Fixed Point Algorithms for Phase Retrieval and Ptychography

Albert Fannjiang

University of California, Davis

Mathematics of Imaging Workshop:
Variational Methods and Optimization in Imaging
IHP, Paris, February 4th-8th 2019

Collaborators: Pengwen Chen (NCHU), Gi-Ren Liu (NCKU), Zheqing Zhang (UCD)
Outline

- Introduction
- Alternating projection for feasibility
- Douglas-Rachford splitting/ADMM
- Convergence analysis
- Initialization methods
- Blind ptychoraphy
- Conclusion
Phase retrieval

- X-ray crystallography: von Laue, Bragg etc. since 1912.
- Non-periodic structures: Gerchberg, Saxton, Fienup etc since 1972, delay due to low SNR.
- Nonlinear signal model: data = diffraction pattern = $|\mathcal{F}(f)|^2$

$\mathcal{F} =$ Fourier transform, $|\cdot|$ = componentwise modulus.
Coded diffraction pattern
Alternating projections
Nonconvex feasibility

- Masking $\mu$ + propagation $F$ + intensity measurement:

  \[
  \text{coded diffraction pattern} = |F(f \circ \mu)|^2. \]

- F (2012): Uniqueness with probability one

  \[
  b = |Ax|, \quad x \in X
  \]

  (1 mask) \quad $X = \mathbb{R}^n$, \quad $A = \Phi \, \text{diag}(\mu)$

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

- Non-convex feasibility:

  Find $\hat{y} \in A\mathcal{X} \cap \mathcal{Y}$

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

  Intersection of $N$-dim torus $\mathcal{Y}$ and $n$- or $2n$-dim subspace $A\mathcal{X}$
Alternating projections

von Neuman 1933

Cheney-Goldstein 1959
Bregman 1965

Non convex: local convergence?
Coded vs plain diffraction pattern

(a) coded; 40 iter

(b) error

(c) plain; 1000 iter

(d) error

- AP: real-valued Cameraman with one diffraction pattern.
- Plain diffraction pattern allows ambiguities such as translation, twin-image which are forbidden by the presence of a random mask.
Douglas-Rachford splitting
Minimization with a sum of two objective functions

$$\arg\min_u K(u) + L(v), \quad u = v$$

where

$$K = \text{Indicator function of } \{Ax : x \in \mathbb{C}^n\}$$

$$L(v) = \sum_i |v[i]|^2 - b^2[i] \ln |v[i]|^2 \quad (\text{Poisson log-likelihood})$$

- Projection onto $K = AA^\dagger u$.
- Linear constraint $u = v$.
- $L$ has a simple asymptotic form
Gaussian log-likelihood

- High SNR: Gaussian distribution with variance = mean: $\frac{e^{-(b^2 - \lambda)^2 / (2\lambda)}}{\sqrt{2\pi\lambda}}$.
- Gaussian log-likelihood: $\lambda = |v|^2$

$$\sum_j \ln |v[j]| + \frac{1}{2} \left| \frac{b^2[j]}{|v[j]|} - |v[j]| \right|^2 \rightarrow L$$

- In the vicinity of $b$, we make the substitution

$$\frac{b[j]}{|v[j]|} \rightarrow 1, \quad \ln |v[j]| \rightarrow \ln \sqrt{b[j]}$$

$$\text{const.} + \frac{1}{2} \sum_j |b[j] - |v[j]||^2 \rightarrow L$$

which is the smoothest of the 3 functions.
Alternating projections revisited

- Hard constraint $u = v$

$$\arg \min_u K(u) + \mathcal{L}(u) = \arg \min_x \mathcal{L}(u), \quad u = Ax$$

where

$$K = \text{Indicator function of } \{Ax : x \in \mathbb{C}^n\}$$

$$\mathcal{L}(u) = \frac{1}{2} \|b - |u|\|^2 \quad \text{(Gaussian log-likelihood)}.$$ 

- $\mathcal{L}$ non-smooth where $b$ vanishes.
- AP = gradient descent with unit stepsize: $x^{k+1} = x^k - \nabla \mathcal{L}(x^k)$.
- Wirtinger flow = gradient descent with

$$\mathcal{L} = \frac{1}{2} \||Ax|^2 - b\|^2 \quad \text{(additive i.i.d. Gaussian noise)}.$$
Proximal optimality

- Proximity operators are generalization of projections:

\[
\text{prox}_{\mathcal{L}/\rho}(u) = \arg \min_x \mathcal{L}(x) + \frac{\rho}{2} \|x - u\|^2
\]

\[
\text{prox}_{K/\rho}(u) = AA^\dagger u.
\]

For simplicity, set \( \rho = 1 \).

- Proximal reflectors \( R_{\mathcal{L}} = 2 \text{prox}_{\mathcal{L}} - I \), \( R_K = 2 \text{prox}_K - I \)

- Proximal optimality:

\[
0 \in \partial \mathcal{L}(x) + \partial K(x) \iff \xi = R_{\mathcal{L}}R_K(\xi), \quad x = \text{prox}_K(\xi)
\]
Let $\eta = R_K(\xi)$. Then $\xi = R_L(\eta)$.

Also $\zeta := \frac{1}{2}(\xi + \eta) = \prox_L(\eta) = \prox_K(\xi)$. Equivalently

$$\xi \in \partial K(\zeta) + \zeta, \quad \eta \in \partial L(\zeta) + \zeta$$

Adding the two equations: $0 \in \partial K(\zeta) + \partial L(\zeta)$.

Finally $\zeta = \prox_K(\xi)$ is a stationary point.
Douglas-Rachford splitting (DRS)

- Optimality leads to Peaceman-Rachford splitting:
  \[ z^{k+1} = R_{\mathcal{L}/\rho} R_{\mathcal{K}/\rho}(z^k). \]

- DRS \( z^{l+1} = \frac{1}{2} z^l + \frac{1}{2} R_{\mathcal{L}/\rho} R_{\mathcal{K}/\rho}(z^l) \): for \( l = 1, 2, 3 \cdots \)
  \[
  y^{l+1} = \text{prox}_{\mathcal{K}/\rho}(u^l);
  
  z^{l+1} = \text{prox}_{\mathcal{L}/\rho}(2y^{l+1} - u^l);
  
  u^{l+1} = u^l + z^{l+1} - y^{l+1}.
  \]

- \( \gamma = 1/\rho = \) stepsize; \( \rho = 0 \) the classical DR algorithm.

- Alternating Direction Method of Multipliers (ADMM) applied to the dual problem
  \[
  \max_{\lambda} \min_{y,z} \mathcal{L}^*(y) + K^*(-A^*z) + \langle \lambda, y - A^*z \rangle + \frac{\rho}{2} \| A^*z - y \|^2
  \]
Object update: \( f = A^\dagger u^\infty \) where \( u^\infty \) is the terminal value of

\[
u^{l+1} = \frac{1}{\rho + 1} u^l + \frac{\rho - 1}{\rho + 1} Pu^l + \frac{1}{\rho + 1} b \odot \text{sgn}(2Pu^l - u^l) = \frac{1}{2} u^l + \frac{\rho - 1}{2(\rho + 1)} Ru^l + \frac{1}{\rho + 1} b \odot \text{sgn}(Ru^l)
\]

where \( P = AA^\dagger \) is the orthogonal projection onto the range of \( A \) and \( R = 2P - I \) is the corresponding reflector.

\( \rho = 0 \): the classical Douglas-Rachford algorithm

\[
u^{l+1} = \frac{1}{2} u^l - \frac{1}{2} Ru^l u^l + b \odot \text{sgn}(Ru^l) = u^l - Pu^l + b \odot \text{sgn}(Ru^l).
\]
Convergence analysis
Convergence analysis

- Aragoón-Borwein (2012): global convergence of DR ($\rho = 0$) for intersection of a line and a circle.
- Li-Pong (2016):
  - $\mathcal{L}$ has uniformly Lipschitz gradient (ULG).
  - DRS with $\rho$ sufficiently large, depending on Lipschitz constant.
  - Global convergence: cluster point = stationary point.
  - Local geometric convergence for semi-algebraic case.

$K$ and $\mathcal{L}$ don’t have ULG and optimal performance is with $\rho \sim 1$.

Fixed point equation

$$u = \frac{1}{2}u + \frac{\rho - 1}{2(\rho + 1)}R_\infty u + \frac{1}{\rho + 1}b \odot \text{sgn}(R_\infty u)$$

The differential map is given by $$\Omega J_A(\eta)$$ where

$$J_A(\eta) = CC^\dagger \eta - \frac{1}{1 + \rho} \left[ \Re(2CC^\dagger \eta - \eta) ight. \right.$$

$$\left. + \imath \left( I - \text{diag}(b/|Ru|) \right) \Im \left( 2CC^\dagger \eta - \eta \right) \right]$$

where

$$\Omega = \text{diag}(\text{sgn}(Ru)), \quad C = \Omega^* A.$$
Fixed point analysis

Two randomly coded diffraction patterns:

- F (2012) – intersection $\sim S^1$ (arbitrary phase factor).
- Chen & F (2016) – DR ($\rho = 0$) fixed points $u$ take the form

\[ u = e^{i\theta} (b + r) \odot \text{sgn}(Af), \quad r \in \mathbb{R}^N, \quad b + r \geq 0 \]

\[ \Rightarrow \text{sgn}(u) = \theta + \text{sgn}(Af) \]

where $r$ is a real null vector of $A^\dagger \text{diag}[\text{sgn}(Af)]$

\[ \Rightarrow \text{DR fixed point set has real dimension } N - n. \]

- Chen, F & Liu (2016) – AP based on the hard constraint $u = v$

AP fixed point $x_*$:  \[ \|Ax_*\| = \|Af\| \quad \text{iff} \quad x_* = \alpha f, \quad |\alpha| = 1. \]
Spectral gap and linear convergence rate

$J_A$ can be analyzed by the eigen-structure of

$$H := \begin{bmatrix} \Re[A^\dagger \Omega] \\ \Im[A^\dagger \Omega] \end{bmatrix}, \quad \Omega = \text{diag}(\text{sgn}(Af)).$$

- $\|J_A(\eta)\| = \|\eta\|$ occurs at $\eta = \pm ib$.
- Linear convergence rate is related to the spectral gap of $H$.
- One randomly coded diffraction pattern:
  - Chen & F (2016) – the differential map at $Af$ has the largest singular value 1 corresponding to the constant phase and a positive spectral gap $\Rightarrow$ the true solution is an attractor (local linear convergence).
  - F & Zhang (2018) – the differential map at any DR fixed point has a spectral radius $= 1$.
  - Chen, F & Liu (2016) – same for AP (parallel or serial).
Proposition

Let $u$ be a fixed point and $f_\infty := A^\dagger u$.

(i) $\rho \geq 1$: If $\|J_A(\eta)\|_2 \leq \|\eta\|_2$ then $|\mathcal{F}(\mu, f_\infty)| = b$.

(ii) $\rho \geq 0$: If $|\mathcal{F}(\mu, f_\infty)| = b$ then $\|J_A(\eta)\|_2 \leq \|\eta\|_2$. where the equality holds iff $\eta$ parallels $\nu b$.

Summary:

- DRS ($\rho \geq 1$) fixed point is linearly stable iff it is a true solution
- DR ($\rho = 0$) introduces harmless, stable fixed points.
- AP likely introduces spurious nonsolution fixed points.
- Linear convergence rate:

  Serial AP < parallel AP \sim DRS (\rho = 1) < DR (\rho = 0).
Initialization
Initialization by feature extraction

\[ b = |Af| \] where \( A \in \mathbb{C}^{N \times n} \) is the measurement matrix.

**Feature:** two sets of signals, weak and strong.
- Weak signals selected by a threshold \( \tau \), i.e. \( b_i \leq \tau, i \in I \).
- \( x_{\text{null}} := \) ground state of \( A_I \).

**Isometry:** \( \|Ax\|^2 = \|A_I x\|^2 + \|A_{I_c} x\|^2 = \|x\|^2 \implies \)

\[
x_{\text{null}} = \arg \min \left\{ \|A_I x\|^2 : \|x\| = \|f\| \right\}
= \arg \max \left\{ \|A_{I_c} x\|^2 : \|x\| = \|f\| \right\}
\]
solved by the power method efficiently.

**Non-isometry \( \implies QR: A = QR \)**
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. $x_{k+1} \leftarrow \left[ x'_k \right]_\chi / \| x'_k \|_\chi$
   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. $x_{k+1} \leftarrow \left[ x'_k \right]_\chi / \| x'_k \|_\chi$
   6. end
7. **Output:** $x_{\text{spec}} = x_{k_{\text{max}}}$.

**Truncated spectral vector**

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

\[ \{ i : |A^* x(i)| \leq \tau \|b\| \} \]

Netrapalli-Jain-Sanghavi 2015

Candes-Chen 2015
Theorem (Chen-F.-Liu 2016)

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

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

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

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

holds with probability at least

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

- Non-asymptotic estimate: $n < |I| < N < |I|^2$, \hspace{1em} $L = N/n$

\[|I| = N^\alpha n^{1-\alpha} \implies \text{RE} \sim L^{(\alpha-1)/2}, \quad \alpha \in [1/2, 1)\]
2 CDPs, \(|I| = \sqrt{nN}\).

Uniqueness of phase retrieval with 2 CDPs (F. 2012).

Figure: Noisy estimation by Algorithm 1 with \(|I| = \sqrt{nN}\) at various NSRs.
Experiments: with null initialization

- PAP: two diffraction patterns used in parallel
- SAP: two diffraction patterns used in serial
Comparison with Wirtinger flow
Complex Gaussian noise

\[ b = |Af + \text{complex Gaussian noise}| \]

\[ \text{NSR} = \text{noise/signal} \]
Blind ptychography
Ptychography: extended objects

Hoppe (1969), Nellist-Rodenburg (95), Faulkner-Rodenburg (04, 05), Thibault et al. (08, 09)

- Inverse problem with shifted *windowed* Fourier intensities.
- Unlimited, extended objects: structural biology, materials science etc.
Linear phase ambiguity

Consider the probe and object estimates

\[ \nu^0(n) = \mu^0(n) \exp(-ia - iw \cdot n), \quad n \in \mathcal{M}^0 \]
\[ g(n) = f(n) \exp(ib + iw \cdot n), \quad n \in \mathbb{Z}_n^2 \]

for any \( a, b \in \mathbb{R} \) and \( w \in \mathbb{R}^2 \). We have all \( n \in \mathcal{M}^t, t \in \mathcal{T} \)

\[ \nu^t(n) g^t(n) = \mu^t(n) f^t(n) \exp(i(b - a)) \exp(iw \cdot t). \]
Raster scan pathology

Raster scan: $t_{kl} = \tau(k, l), k, l \in \mathbb{Z}$ where $\tau$ is the step size.

$\mathcal{M} = \mathbb{Z}_n^2, \mathcal{M}^0 = \mathbb{Z}_m^2, n > m$, with the periodic boundary condition.
Mixing schemes

- **Partial perturbation** $t_{kl} = \tau(k, l) + (\delta^1_k, \delta^2_l)$.
- **Full perturbation** $t_{kl} = \tau(k, l) + (\delta^1_{kl}, \delta^2_{kl})$. 
Mask phase constraint (MPC)

- $\mu^0$: independent phases with range $\geq \pi$.
- $\nu^0$ satisfies MPC if $\nu_0(n)$ and $\mu^0(n)$ form an acute angle

$$|\arg[\nu^0(n)/\mu^0(n)]| < \pi/2$$
Global uniqueness

**Theorem (F 2018)**

Suppose $f$ does not vanish in $\mathbb{Z}_n^2$. Let $a^i_j = 2\delta_{j+1}^i - \delta_j^i - \delta_{j+2}^i$ and let $\{\delta^i_{jk}\}$ be the subset of perturbations satisfying $\gcd_{jk}\{|a^i_{jk}|\} = 1$, $i = 1, 2$, and

$$2\tau \leq m - \max_{i=1,2}\{\delta^i_{jk+2} - \delta^i_{jk}\} \quad (Overlap > 50\%)$$

$$\max_{i=1,2}[|a^i_{jk}| + \max_{k'}\{\delta^i_{k'+1} - \delta^i_{k'}\}] \leq m - \tau$$

$$\delta^i_{jk+1} - \delta^i_{jk+2} \leq \tau \leq m - 1 + \delta^i_{jk+1} - \delta^i_{jk+2}.$$

Then APA and SF are the only ambiguities, i.e. for some explicit $r$

$$g(n)/f(n) = \alpha^{-1}(0) \exp(\text{i}n \cdot r),$$

$$\nu^0(n)/\mu^0(n) = \alpha(0) \exp(\text{i}\phi(0) - \text{i}n \cdot r)$$

$$\theta_{kl} = \theta_{00} + t_{kl} \cdot r.$$
Initialization with mask phase constraint

- Mask/probe initialization

\[ \mu_1(n) = \mu^0(n) \exp[i\phi(n)], \]

where \( \phi(n) \) i.i.d. uniform on \((-\pi/2, \pi/2)\)

Relative error of the mask estimate

\[
\sqrt{\frac{1}{\pi} \int_{-\pi/2}^{\pi/2} |e^{i\phi} - 1|^2 d\phi} = \sqrt{2(1 - \frac{2}{\pi})} \approx 0.8525
\]

- Object initialization: \( f_1 = \) constant or random phase object.
Alternating minimization

$|\mathcal{F}(\mu, f)| = b$ : the ptychographic data. Define $A_k h := \mathcal{F}(\mu_k, h)$, $B_k \eta := \mathcal{F}(\eta, f_{k+1})$. We have $A_k f_{j+1} = B_j \mu_k$.

1. Initial guess $\mu_1$.

2. Update the object estimate $f_{k+1} = \text{argmin}_{g \in \mathbb{C}^{n \times n}} \mathcal{L}(A_k^* g)$

3. Update the probe estimate $\mu_{k+1} = \text{argmin}_{\nu \in \mathbb{C}^{m \times m}} \mathcal{L}(B_k^* \nu)$

4. Terminate when $\|B_k^* \mu_{k+1} - b\|$ is less than tolerance or stagnates. If not, go back to step 2 with $k \rightarrow k + 1$. 
Fixed point algorithm with $\rho = 1$

- $\rho = 1$
- Reflectors: $R_k = 2P_k - I, S_k = 2Q_k - I$.
- Gaussian:

  $$u_k^{l+1} = \frac{1}{2} u_k^l + \frac{1}{2} b \odot \text{sgn}(R_k u_k^l)$$

  $$v_k^{l+1} = \frac{1}{2} v_k^l + \frac{1}{2} b \odot \text{sgn}(S_k v_k^l)$$.

- Poisson:

  $$u_k^{l+1} = \frac{1}{2} u_k^l - \frac{1}{3} R_k u_k^l + \frac{1}{6} \sqrt{|R_k u_k^l|^2 + 24 b^2} \odot \text{sgn}(R_k u_k^l)$$

  $$v_k^{l+1} = \frac{1}{2} v_k^l - \frac{1}{3} S_k v_k^l + \frac{1}{6} \sqrt{|S_k v_k^l|^2 + 24 b^2} \odot \text{sgn}(S_k v_k^l)$$. 
correlation length $c = 0, 0.4m, 0.7m, 1m$
Rank-one vs. full-rank
Independent vs. correlated mask

![Graph showing the comparison between independent and correlated masks over epochs, with different markers and lines for different correlation lengths.](image-url)
Poisson noise

- Photon counting noise: \( b^2 = \text{Poisson r.v. with mean} = |Af|^2 \).
- Gaussian log-likelihood outperforms Poisson log-likelihood.
Disorder can better condition measurement schemes: random mask, random perturbation to raster scan

Analytical and statistical considerations can guide our way to a better objective function

Fixed point analysis can help determine parameters or select algorithms

Initialization by feature extraction

Thank you!


