, Initial estimate $\mathbf{x}_0 \in \mathbb{C}^d$.

1: **for** $i = 0$ to $N - 1$ **do**
2: Compute $\mathbf{y} = M\mathbf{x}_i$
3: Set $\tilde{\mathbf{y}} = \mathbf{b} \angle\mathbf{y}$
4: Compute $\mathbf{x}_{i+1} = M^\dagger \tilde{\mathbf{y}}$
5: **end for**

- N is the number of iterations
- $M^\dagger$ is the Moore-Penrose pseudo-inverse

Alternating Projections with Sparsity Constraints

[Mukherjee and Seelamantula, 2012]

- Assume sparsity in some basis.
- Introduce a thresholding step where you retain only k coefficients of $\mathbf{x}$ in this basis representation.
- Similar performance as with Gerchberg–Saxton.

Recall – PhaseLift [Candes et. al., 2012]

- Let $\mathbf{w}^m$ be a mask. The measurements may be written as

$$\begin{aligned} |\langle \mathbf{w}^m, \mathbf{x} \rangle|^2 = \mathsf{Tr}(\mathbf{x}^* \mathbf{w}^m (\mathbf{w}^m)^* \mathbf{x}) &= \mathsf{Tr}(\mathbf{w}^m (\mathbf{w}^m)^* \mathbf{x}\mathbf{x}^*) \\ &:= \mathsf{Tr}(W^m X). \end{aligned}$$

- Let $\mathcal{W}$ be the linear operator mapping positive semidefinite matrices into $\{\mathsf{Tr}(W^m X) : k = 0, \ldots, L\}$.
- Instead, use the convex relaxation

$$\begin{array}{ll} \text{minimize} & \text{trace}(X) \\ \text{subject to} & \begin{array}{l} \mathcal{W}(X) = b \\ X \succeq 0 \end{array} \end{array}$$

- Implemented using a semidefinite program (SDP).

Extensions to Sparse Signals [Ohlsson et. al., 2012]

Compressive Phase Retrieval via Lifting (CPRL)

$$\begin{array}{ll} \text{minimize} & \text{trace}(X) + \lambda ||X||_1 \\ \text{subject to} & \begin{array}{l} \mathcal{W}(X) = b \\ X \succeq 0. \end{array} \end{array}$$

- λ is a regularization parameter.
- Can be shown [Moravec et. al., 2007] that $\mathbf{x}$ can be uniquely recovered if $d = O\left(k^2 \log(4D/k^2)\right)$.

Other Methods

- GrEedy Sparse PhAse Retrieval (GESPAR) algorithm [Shechtman et. al. 2014]
  - Fast 2-opt local search applied to sparsity constrained non-linear optimization
  - Works with magnitude measurements of any linear transform (including Fourier)
  - No theoretical error guarantees
  - No. of required measurements scales as k^3

- Compressive Phase Retrieval via Generalized Approximate Message Passing [Schniter, Rangan 2014]
  - Probabilistic, graph-based message passing techniques
  - Only numerical study; no theoretical error guarantees
  - No. of measurements required: $D \geq 2k \log_2(N/k)$
  - Efficient implementation

Computational Framework

Let the measurement matrix $\mathcal{M}$ be of the form

$$\mathcal{M} = \mathcal{P}\mathcal{C},$$

where

- $\mathcal{P} \in \mathbb{C}^{D \times \tilde{d}}$ is an admissible phase retrieval matrix, and
- $\mathcal{C} \in \mathbb{C}^{\tilde{d} \times d}$ is an admissible compressive sensing matrix.

Note: We typically have $D = O(\tilde{d})$ and $\tilde{d} = O\left(k \log(d/k)\right)$, where k is the sparsity of $\mathbf{x}$.

Computational Framework

1. Solve a (non-sparse) phase retrieval problem

$$\begin{array}{ll} \text{minimize} & \text{trace}(Y) \\ \text{subject to} & \begin{array}{l} \mathcal{W}(Y) = b \\ Y \succeq 0, \end{array} \end{array}$$

   where $Y = \mathbf{y}\mathbf{y}^*$ and $\mathbf{y} \in \mathbb{C}^{\tilde{d}}$ is an intermediate solution.

2. Recover $\mathbf{x}$ using a compressive sensing formulation

$$\begin{array}{ll} \text{minimize} & ||\mathbf{x}||_1 \\ \text{subject to} & \mathcal{C}\mathbf{x} = b. \end{array}$$

Computational Framework

1. Solve a (non-sparse) phase retrieval problem

$$\begin{array}{ll} \text{minimize} & \text{trace}(Y) \\ \text{subject to} & \begin{array}{l} \mathcal{W}(Y) = b \\ Y \succeq 0, \end{array} \end{array}$$

where $Y = \mathbf{y}\mathbf{y}^*$ and $\mathbf{y} \in \mathbb{C}^{\tilde{d}}$ is an intermediate solution.

2. Recover $\mathbf{x}$ using a compressive sensing formulation

$$\begin{array}{ll} \text{minimize} & ||\mathbf{x}||_1 \\ \text{subject to} & \mathcal{C}\mathbf{x} = b. \end{array}$$

Advantages

- Dramatically reduces the problem dimension.
  - In CPRL, the single SDP is in d^2 variables, with D equality constraints.
  - In the 2-stage formulation, the first SDP is in $\tilde{d}^2 = O(k^2 \log^2 d)$ variables, with D equality constraints. The second optimization program is much simpler (LP).
- Recovery guarantees follow naturally from guarantees for the Phase Retrieval method employed and Compressive Sensing
- Not limited to Phase Lift – Can work with any phase retrieval method (Step 1)

Error Guarantee

- Consider noisy measurements of the form

$$\mathbf{b} := |\mathcal{P}\mathcal{C}\mathbf{x}|^2 + \mathbf{n}$$

- $\mathcal{P} \in \mathbb{C}^{D \times \tilde{d}}$ is any phase retrieval matrix with an associated recovery algorithm $\Phi_{\mathcal{P}} : \mathbb{R}^D \to \mathbb{C}^{\tilde{d}}$ (and error guarantee)
- $\mathcal{C} \in \mathbb{C}^{\tilde{d} \times d}$ is any compressive sensing matrix with an associated recovery algorithm $\Delta_{\mathcal{C}} : \mathbb{C}^{\tilde{d}} \to \mathbb{C}^d$ (and error guarantee)
- Composition of the two recovery algorithms, $\Delta_{\mathcal{C}} \circ \Phi_{\mathcal{P}} : \mathbb{R}^D \to \mathbb{C}^d$, should accurately approximate $\mathbf{x} \in \mathbb{C}^d$, up to a global phase factor, from $\mathbf{b}$ whenever $\mathbf{x}$ is sufficiently sparse or compressible.

Error Guarantee

Theorem (PhaseLift: Candes, Li 2014)

Let $\mathcal{P} \in \mathbb{C}^{D \times \tilde{d}}$ have its D rows be independently drawn either uniformly at random from the sphere of radius $\sqrt{\tilde{d}}$ in $\mathbb{C}^{\tilde{d}}$, or else as complex normal random vectors from $\mathcal{N}(0, \mathcal{I}_{\tilde{d}}/2) + \mathbb{i}\mathcal{N}(0, \mathcal{I}_{\tilde{d}}/2)$. Then, $\exists$ universal constants $\tilde{B}, \tilde{C}, \tilde{A} \in \mathbb{R}^+$ such that the PhaseLift procedure $\Phi_{\mathcal{P}} : \mathbb{R}^D \to \mathbb{C}^{\tilde{d}}$ satisfies

$$\min_{\theta \in [0, 2\pi]} \left\| \, \Phi_{\mathcal{P}}\left(\mathbf{b}\right) - \mathbb{e}^{\mathbb{i}\theta}\mathbf{x} \, \right\|_2 \leq \tilde{C} \cdot \min\left(\|\mathbf{x}\|_2, \frac{\|\mathbf{n}\|_1}{D\|\mathbf{x}\|_2} \right)$$

for all $\mathbf{x} \in \mathbb{C}^{\tilde{d}}$ with probability $1 - \mathcal{O}(\mathbb{e}^{-\tilde{B}D})$, provided that $D \geq \tilde{A}\tilde{d}$.

Error Guarantee

Theorem (from Foucart, Rauhut Theorem 4.22)

Suppose that the matrix $\mathcal{C} \in \mathbb{C}^{\tilde{d} \times d}$ satisfies the ℓ_2-robust null space property of order k with constants $0 < \rho < 1$ and $\tau > 0$. Then, for any $\mathbf{x} \in \mathbb{C}^d$, the vector

$$\tilde{\mathbf{x}} := \underset{\mathbf{z} \in \mathbb{C}^d}{\arg\min} \ \|\mathbf{z}\|_1 \quad \textit{subject to} \quad \|\mathcal{C}\mathbf{z} - \mathbf{y}\|_2 \leq \eta,$$

where $\mathbf{y} := \mathcal{C}\mathbf{x} + \mathbf{e}$ for some $\mathbf{e} \in \mathbb{C}^{\tilde{d}}$ with $\|\mathbf{e}\|_2 \leq \eta$, will satisfy

$$\|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq \frac{C}{\sqrt{k}} \cdot \left(\inf_{\mathbf{z} \in \mathbb{C}^d, \|\mathbf{z}\|_0 \leq k} \|\mathbf{x} - \mathbf{z}\|_1 \right) + A\eta$$

for some constants $C, A \in \mathbb{R}^+$ that only depend on ρ and τ.

Error Guarantee

Theorem (Iwen, V., Wang 2014)

Let $\mathcal{P} \in \mathbb{C}^{D\times\tilde{d}}$ have its D rows be independently drawn either uniformly at random from the sphere of radius $\sqrt{\tilde{d}}$ in $\mathbb{C}^{\tilde{d}}$, or else as complex normal random vectors from $\mathcal{CN}(0, \mathcal{I}_{\tilde{d}})$.
Furthermore, suppose that $\mathcal{C} \in \mathbb{C}^{\tilde{d}\times d}$ satisfies the ℓ_2-robust null space property of order k with constants $0 < \rho < 1$ and $\tau > 0$. Then,

$$\min_{\theta\in[0,2\pi]} \left\| \mathbb{e}^{\mathbb{i}\theta}\mathbf{x} - \Delta_{\mathcal{C}}\left(\Phi_{\mathcal{P}}\left(\mathbf{b}\right)\right)\right\|_2 \le \frac{C}{\sqrt{k}} \cdot \left(\inf_{\mathbf{z}\in\mathbb{C}^d, \|\mathbf{z}\|_0\le k} \|\mathbf{x}-\mathbf{z}\|_1 \right) + A \cdot \min\left(\|\mathcal{C}\mathbf{x}\|_2, \frac{\|\mathbf{n}\|_1}{D\|\mathcal{C}\mathbf{x}\|_2} \right)$$

holds for all $\mathbf{x} \in \mathbb{C}^d$ with probability $1 - \mathcal{O}(\mathbb{e}^{-BD})$, provided that $D \ge E \cdot \tilde{d}$. Here $B, E \in \mathbb{R}^+$ are universal constants, while $C, A \in \mathbb{R}^+$ are constants that only depend on ρ and τ.

Error Guarantee

When $\mathcal{C}$ is a random matrix with i.i.d. subGaussian random entries, we can further show that

$$\min_{\theta\in[0,2\pi]} \left\| \mathbb{e}^{\mathbb{i}\theta}\mathbf{x} - \Delta_{\mathcal{C}}\left(\Phi_{\mathcal{P}}\left(\mathbf{b}\right)\right) \right\|_2 \leq \frac{C}{\sqrt{k}} \cdot \left(\inf_{\mathbf{z}\in\mathbb{C}^d, \|\mathbf{z}\|_0 \leq k} \|\mathbf{x} - \mathbf{z}\|_1 \right) + A\|\mathbf{n}\|_2$$

Numerical Results

- Test signals – sparse, unit-norm complex vectors
  - non-zero indices are independently and randomly chosen
  - non-zero entries are i.i.d. standard complex Gaussians

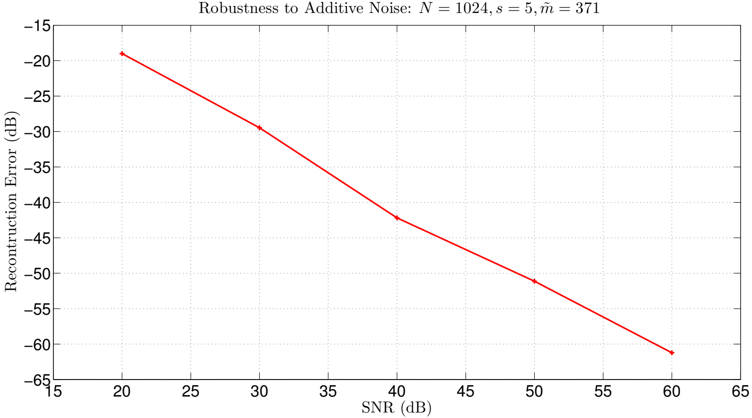
- Noise model – i.i.d. zero-mean additive Gaussian noise at different SNRs

- Errors reported as SNR (dB)

$$\text{Error (dB)} = 10\log_{10}\left(\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}\right)$$

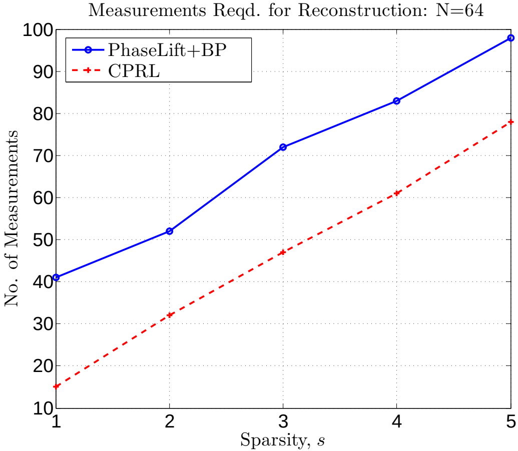
($\hat{\mathbf{x}}$ – recovered signal, $\mathbf{x}$ – true signal)

Robustness



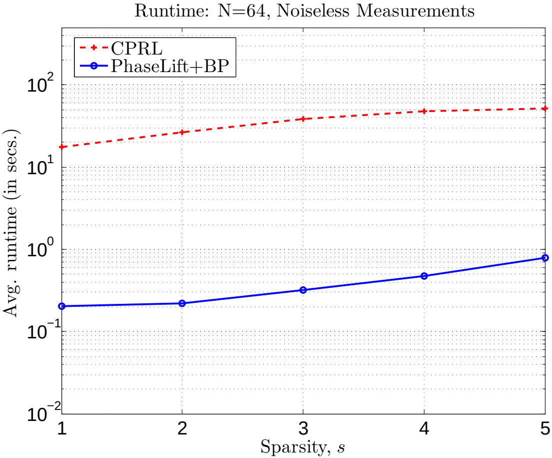
- Signal size: $d = 1024$
- Sparsity: $k = 5$
- 371 measurements $(14k \log(d/k))$

No. of measurements – Comparison with SDP-based CPRL



- $d = 64$, Noiseless measurements
- No. of measurements required for successful (relative ℓ_2-norm error $\leq 10^{-5}$) reconstruction

Corresponding Runtime



- $d = 64$
- Noiseless measurements
- Averaged over 100 trials

Discussion

(+)

- Requires $O(k \log(d/k))$ measurements.
- Significantly faster than comparable (SDP-based) methods.
- Recovery guarantee

(-)

- May require more (a small linear factor) measurements for small problems.

Summary

- Robust, efficient (FFT-time) phase retrieval algorithm
- Uses compactly supported masks and a block circulant construction in conjunction with angular synchronization
- Deterministic, well conditioned measurements masks
- Simple 2-stage method for sparse signals

References – Phase Retrieval

1. R.W. Gerchberg and W.O. Saxton. *A practical algorithm for the determination of phase from image and diffraction plane pictures*. Optik, 35:237–246, 1972

2. J.R. Fienup. *Reconstruction of an object from the modulus of its Fourier transform*. Optics Letters, 3:27–29, 1978.

3. E. J. Candes, T. Strohmer, and V. Voroninski. *Phaselift: exact and stable signal recovery from magnitude measurements via convex programming*. Comm. Pure Appl. Math., 66, 1241–1274, 2012.

4. I. Waldspurger, A. d'Aspremont, and S. Mallat. *Phase recovery, MaxCut and complex semidefinite programming*. Math. Prog., 2013.

5. B. Alexeev, A. S. Bandeira, M. Fickus, and D. G. Mixon. *Phase retrieval with polarization*. SIAM J. Imag. Sci., 7(1), 35–66.

References – Sparse Phase Retrieval

1. H. Ohlsson, A. Yang, R. Dong, and S. Sastry. *CPRL: An Extension of Compressive Sensing to the Phase Retrieval Problem*. Proc. 26^{th} Conf. Adv. Neural Inf. Proc. Sys., 1376–1384, 2012.

2. Y. Shechtman, A. Beck, and Y. C. Eldar. *GESPAR: Efficient Phase Retrieval of Sparse Signals*. IEEE Tran. Sig. Proc., 62(4):928–938, 2014.

3. P. Schniter and S. Rangan, *Compressive Phase Retrieval via Generalized Approximate Message Passing. arXiv:1405.5618, 2014.*

4. *M. Iwen, A. Viswanathan, an Y. Wang, Robust Sparse Phase Retrieval Made Easy. arXiv:1410.5295, 2014.*

Software Repository