Phase Retrieval in Coherent Diffractive Imaging

Albert Fannjiang

University of California, Davis

National Center for Theoretical Science, Taiwan, 2016

Collaborators: Peng-Wen Chen (NCHU), Gi-Ren Liu (NCKU)
Coherent diffractive imaging

- Linear propagation + intensity measurement: \( b(j)^2 = |a_j^* x_0|^2 \)
- Phase retrieval: Given \( b = (b(j)) \in \mathbb{R}_+^N \) and \( A^* = [a_j^*] \in \mathbb{C}^{N \times M} \), determine \( x_0 \).
- Geometry: Intersection of \( N \)-dim real torus of radii \( \{b(j)\} \) and complex linear subspace \( A^* \mathbb{C}^M (N > M) \).
Uniqueness for generic frames (Balan-Casazza-Edidin 06)

- Full-rank $A \in \mathbb{C}^{M \times N}, N > M$: $\{\text{col}(A)\} =$ frame
- Frames form a metric space.
- Necessary condition for injectivity (left inverse exists): $N \geq 2M$.
- Sufficient condition: If $N \geq 4M - 2$ then generic (i.e. an open dense set) frames are injective.

Fourier frame is exceptional!
Diffraction = Fourier transform

Let \( x_0(n) \) be a discrete object function with \( n = (n_1, n_2, \cdots, n_d) \in \mathbb{Z}^d \). We assume \( d \geq 2 \). \( \mathcal{M} = \{0 \leq m_1 \leq M_1, 0 \leq m_2 \leq M_2, \cdots, 0 \leq m_d \leq M_d\} \)

**Diffraction pattern**

\[
\left| \sum_{m \in \mathcal{M}} x_0(m) e^{-i2\pi m \cdot \omega} \right|^2 = \sum_{n=-M}^{M} \sum_{m \in \mathcal{M}} x_0(m + n) \overline{x_0(m)} e^{-i2\pi n \cdot w}
\]

\( w = (w_1, \cdots, w_d) \in [0, 1]^d \), \( \mathcal{M} = (M_1, \cdots, M_d) \)

**Autocorrelation**

\[
R(n) = \sum_{m \in \mathcal{M}} x_0(m + n) \overline{x_0(m)}.
\]

\( \mathcal{\tilde{M}} = \{(m_1, \cdots, m_d) \in \mathbb{Z}^d : -M_1 \leq m_1 \leq M_1, \cdots, -M_d \leq m_d \leq M_d\} \)

**Oversampling ratio = \( 2^d \)**
Ambiguities (Bruck-Sodin 1979, Hayes 1982)

- Oversampling: \( N \geq 4M - 4\sqrt{M} + 1 \).
- Global ambiguities for generic objects \( x_0 \in \mathbb{R}^M \)

  - (harmless) global phase \( x_0(\cdot) \longrightarrow e^{i\theta} x_0(\cdot) \)
  - translation \( x_0(\cdot) \longrightarrow x_0(\cdot + n), \forall n \)
  - conjugate inversion \( x_0(\cdot) \longrightarrow \overline{x_0}(\cdot) \)

- Generic objects = random vectors according to continuous prior distribution \( \Longrightarrow \) nongeneric objects \( \in \) a measure zero set.
- Problems:
  - You can not determine if a given object is generic or not since the “world ensemble” may not be absolutely continuous w.r.t. your prior distribution.
  - Global ambiguities may lead to poor reconstruction: bad algorithm or measurement scheme?
Coded diffraction pattern
Measurement matrix

- Mask function: $\mu(n)$.
- Masked object: $\tilde{x}_0(n) = \mu(n)x_0(n)$
- Randomly phased mask: $\mu(n) = \exp(i\phi(n))$ where $\phi(n)$ are random variables.
- Measurement matrix: $\Phi = \text{discrete Fourier transform}$

\[
(1 \text{ mask}) \quad A^* = \Phi \text{ diag}(\mu)
\]
\[
(2 \text{ masks}) \quad A^* = \begin{bmatrix}
\Phi \text{ diag}(\mu_1) \\
\Phi \text{ diag}(\mu_2)
\end{bmatrix}
\]
Theorem (F. 2012)

Suppose \( x_0 \in \mathbb{C}^M \) is rank \( \geq 2 \) and \( \arg(x_0) \) belongs in a proper sub-interval \( [a, b] \subset [0, 2\pi) \). Then the object is determined by one coded diffraction pattern up to a constant phase factor with probability at least

\[
1 - M \left| \frac{b - a}{2\pi} \right|^{s/2}
\]

where \( s \) is the number of nonzero pixels.

Corollary

Suppose \( x_0 \in \mathbb{R}^M \) and is rank \( \geq 2 \). Then with probability one the object is determined by one coded diffraction pattern up to \( \pm \) sign.
Uniqueness (continued)

**Theorem (F. 2012)**

Suppose $x_0 \in \mathbb{C}^M$ and is rank $\geq 2$. Then the object is determined by \textit{two} coded diffraction patterns up to a constant phase factor with probability one.

**vs Candes-Li-Soltanolkotabi 2015:**

- PhaseLift: convex programming.
- Large number of regularly sampled patterns.
- Candes-Strohmer-Voroninski 2013: Gaussian random measurement.
- Lifting $\Rightarrow$ huge increase of dimensionality & unpractical computation
Nonconvex constraint

- Non-linear system:

\[ b = |A^*x|, \quad x \in \mathcal{X} \]

(1 mask) \( \mathcal{X} = \mathbb{R}^M, \quad A^* = \Phi \text{ diag}(\mu) \)

(2 masks) \( \mathcal{X} = \mathbb{C}^M, \quad A^* = \begin{bmatrix} \Phi \text{ diag}(\mu_1) \\ \Phi \text{ diag}(\mu_2) \end{bmatrix} \)

- Non-convex feasibility problem:

Find \( \hat{y} \in A^*\mathcal{X} \cap \mathcal{Y} \)

\( \mathcal{Y} := \{y \in \mathbb{C}^N : |y| = b\} \)

\( \hat{x} = (A^*)^\dagger \hat{y} \)

- Geometry: Intersection of \( N \)-dim torus of radii \( \{b_j\} \) and linear subspace \( A^*\mathcal{X} \)
Alternating projections: feasibility problem

Two constraints: Fourier magnitude data ($N$-dim torus of uneven radii) $\cap$ oversampled Fourier matrix ($2M$-dim subspace)

von Neuman 1933

Cheney-Goldstein 1959
Bregman 1965

Non convex: local convergence?
Experiments: plain diffraction pattern

Original images  AP  HIO (Fienup 1982)
Reconstruction with coded diffraction patterns

- Convex method converges surely but (extremely) slowly.
- Nonconvex methods converge fast (with good measurement) without guarantee.
  1. Gradient descent algorithms: e.g. Wirtinger flow (Candes-Li-Soltanolkotabi 2015).
  2. Iterative projection/fixed point algorithms.
- Initial guess is crucial for non-convex methods: How to put the initial guess in the basin of attraction of the global minimizer?
Null vector method (Chen-F.-Liu 2015)

\[ A^* = [a_j^*] \]
\[ a_j^* x_0 = 0 \quad \Rightarrow \quad b_j = |a_j^* x_0| = a_j^* x_0. \]

If there are sufficiently many data that are small, then the unique null vector of the row sub-matrix may be a good bet.

\[ x_{null} := \arg \min \left\{ \sum_{i \in I} \|a_i^* x\|^2 : x \in \mathcal{X}, \|x\| = \|x_0\| \right\} \]

\[ x_{dual} := \arg \max \left\{ \|A_{Ic}^* x\|^2 : x \in \mathcal{X}, \|x\| = \|x_0\| \right\} \]

Isometry \[ \|A_I^* x\|^2 + \|A_{Ic}^* x\|^2 = \|x\|^2 \]

\[ x_{null} = x_{dual} \]

power method  

14 / 38
Null vector algorithm

Let $1_c$ be the characteristic function of the complementary index $I_c$ with $|I_c| = \gamma N$.

**Algorithm 1: The null vector method**

1. Random initialization: $x_1 = x_{\text{rand}}$
2. Loop:
   3. for $k = 1 : k_{\text{max}} - 1$ do
      4. $x'_k \leftarrow A(1_c \odot A^* x_k)$;
      5. $x_{k+1} \leftarrow \frac{x'_k}{\|x'_k\|} / \|x'_k\|$
   6. end
7. Output: $x_{\text{null}} = x_{k_{\text{max}}}$.

**Algorithm 2: The spectral vector method**

1. Random initialization: $x_1 = x_{\text{rand}}$
2. Loop:
   3. for $k = 1 : k_{\text{max}} - 1$ do
      4. $x'_k \leftarrow A(|b|^2 \odot A^* x_k)$;
      5. $x_{k+1} \leftarrow \frac{x'_k}{\|x'_k\|} / \|x'_k\|$
   6. end
7. Output: $x_{\text{spec}} = x_{k_{\text{max}}}$.

**Truncated spectral vector**

$$x_{t\text{-spec}} = \arg \max_{\|x\|=1} \|A \left(1_{\tau} \odot |b|^2 \odot A^* x \right) \| \quad \{i : |A^* x(i)| \leq \tau \|b\| \}$$

Netrapalli-Jain-Sanghavi 2015

Candes-Chen 2015
Experiments: Fourier case with two masks
Experiments: Fourier case with one mask

Error metrics often poorly reflect the quality of initialization
Performance guarantee: Gaussian case

Theorem (Chen-F.-Liu 2016)

Let $A$ be drawn from the $M \times N$ standard complex Gaussian ensemble. Let

$$\sigma := |I|/N < 1, \quad \nu = M/|I| < 1.$$ 

Then for any $x_0 \in \mathbb{C}^n$ the following error bound

$$\|x_0 x_0^* - x_{\text{null}} x_{\text{null}}^*\|_2^2 \leq c_0 \sigma \|x_0\|^4$$

holds with probability at least

$$1 - 5 \exp \left(-c_1 |I|^2/N\right) - 4 \exp(-c_2 M).$$

- Nonasymptotic estimate
- Asymptotic regime: $|I|/N \ll 1, \quad |I|^2/N \gg 1$
  \[ \implies |I| = N^\alpha, \quad \text{error} \sim N^{(\alpha-1)/2}, \quad \alpha \in (1/2, 1) \]
Experiments: Gaussian case

Empirical scaling law: Relative error $\sim L^{-\beta}$ where $L = N/M$ and $\beta \approx 1/2$.

Theoretical bound: $\text{RE} \sim \sqrt{|I|/N} = L^{(\alpha-1)/2}$ where $1/2 < \alpha < 1$. 

(a) White noise  
(b) Low-pass noise  
(c) Randomly phased Phantom
Alternating projections

- Non-convex feasibility problem:
  \[ \text{Find } \hat{y} \in A^* \mathcal{X} \cap \mathcal{Y} \]
  \[ \mathcal{Y} := \{ y \in \mathbb{C}^N : |y| = b \} \]
  \[ \hat{x} = (A^*)^\dagger \hat{y} \]

- Let \( P_1 \) and \( P_2 \) be projections onto \( A^* \mathcal{X} \) and \( \mathcal{Y} \), respectively.
  \[
  \text{(AP)} \quad P_1 P_2 y = \left[ (A^*)^\dagger \left( b \odot \frac{y}{|y|} \right) \right] x
  \]
  with initial guess \( y^{(1)} = A^* x^{(1)} \), \( x^{(1)} \in \mathcal{X} \).

- Nonconvex optimization: \( U = \{ u \in \mathbb{C}^N : |u(j)| = 1 \} \) \( N \)-torus.
  \[
  f(x, u) = \frac{1}{2} \| A^* x - u \odot b \|_2^2
  \]
  \[
  u^{(k)} = \arg \min_{u \in U} f(x^{(k)}, u) \quad \text{(non-convex)}
  \]
  \[
  x^{(k+1)} = \arg \min_{x \in \mathcal{X}} f(x, u^{(k)}) \quad \text{(non-smooth)}
  \]
Parallel AP (PAP)

\[ x^{(k+1)} = \mathcal{F}(x^{(k)}) \]

\[ \mathcal{F}(x) = \left[ (A^*)^\dagger (b \odot \frac{A^*x}{|A^*x|}) \right] x \]

\[ (A^*)^\dagger = (AA^*)^{-1}A \]

(2-mask case) \[ A^* = c \begin{bmatrix} \Phi & \text{diag}\{\mu_1\} \\ \Phi & \text{diag}\{\mu_2\} \end{bmatrix} \]

**Fact** every limit point of \( \{x^{(k)}\} \) is a fixed point of the map \( \mathcal{F} \)

**Proposition** A fixed point preserves the total signal strength, iff it is the true solution up to a global phase.

\[ \|A^*x_*\| = \|b\| \quad \text{iff} \quad x_* = \alpha x_0 \text{ with } |\alpha| = 1. \]

Otherwise \( \|A^*x_*\| < \|b\| \).
Serial AP (SAP)

Find \( \hat{y} \in \cap_{i=1}^{2} (A_i^* X \cap \mathcal{Y}_i), \quad \mathcal{Y}_i := \{y_l \in \mathbb{C}^{N/2} : |y_l| = b_l\} \)

**SAP** \( \mathcal{F}_2 \mathcal{F}_1(x) \)

\[ \mathcal{F}_l(x) = A_l \left( b_l \odot \frac{A_i^* x}{|A_i^* x|} \right), \quad l = 1, 2, \]

**PAP** \( \mathcal{F}(x) = A \left( b \odot \frac{A^* x}{|A^* x|} \right) = \frac{1}{2} (\mathcal{F}_1(x) + \mathcal{F}_2(x)) \)
Gradient map

\[ B := A \text{ diag} \left\{ \frac{A^* x_0}{|A^* x_0|} \right\} \]
\[ \mathcal{B} := \left[ \begin{array}{c} \Re[B] \\ \Im[B] \end{array} \right] \in \mathbb{R}^{2n,N} \]

\[ G(-i \text{dF} \xi) = \mathcal{B} B^\top G(-i \xi), \quad \forall \xi \in \mathbb{C}^n \]

Isomorphism \[ G(-iv) := \left[ \begin{array}{c} \Re(v) \\ -\Im(v) \end{array} \right], \quad \forall v \in \mathbb{C}^n \]

Let \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{2n} \geq \lambda_{2n+1} = \cdots = \lambda_N = 0 \) be the singular values of \( \mathcal{B} \) with the corresponding right singular vectors \( \{\eta_k \in \mathbb{R}^N\}_{k=1}^N \) and left singular vectors \( \{\xi_k \in \mathbb{R}^{2n}\}_{k=1}^{2n} \).

**Proposition**

We have \( \xi_1 = G(x_0) \), \( \xi_{2n} = G(-ix_0) \), \( \lambda_1 = 1 \), \( \lambda_{2n} = 0 \) and \( \eta_1 = |A^* x_0| \).

\[ u^{(k)} := -i(\alpha^{(k)} x^{(k)} - x_0) \quad \xi_1 \perp G(u^{(k)}), \quad \forall k \]
Spectral gap

\[ \lambda_2 = \max\{\|\mathcal{S}[B^*u]\| : u \in \mathbb{C}^n, iu \perp x_0, \|u\| = 1\} \]
\[ = \max\{\|B^\top u\| : u \in \mathbb{R}^{2n}, u \perp \xi_1, \|u\| = 1\}. \]

**Proposition**

Suppose \( x_0 \in \mathbb{C}^n \) is rank-2. Then \( \lambda_2 < 1 \) with probability one.

**Uniqueness theorem for magnitude retrieval**

If

\[ \angle A^*\hat{x} = \pm \angle A^*x_0 \]

where the \( \pm \) sign may be pixel-dependent, then almost surely

\( \hat{x} = cx_0 \) for some constant \( c \in \mathbb{R} \).

One random mask suffices!
Local geometric convergence

Theorem (Chen-F.-Liu 2015)

For any given $0 < \epsilon < 1 - \lambda_2^2$, if $x^{(1)}$ is sufficiently close to $x_0$, then with probability one PAP converges to $x_0$ geometrically after global phase adjustment

$$\|\alpha^{(k+1)}x^{(k+1)} - x_0\| \leq (\lambda_2^2 + \epsilon)\|\alpha^{(k)}x^{(k)} - x_0\|$$

where $\alpha^{(k)} = x^{(k)*}x_0 / |x^{(k)*}x_0|$.

Theorem (Chen-F.-Liu 2015)

For any given $0 < \epsilon < 1 - (\lambda_2^{(2)} \lambda_2^{(1)})^2$, if $x^{(1)}$ is sufficiently close to $x_0$ then with probability one SAP converges to $x_0$ geometrically after global phase adjustment,

$$\|\alpha^{(k+1)}x^{(k+1)} - x_0\| \leq ((\lambda_2^{(2)} \lambda_2^{(1)})^2 + \epsilon)\|\alpha^{(k)}x^{(k)} - x_0\|.$$
Experiments: with null initialization

(a) RSCB

(b) RPP
Experiments: null vector with noisy data

- **Case 1:** \( \|x_{\text{null}}\| = \|b\| \).
- **Case 2:** \( \|x_{\text{null}}\| = \|x_0\| \).
Experiments: noise stability

(a) Cameraman

(b) Phantom

(c) RSCB

(d) RPP
Douglas-Rachford splitting

▶ **Feasibility:** \( \mathcal{Y} \cap \mathcal{Z} \implies \min_{y \in \mathcal{Y}, z \in \mathcal{Z}} \frac{1}{2} \| y - z \|^2, \quad y = z \).

▶ **ADMM** (alternating direction method of multiplier)

\[
\max_{\lambda} \min_{y \in \mathcal{Y}, z \in \mathcal{Z}} \mathcal{L} := \frac{1}{2} \| y - z \|^2 + \langle \lambda, (y - z) \rangle
\]

\[
= \max_{\lambda} \min_{y \in \mathcal{Y}, z \in \mathcal{Z}} \mathcal{L} := \frac{1}{2} \| y - z + \lambda \|^2 - \frac{1}{2} \| \lambda \|^2
\]

\[
\begin{aligned}
y^{t+1} &= \arg \min_{y \in \mathcal{Y}} \frac{1}{2} \| y - z^t + \lambda^t \|^2 \\
z^{t+1} &= \arg \min_{z \in \mathcal{Z}} \frac{1}{2} \| y^{t+1} - z + \lambda^t \|^2 \\
\lambda^{t+1} &= \lambda^t + \nabla_{\lambda} \mathcal{L}(y^{t+1}, z^{t+1})
\end{aligned}
\]

▶ **DR:** \( x^t := y^{t+1} + \lambda^t \implies \)

\[
x^{t+1} = x^t + P_{\mathcal{Y}}(2P_{\mathcal{Z}} - I)x^t - P_{\mathcal{Z}}x^t
\]
Fourier domain Douglas-Rachford

\[ \mathcal{Y} = \{ y \in \mathbb{C}^N : |y| = b \}, \quad \mathcal{Z} = A^* \mathcal{X} \]

\[ \implies P_{\mathcal{Y}}(y) = b \odot \frac{y}{|y|}, \quad P_{\mathcal{Z}}(y) = A^* Ay \]

\[
S_f(y) = y + A^* \left[ A \left( 2b \odot \frac{y}{|y|} - y \right) \right] x - b \odot \frac{y}{|y|}
\]

Gradient

\[ J_f \nu = (I - B^* B) \mathcal{R}(\nu) + iB^* B \mathcal{S}(\nu) \]

\( J_f \) is a real, but not complex, linear map

\[
S(x) = x + \left[ \tilde{A} \left( 2b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|} \right) - x \right] x - \tilde{A} \left( b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|} \right)
\]
Fixed point with two masks

\[ S_f(y_\infty) = y_\infty, \quad x_\infty = A y_\infty. \]

\[ y_\infty = e^{i\theta} (|y_0| + v) \odot \frac{y_0}{|y_0|} \]

\(|y_0| + v \text{ has all nonnegative components}\)

\[ v \in \text{null}_\mathbb{R}(\mathcal{B}) \subset \mathbb{R}^N \]

Theorem (Chen-F. 2016)

The projected fixed point is unique, i.e. \( x_\infty = e^{i\theta} x_0 \) almost surely.
FDR locally converges geometrically

Theorem (Chen-F. 2016)

For $0 < \epsilon < 1 - \lambda_2$, if $\alpha^{(1)} x^{(1)}$ is sufficient close to $x_0$, then FDR converges geometrically to the solution

$$\|\alpha^{(k)} x^{(k)} - x_0\| \leq (\lambda_2 + \epsilon)^{k-1} \|\alpha^{(1)} x^{(1)} - x_0\|.$$

▶ Explicit measurement schemes.
▶ Explicit characterization of $\lambda_2 < 1$.
▶ No hard-to-verify assumptions.
▶ Convex setting (He-Yuan 2012, 2015): $k$-th error $= \mathcal{O}(1/k)$. 
Experiments: Two patterns
Fourier domain vs. object domain DR

\begin{align*}
\text{(FDR)} \quad S_f(x) &= y + A^* A \left( 2b \odot \frac{y}{|y|} - y \right) - b \odot \frac{y}{|y|} \\
\text{(ODR)} \quad S(x) &= x + \tilde{A} \left( 2b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|} \right) - x - \tilde{A} \left( b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|} \right)
\end{align*}

\tilde{A} : \text{various extensions of } A
Conclusion

- Two globally convergent schemes in practice:
  1. AP + null initialization
  2. FDR
- Open problem: proof of global convergence.
References


5. P. Chen, A. Fannjiang and G. Liu, “Phase retrieval with one or two coded diffraction patterns by alternating projection with the null initialization,” arxiv:1510.07379.


