Fixed Point Algorithms for Phase Retrieval

Albert Fannjiang, UC Davis

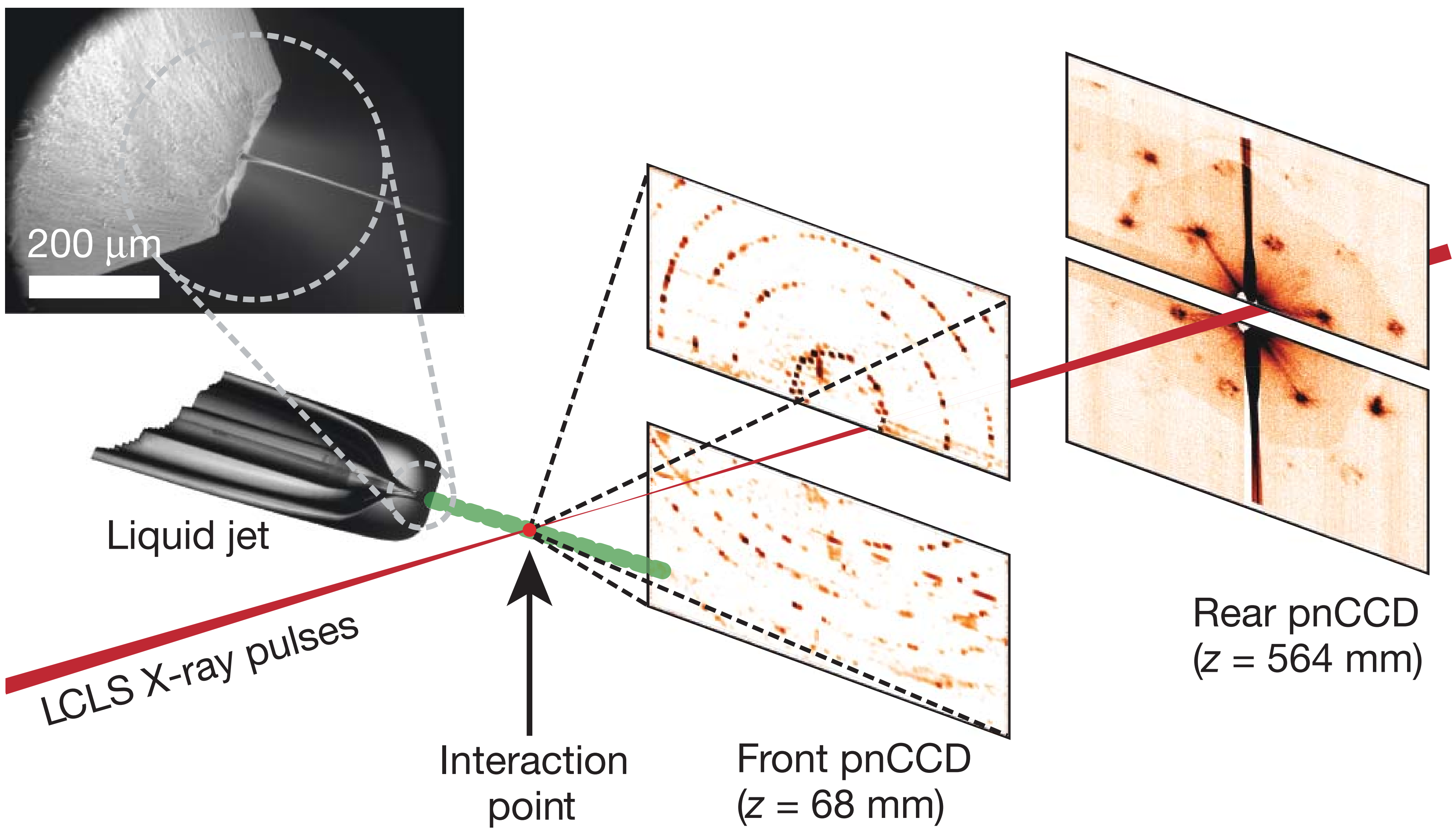
Collaborators: Pengwen Chen (NCHU), Gi-Ren Liu (NTHU)

Stanford, Feb 3, 2016

Outline

- Coded diffraction patterns: one or two patterns
- Fixed point algorithms
- Alternating projections, Douglas-Rachford etc.
- Fixed point: uniqueness
- Convergence: local vs. global
- Simulations

Coherent X-ray diffraction



Chapman et al. 2011

Diffract before destruct

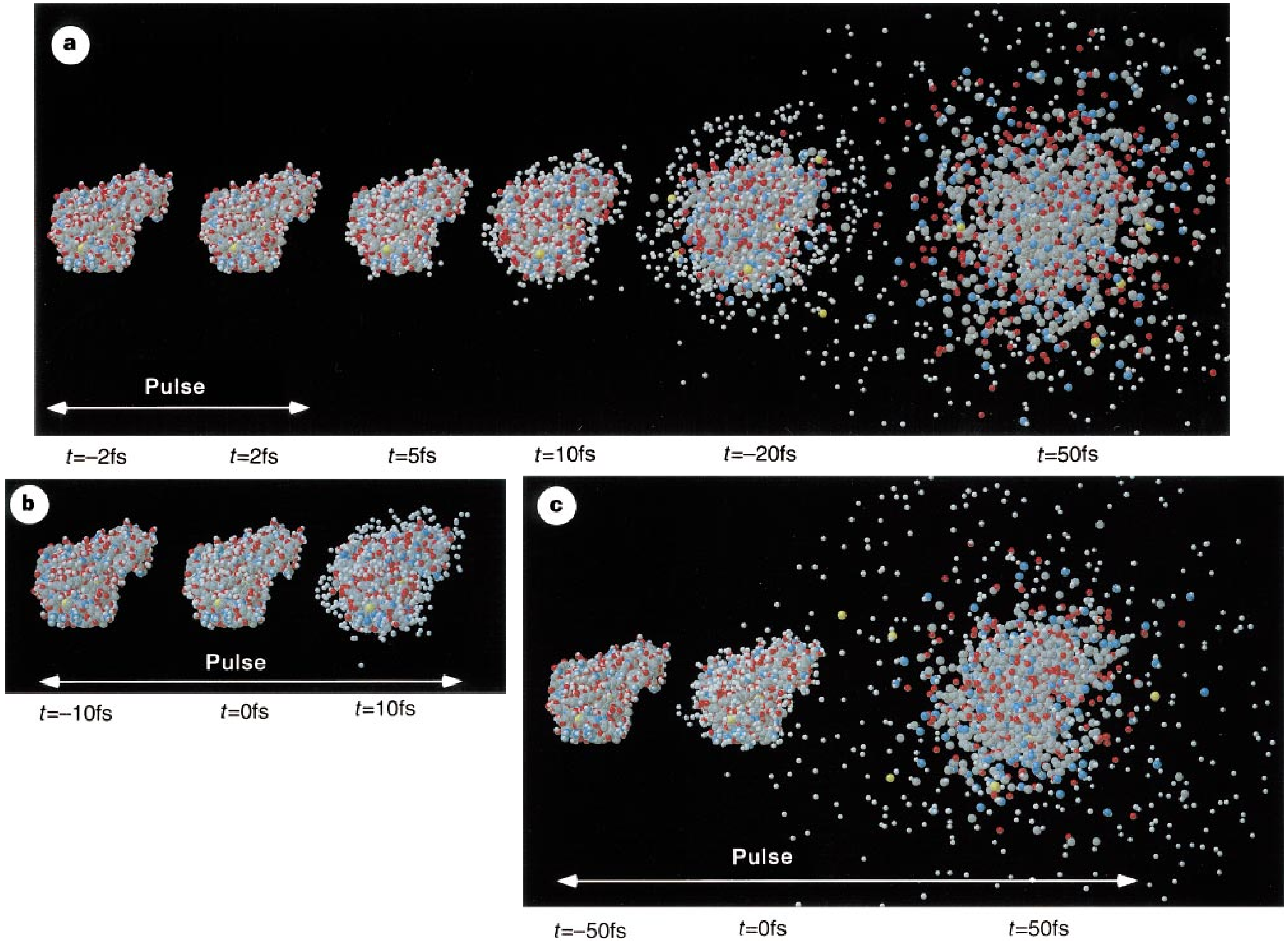


Figure 2 Explosion of T4 lysozyme (white, H; grey, C; blue, N; red, O; yellow, S) induced by radiation damage. The integrated X-ray intensity was 3×10^{12} (12 keV) photons per 100-nm diameter spot (3.8×10^{6} photons per Å^{2}) in all cases. **a**, A protein exposed to an X-ray pulse with an FWHM of 2 fs, and disintegration followed in time. Atomic positions in the first two structures (before and after the pulse) are practically identical at this pulse length because of an inertial delay in the explosion. $R_{\mathrm{nucl}} = 3\%$, $R_{\mathrm{elec}} = 11\%$ **b**, Lysozyme exposed to the same number of photons as in **a**, but the FWHM of the pulse was 10 fs. Images show the structure at the beginning, in the middle and near the end of the X-ray pulse. $R_{\mathrm{nucl}} = 7\%$, $R_{\mathrm{elec}} = 12\%$ **c**, Behaviour of the protein during an X-ray pulse with an FWHM of 50 fs. $R_{\mathrm{nucl}} = 26\%$, $R_{\mathrm{elec}} = 30\%$.

Neutze et al. 2000

Diffraction pattern

Let $x_0(\mathbf{n})$ be a discrete object function with $\mathbf{n} = (n_1, n_2, \cdots, n_d) \in \mathbb{Z}^d$. 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

$$\left|\sum_{\mathbf{m}\in\mathcal{M}} x_0(\mathbf{m}) e^{-\mathrm{i}2\pi\mathbf{m}\cdot\boldsymbol{\omega}}\right|^2 = \sum_{\mathbf{n}=-\mathbf{M}}^{\mathbf{M}} \boxed{\sum_{\mathbf{m}\in\mathcal{M}} x_0(\mathbf{m}+\mathbf{n})\overline{x_0(\mathbf{m})}} e^{-i2\pi\mathbf{n}\cdot\mathbf{w}}$$

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

Autocorrelation

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

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

Oversampling ratio = 2^d

Phase information

$$x_L = \textbf{Lena}$$
$$y_L(\mathbf{w}) = |y_L(\mathbf{w})| e^{\mathrm{i}\theta_L(\mathbf{w})}$$

$$y_1(\mathbf{w}) = |y_B(\mathbf{w})| e^{\mathrm{i}\theta_L(\mathbf{w})}$$
$$x_1 = |\boldsymbol{\Phi}^* y_1|$$

$$x_B = \textbf{Barbara}$$
$$y_B(\mathbf{w}) = |y_B(\mathbf{w})| e^{\mathrm{i}\theta_B(\mathbf{w})}$$

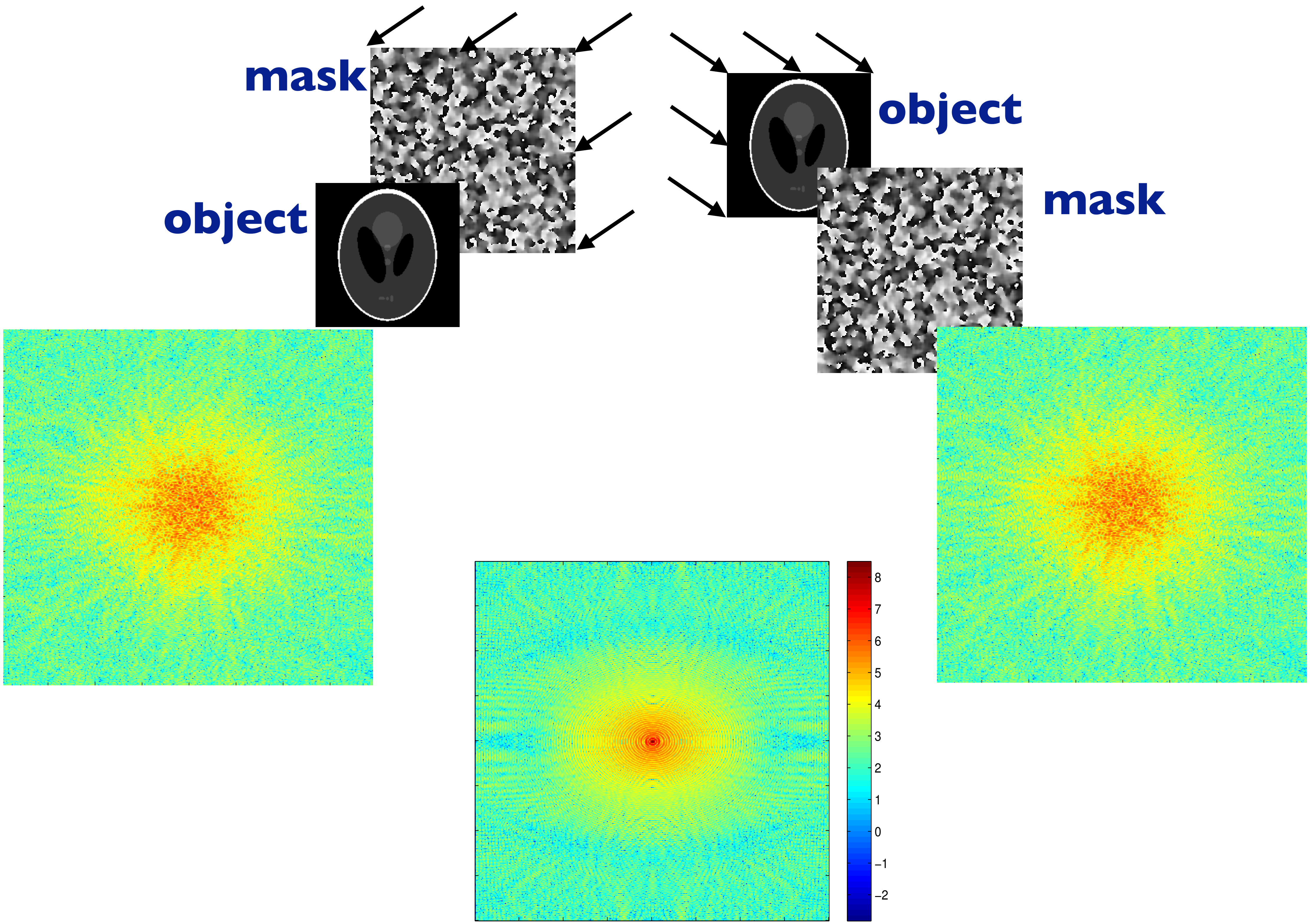
$$y_2(\mathbf{w}) = |y_L(\mathbf{w})| e^{\mathrm{i}\theta_B(\mathbf{w})}$$
$$x_2 = |\boldsymbol{\Phi}^* y_2|$$

Phase= Face

$$y_1(\mathbf{w}) = |y_B(\mathbf{w})| e^{\mathrm{i}\theta_L(\mathbf{w})}$$
$$x_1 = |\mathbf{\Phi}^* y_1|$$

$$y_2(\mathbf{w}) = |y_L(\mathbf{w})| e^{\mathrm{i}\theta_B(\mathbf{w})}$$
$$x_2 = |\mathbf{\Phi}^* y_2|$$

Coded diffraction patterns



Mask function $\{\mu(\mathbf{n})\}$

Masked object $\tilde{x}_0(\mathbf{n}) = x_0(\mathbf{n})\mu(\mathbf{n})$

Phase mask $e^{i\phi(\mathbf{n})}$

$$\text{(1-mask)} \quad A^* = c\Phi \ \operatorname{diag}\{\mu\}$$

$$\text{(2-mask case)} \quad A^* = c \begin{bmatrix} \Phi \ \operatorname{diag}\{\mu_1\} \\ \Phi \ \operatorname{diag}\{\mu_2\} \end{bmatrix}$$

With proper normalization, A^* is isometric.

Asymptotic: Chai-Moscoso-Papanicolaou 2011 (large aperture, no mask)

Uniqueness: F. 2012 (1 or 2 oversampled patterns)
Candes-Li-Soltanolkotabi 2015 (many patterns)

Notation

$$\begin{tikzcd}[column sep=huge] \text{Fourier} \quad y \in \mathbb{C}^N & \text{Object} \quad x \in \mathbb{C}^n \arrow[l, shift right=2, "A^*"'] \arrow[from=1-1, to=1-2, shift right=2, "A"'] \end{tikzcd}$$

We convert the d-dimensional ($d \geq 2$) grid into an ordered set of index and let $n = |\mathcal{M}|$. Let N denote the total number of measured data and hence $A \in \mathbb{C}^{n,N}$.

d=2 One-pattern N=4n Two-pattern N=8n

$\mathbb{C}^n = \mathbb{R}^n \oplus_{\mathbb{R}} i\mathbb{R}^n$ is isomorphic to $\mathbb{R}^{2n}$ via the map

$$G(v) := \begin{bmatrix} \Re(v) \\ \Im(v) \end{bmatrix}, \quad \forall v \in \mathbb{C}^n$$

real inner product

$$\langle u, v \rangle := \Re(u^* v) = G(u)^\top G(v), \quad u, v \in \mathbb{C}^n$$

Phase retrieval

$$b = |A^* x|, \quad x \in \mathcal{X}.$$

One-mask case $\mathcal{X} = \mathbb{R}^n$ or $\mathbb{R}^n_+$

Two-mask case $\mathcal{X} = \mathbb{C}^n$

$$\text{Find} \quad \hat{y} \in A^* \mathcal{X} \cap \mathcal{Y},$$

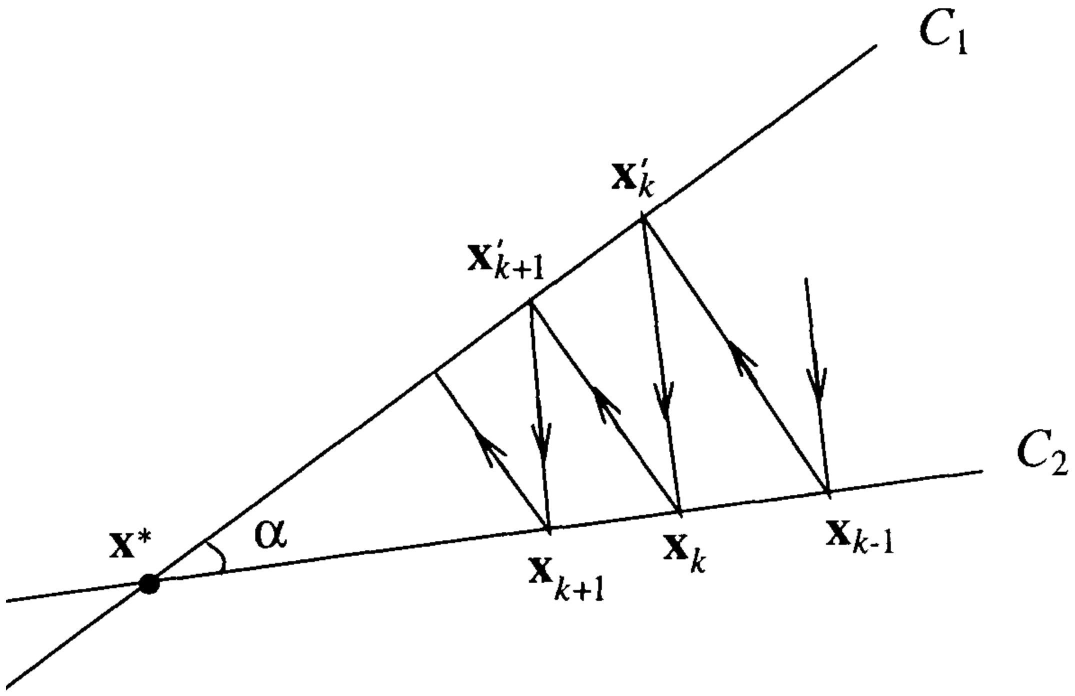
$$\mathcal{Y} := \{y \in \mathbb{C}^N : |y| = b\}$$

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

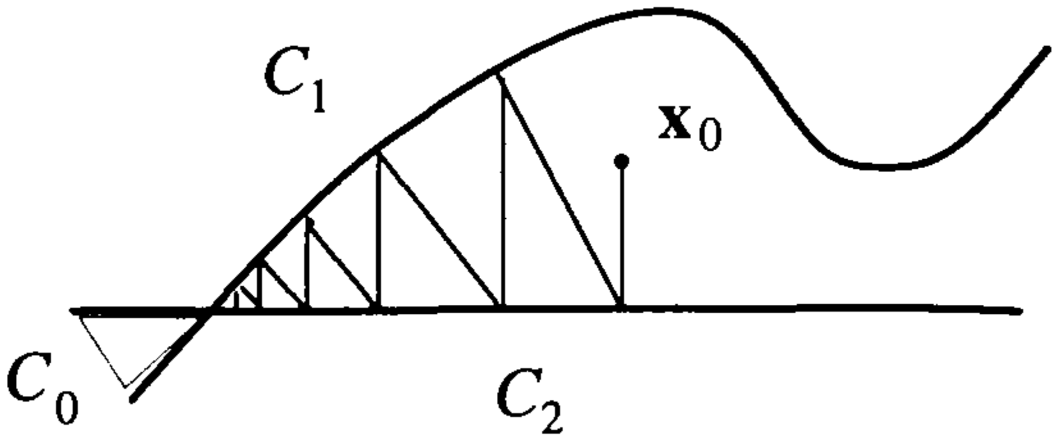
Fixed point algorithms

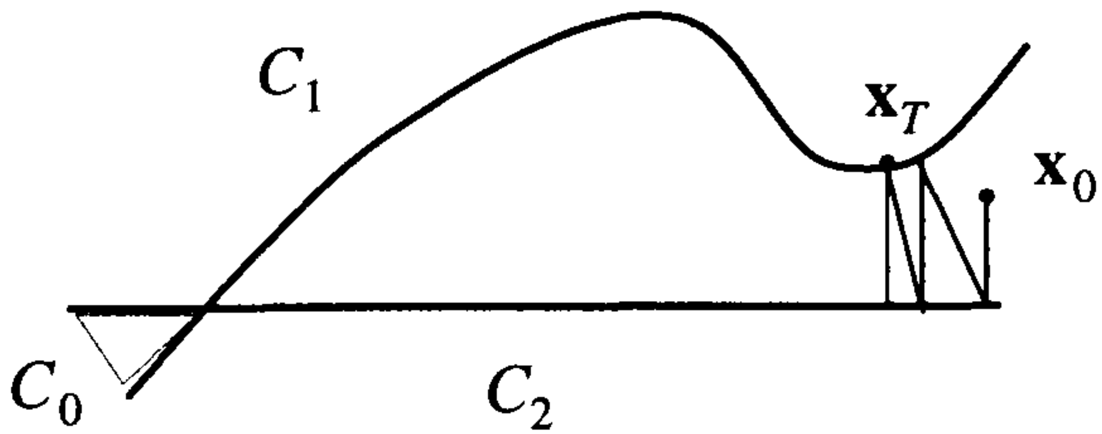
von Neuman 1933

Cheney-Goldstein 1959
Bergman 1965



Non convex: local convergence?





Alternating projection

Let P_1 be the projection onto $A^*\mathcal{X}$ and P_2 the projection onto $\mathcal{Y}$.

$$P_1 P_2 y$$

$$\text{initial guess } y^{(1)} = A^* x^{(1)}, x^{(1)} \in \mathcal{X}$$

Non-convex optimization

$$f(x, u) = \frac{1}{2} \|A^* x - u \odot b\|^2$$

$$f(x^{(k)}, u^{(k)}) = \min_{u \in U} f(x^{(k)}, u),$$

$$f(x^{(k+1)}, u^{(k)}) = \min_{x \in \mathcal{X}} f(x, u^{(k)})$$

Parallel AP (PAP)

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

$$\mathcal{F}(x) = \left[(A^*)^\dagger (b \odot \frac{A^* x}{|A^* x|})\right]_{\mathcal{X}} \qquad (A^*)^\dagger = (AA^*)^{-1}A$$

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

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.

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

Serial AP (SAP)

Find $\quad \hat{y} \in \cap_{l=1}^{2} \left(A_l^* \mathcal{X} \cap \mathcal{Y}_l \right), \quad \mathcal{Y}_l := \{ y_l \in \mathbb{C}^{N/2} : |y_l| = b_l \}$

SAP $\quad \mathcal{F}_2 \mathcal{F}_1 (x)$

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

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

Gradient representation

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

$$\boxed{G(-id\mathcal{F}\xi) \ = \ \mathcal{B}\mathcal{B}^\top G(-i\xi), \quad \forall \xi \in \mathbb{C}^n}$$

Isomorphism $G(-iv) := \begin{bmatrix} \Im(v) \\ -\Re(v) \end{bmatrix}, \quad \forall v \in \mathbb{C}^n$

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* $\eta_1 = |A^* x_0|$.

$$\boxed{u^{(k)} := -i(\alpha^{(k)} x^{(k)} - x_0) \longrightarrow \xi_1 \perp G(u^{(k)}), \quad \forall k}$$

Spectral gap (two patterns)

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

Proposition

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

Uniqueness theorem for magnitude retrieval

If

$$\measuredangle A^* \hat{x} = \pm \measuredangle A^* x_0$$

where the $\pm$ sign may be pixel-dependent, then almost surely $\hat{x} = c x_0$ for some constant $c \in \mathbb{R}$.

One random mask suffices !

Local convergence (PAP)

Theorem (PAP)

For any given $0 < \varepsilon < 1 - \lambda_2^2$*, if* $x^{(1)}$ *is sufficiently close to* x_0 *then with probability one the PAP iterates* $x^{(k)}$ *converges to* x_0 *geometrically after global phase adjustment, i.e.*

$$\|\alpha^{(k+1)}x^{(k+1)} - x_0\| \leq (\lambda_2^2 + \varepsilon)\|\alpha^{(k)}x^{(k)} - x_0\|, \quad \forall k$$

where $\alpha^{(k)} := \arg\min_\alpha\{\|\alpha x^{(k)} - x_0\| : |\alpha| = 1\}$.

Local convergence (SAP)

$$\|\alpha^{(k+1)}x^{(k+1)} - x_0\| \leq (\|\mathcal{D}\|_\perp + \epsilon)\|\alpha^{(k)}x^{(k)} - x_0\|,$$

Convergence rate $\quad \|\mathcal{D}\|_\perp \leq (\lambda_2^{(2)}\lambda_2^{(1)})^2.$

$$\lambda_2^{(l)} = \max\{\|\Im[B_l^* u]\| : u \in \mathbb{C}^n, iu \perp x_0, \|u\| = 1\}, \quad l = 1, 2$$

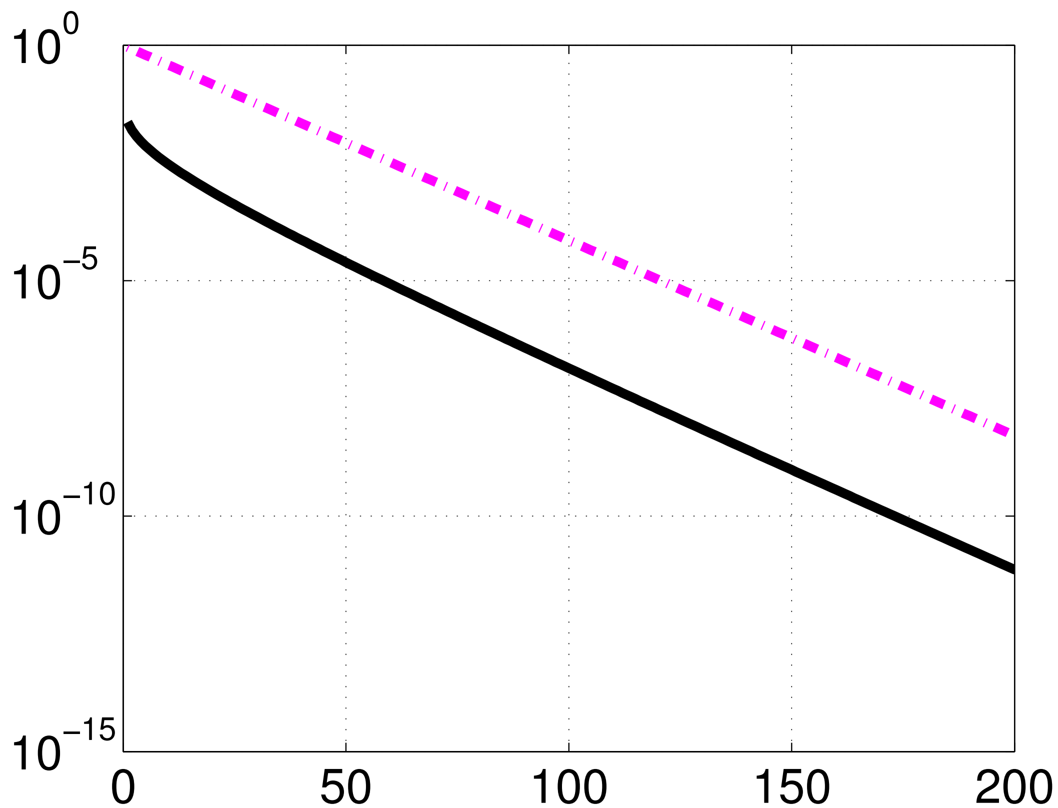
Local convergence (one pattern)

$$\|\alpha^{(k+1)}x^{(k+1)} - x_0\| \leq (\tilde{\lambda}_2^2 + \epsilon)\|\alpha^{(k)}x^{(k)} - x_0\|, \quad \forall k$$

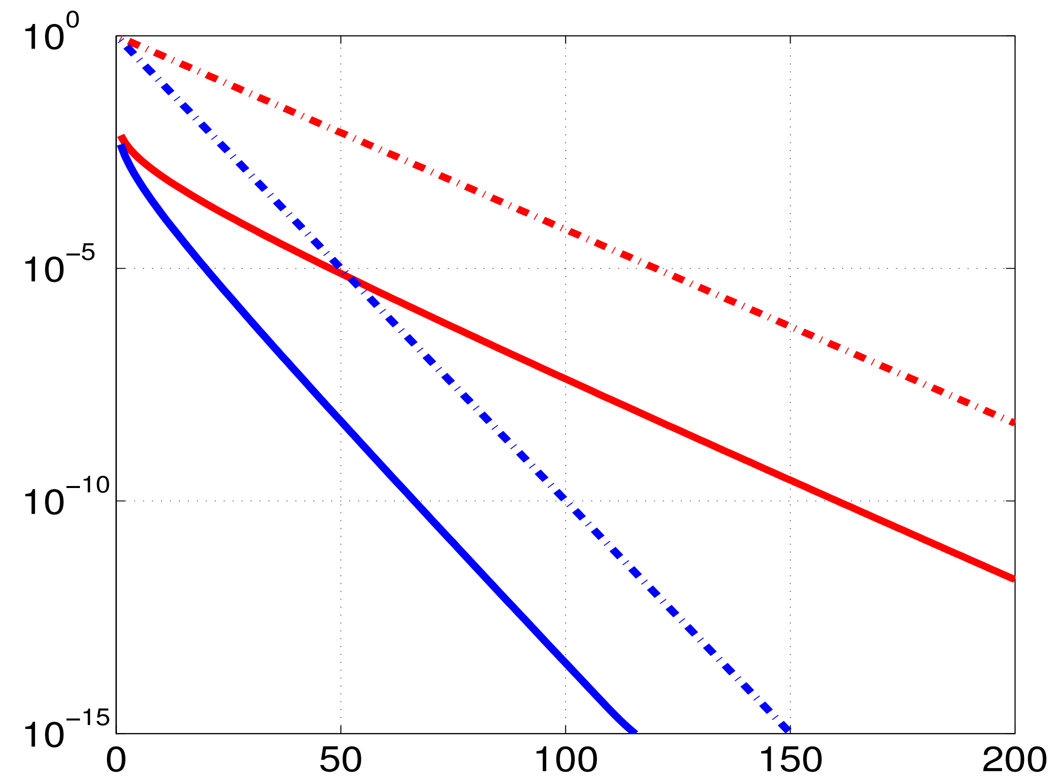
Convergence rate $\tilde{\lambda}_2 \leq \lambda_2$

$$\tilde{\lambda}_2 := \max\{\|\Im(B^*)u\| : u \in \mathbb{R}^n, \langle u, x_0\rangle = 0, \|u\| = 1\}$$

Convergence rate



(a) One-pattern PAP: $\tilde{\lambda}_2^2 = 0.9084$



(b) SAP $\|\mathcal{D}\|_\perp = 0.7946$; PAP $\lambda_2^2 = 0.9086$

$$\mathrm{RE} = \min_{\theta \in [0,2\pi)} \|x_0 - e^{i\theta} x\| / \|x_0\|$$

Initial guess

$$A^* = [a_j^*]$$

$$a_j^* x_0 = 0 \longrightarrow b_j = |a_j^* x_0| = a_j^* x_0.$$

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

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

$$x_{\text{dual}} := \arg\max \left\{ \|A_{I_c}^* x\|^2 : x \in \mathcal{X}, \|x\| = \|x_0\| \right\}$$

Isometry $\longrightarrow \|A_I^* x\|^2 + \|A_{I_c}^* x\|^2 = \|x\|^2$

$\longrightarrow x_{\text{null}} = x_{\text{dual}}$ power method

Null vector method

Theorem

Let $A \in \mathbb{C}^{n\times N}$ be an i.i.d. complex Gaussian matrix. Let $\kappa < 1$ be a fixed constant. Suppose

$$\sigma = \frac{|I|}{N} \le \kappa < 1, \quad \nu = \frac{n}{|I|} < 1.$$

Then for any $\varepsilon \in (0,1), \delta > 0$ and $t \in (0, \nu^{-1/2} - 1)$ the following error bound

$$\|x_0 x_0^* - x_{\mathrm{null}} x_{\mathrm{null}}^*\|^2 \le \left(\left(2 + \frac{t}{1-\epsilon}\right)\sigma + \varepsilon\left(-2\ln(1-\sigma) + \delta\right)\right) \left(1 - (1+t)\sqrt{\nu}\right)^{-2}$$

holds with probability at least

$$1 - 2\exp\left(-c\min(Nt^2/K^2, Nt/K)\right) - 4\exp\left(-nt^2/2\right)$$
$$-2\exp\left(-N\delta^2 e^{-\delta}|1-\sigma|^2/2\right) - 2\exp\left(-4\varepsilon^2|1-\sigma|^2\sigma^2 N\right)$$

where c is an absolute constant and $K \le -4\ln(1-\kappa)/\kappa$.

Scaling

ϵ and t fixed,

$$\boxed{n \gg 1, \quad \frac{n}{|I|} < 1, \quad \frac{|I|}{N} \ll 1, \quad \frac{|I|^2}{N} \gg 1.}$$

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

Algorithm 1: The null vector method

Random initialization: $x_1 = x_{\text{rand}}$
Loop:
for $k = 1 : k_{\max} - 1$ **do**
  $x'_k \leftarrow A(\mathbf{1}_c \odot A^* x_k)$;
  $x_{k+1} \leftarrow \left[x'_k\right]_{\mathcal{X}} / \| \left[x'_k\right]_{\mathcal{X}} \|$
end
Output: $x_{\text{null}} = x_{k_{\max}}$.

Algorithm 2: The spectral vector method

Random initialization: $x_1 = x_{\text{rand}}$
Loop:
for $k = 1 : k_{\max} - 1$ **do**
  $x'_k \leftarrow A(|b|^2 \odot A^* x_k)$;
  $x_{k+1} \leftarrow \left[x'_k\right]_{\mathcal{X}} / \| \left[x'_k\right]_{\mathcal{X}} \|$;
end
Output: $x_{\text{spec}} = x_{k_{\max}}$.

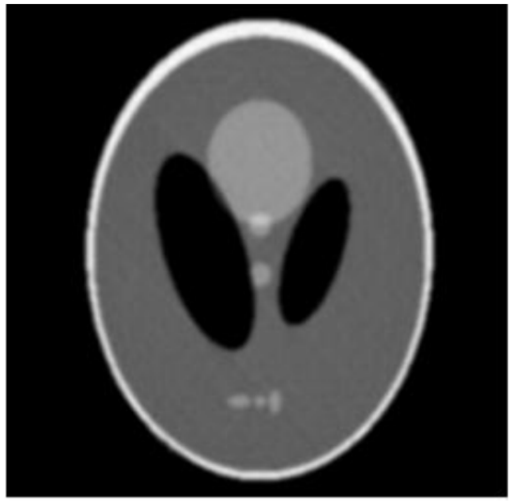
Netrapalli-Jain-Sanghavi 2015

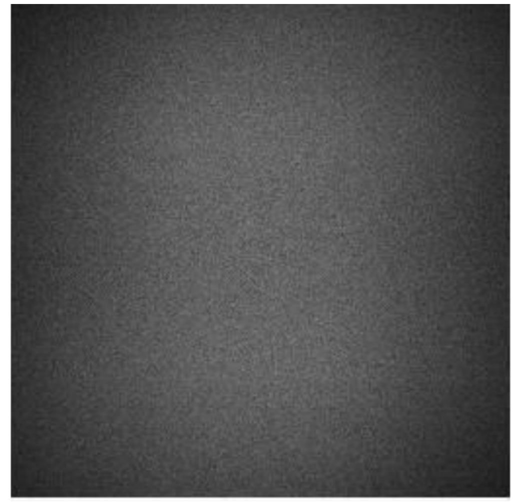
Truncated spectral vector

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

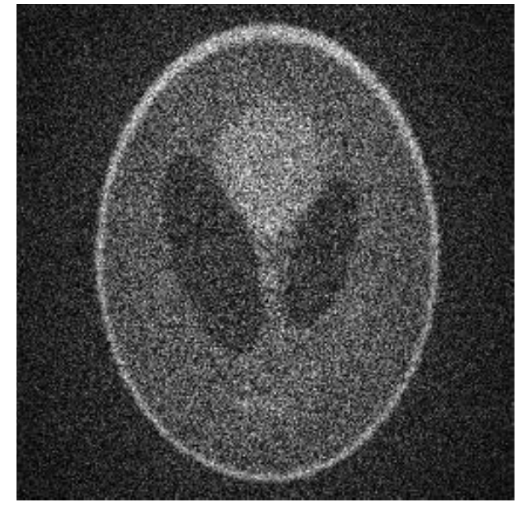
$$\{i : |A^* x(i)| \leq \tau \|b\|\}$$

Candes-Chen 2015

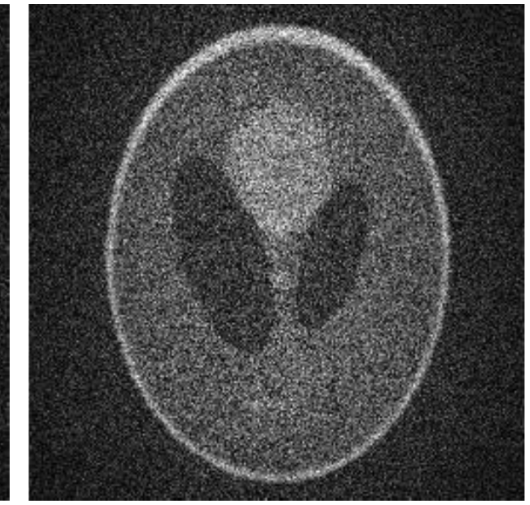




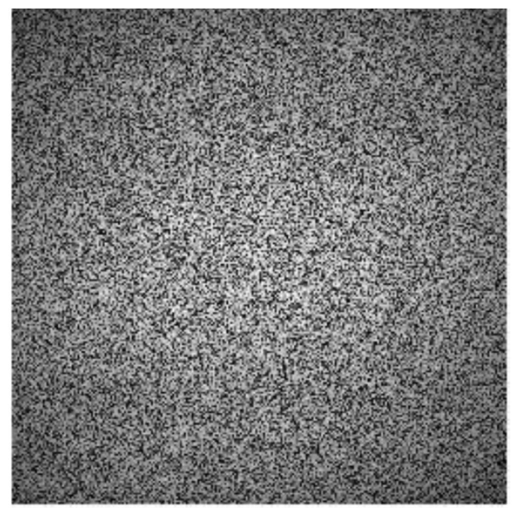
(a) $|x_{\text{t-spec}}|$ $(\tau^2 = 5)$



(b) $|x_{\text{null}}|$ $(\gamma = 0.5)$



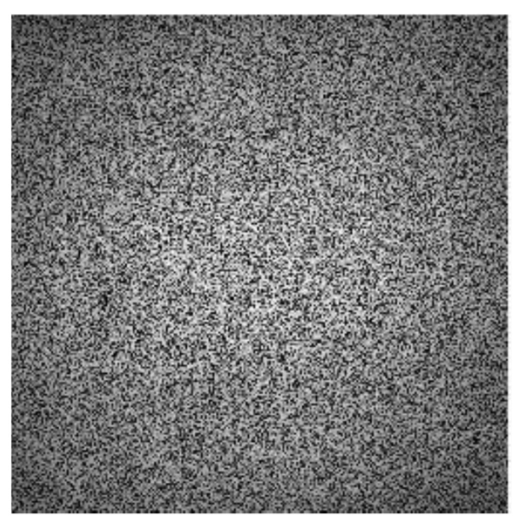
(c) $|x_{\text{null}}|$ $(\gamma = 0.6)$



(a) $|\mathrm{Re}(x_{\text{t-spec}})|$ $(\tau^2 = 5)$

(b) $|\mathrm{Re}(x_{\text{null}})|$ $(\gamma = 0.5)$

(c) $|\mathrm{Re}(x_{\text{null}})|$ $(\gamma = 0.63)$



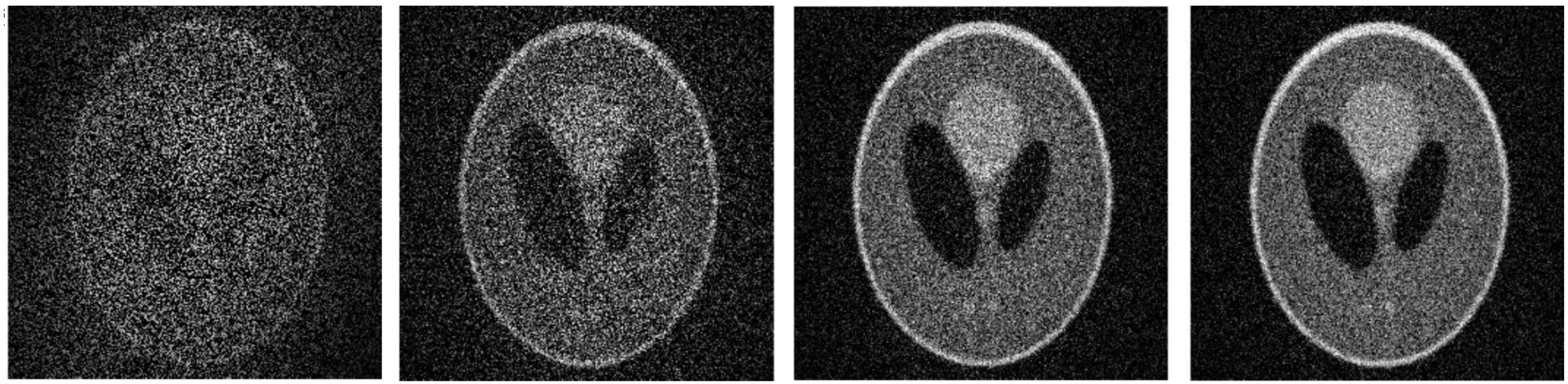
(d) $|\mathrm{Im}(x_{\text{t-spec}})|$ $(\tau^2 = 5)$

(e) $|\mathrm{Im}(x_{\text{null}})|$ $(\gamma = 0.5)$

(f) $|\mathrm{Im}(x_{\text{null}})|$ $(\gamma = 0.63)$

256 x 256

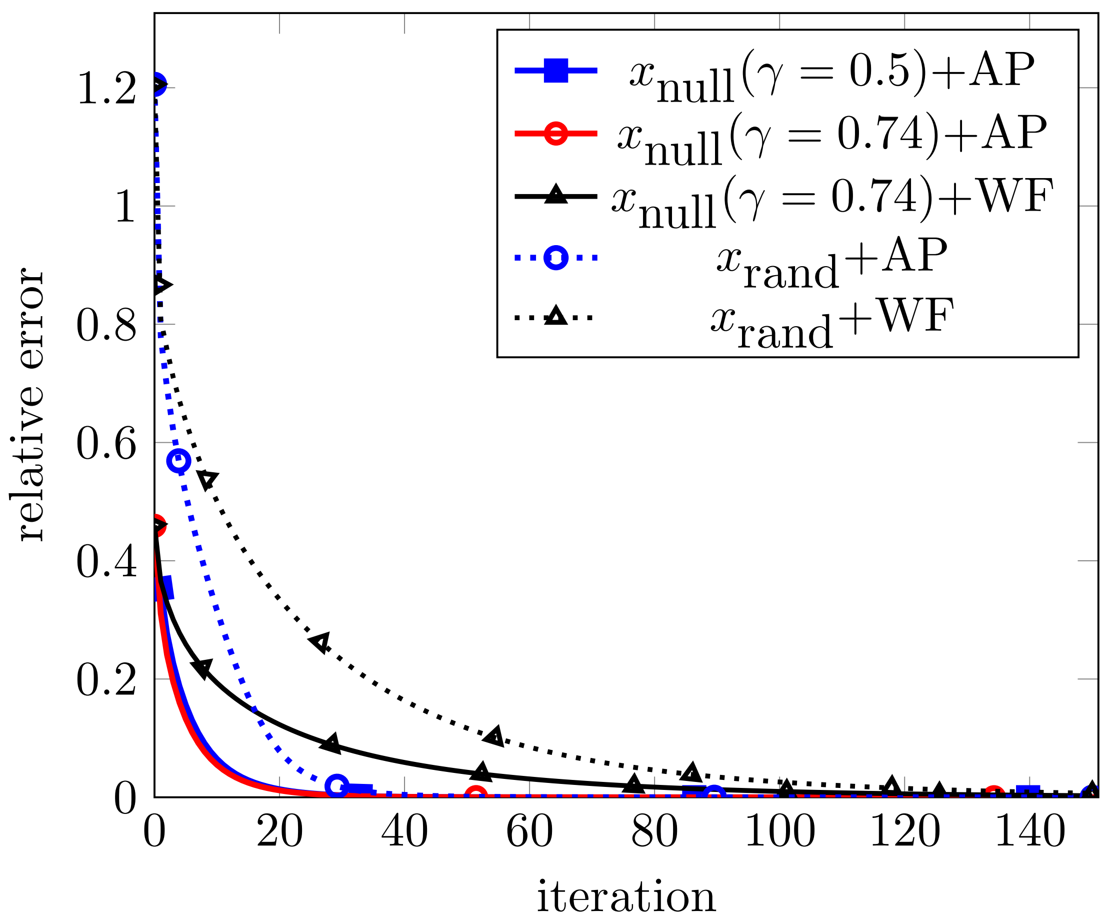
Two-pattern

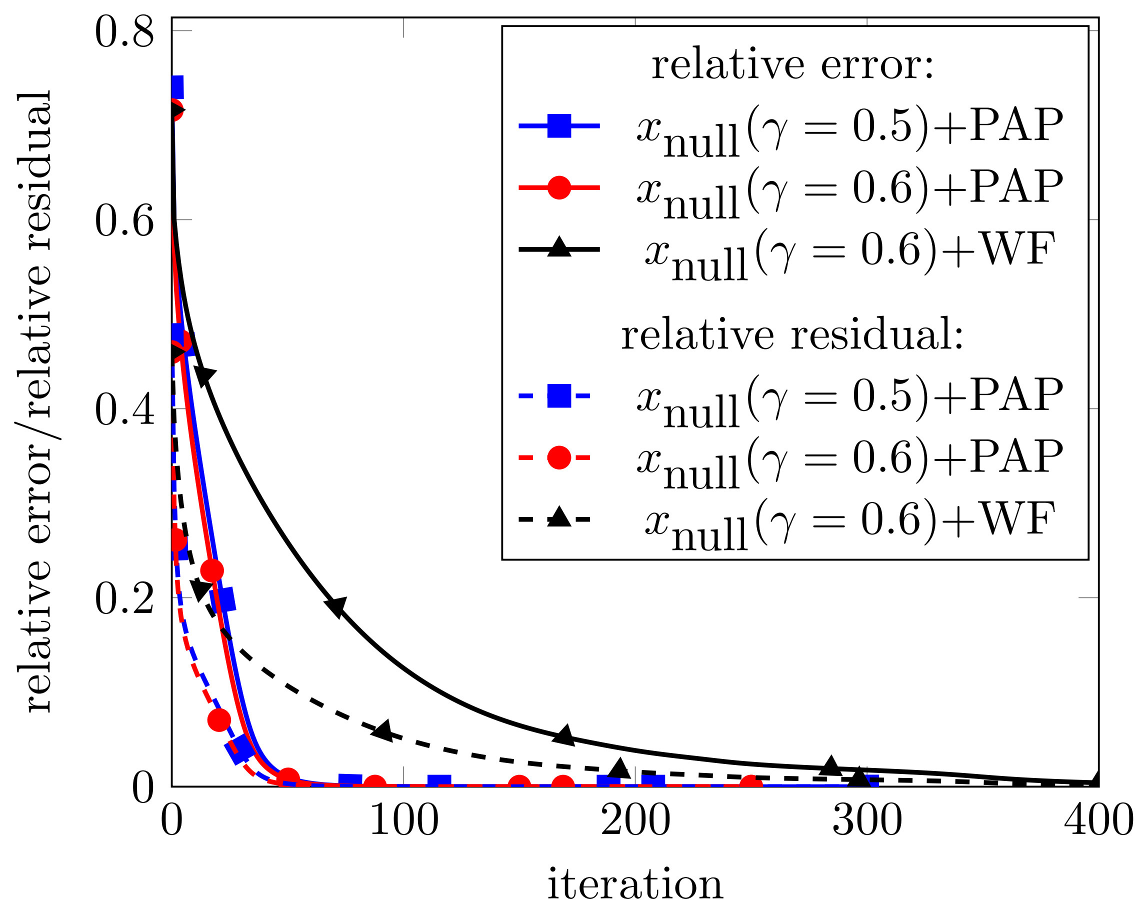


(a) x_{spec} (b) $x_{\text{t-spec}}$ $(\tau^2 = 4.6)$ (c) x_{null} $(\gamma = 0.5)$ (d) x_{null} $(\gamma = 0.74)$

(a) x_{spec} (b) $x_{\text{t-spec}}$ $(\tau^2 = 4.1)$ (c) x_{null} $(\gamma = 0.5)$ (d) x_{null} $(\gamma = 0.7)$

One-pattern





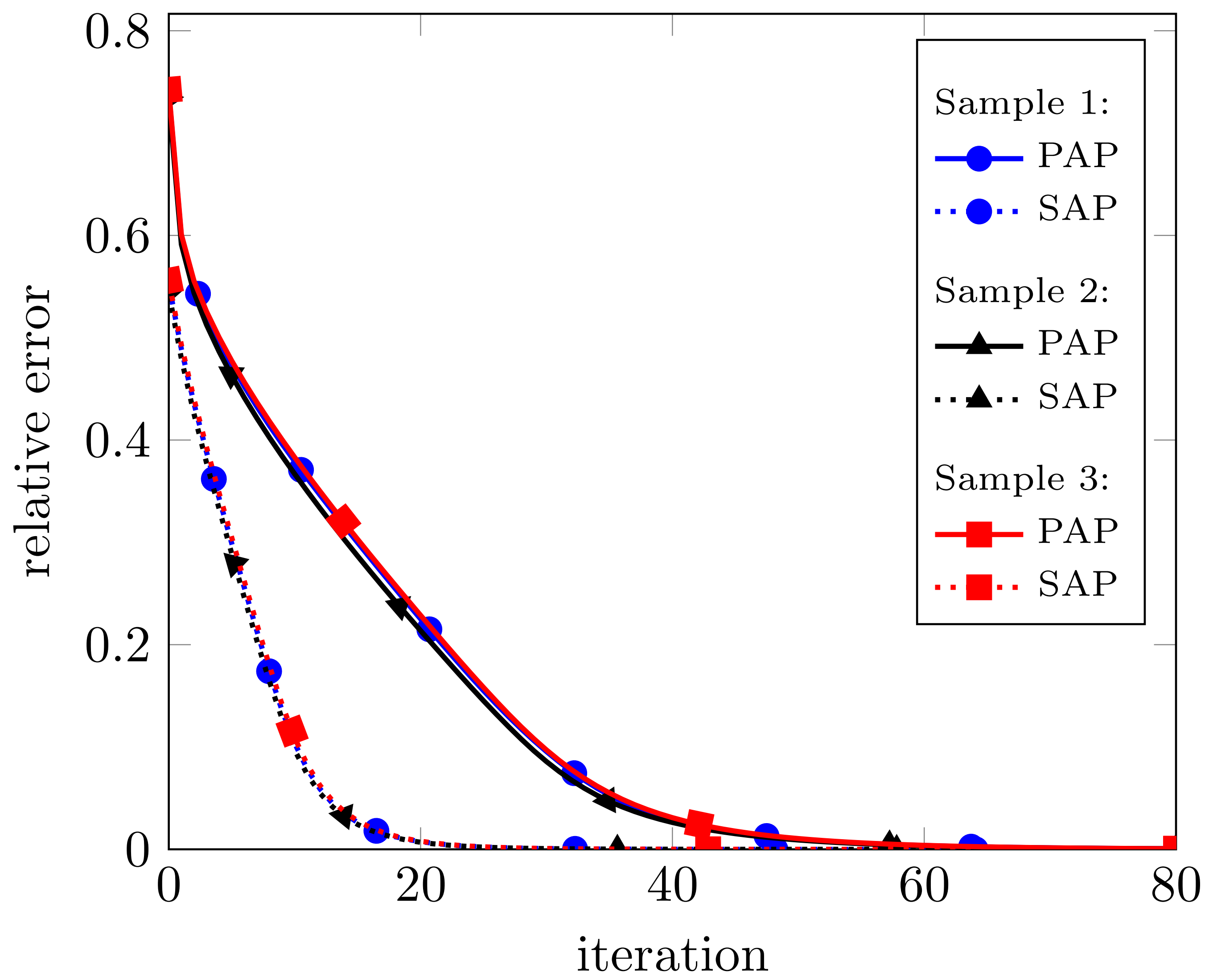
One-pattern

Two-pattern

WF (Wirtinger Flow): Candes-Li-Soltanolkotabi 2015

WF run with an optimized constant step size.

PAP vs SAP



Two-pattern

Fourier Douglas-Rachford

$$R_1 = 2P_1 - I \qquad\qquad R_2 = 2P_2 - I$$

Averaged Alternating Reflection

$$\begin{aligned} y^{(k+1)} &:= \frac{1}{2}(I + R_1 R_2) y^{(k)} \\ &= y^{(k)} + P_1(2P_2 - I)y^{(k)} - P_2 y^{(k)}, \quad k = 1, 2, 3 \cdots \end{aligned}$$

$$\boxed{S_{\mathrm{f}}(y) \;=\; y + A^*\left[A\left(2b \odot \frac{y}{|y|} - y\right)\right]_{\mathcal{X}} - b \odot \frac{y}{|y|}}$$

Gradient $\quad J_{\mathrm{f}} v = (I - B^* B)\Re(v) + iB^* B \Im(v)$

J_{f} is a *real*, but *not complex*, linear map

Convergence to unique fixed point

$$S_{\mathrm{f}}(y_\infty) = y_\infty, \quad x_\infty = A y_\infty.$$

$$\boxed{\begin{array}{l} y_\infty = e^{i\theta}(|y_0| + v) \odot \dfrac{y_0}{|y_0|} \\ |y_0| + v \textit{ has all nonnegative components} \\ \quad v \in \mathrm{null}_{\mathbb{R}}(\mathcal{B}) \subset \mathbb{R}^N \end{array}}$$

Theorem (two-pattern)

Almost surely $\hat{x} = x_\infty = e^{i\theta} x_0$ *for some constant* $\theta \in \mathbb{R}$.

Local convergence

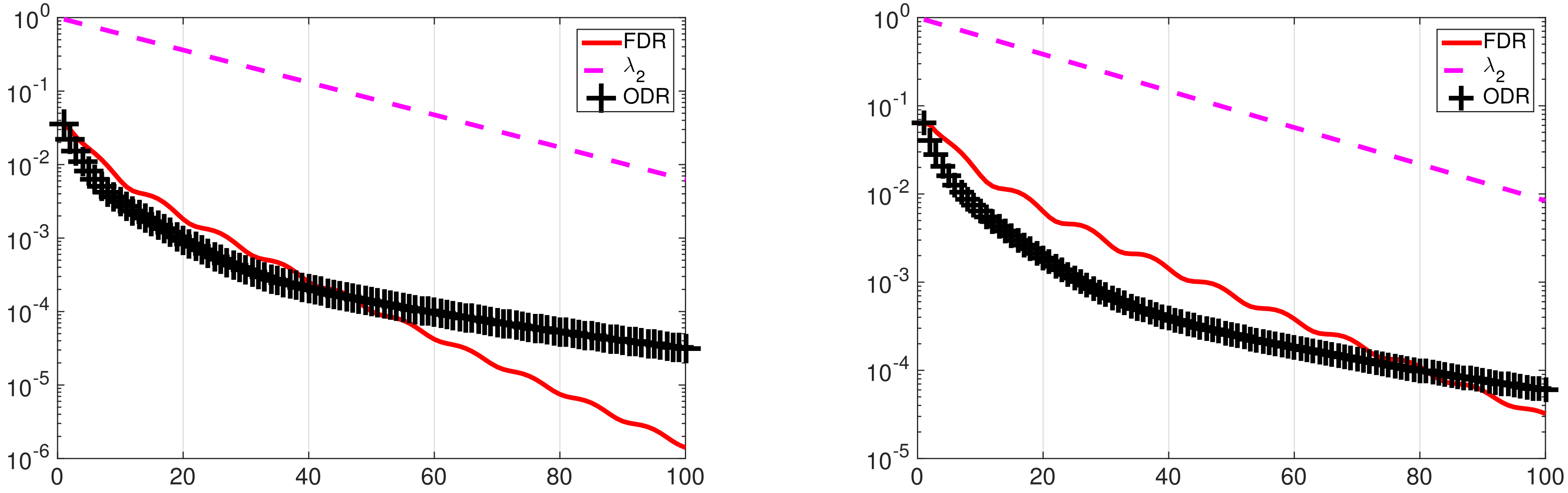
Theorem (two-pattern)

For any given $0 < \varepsilon < 1 - \lambda_2$*, if* $x^{(1)}$ *is sufficiently close to* x_0 *then with probability one* $x^{(k)}$ *converges to* x_0 *geometrically after global phase adjustment, i.e.*

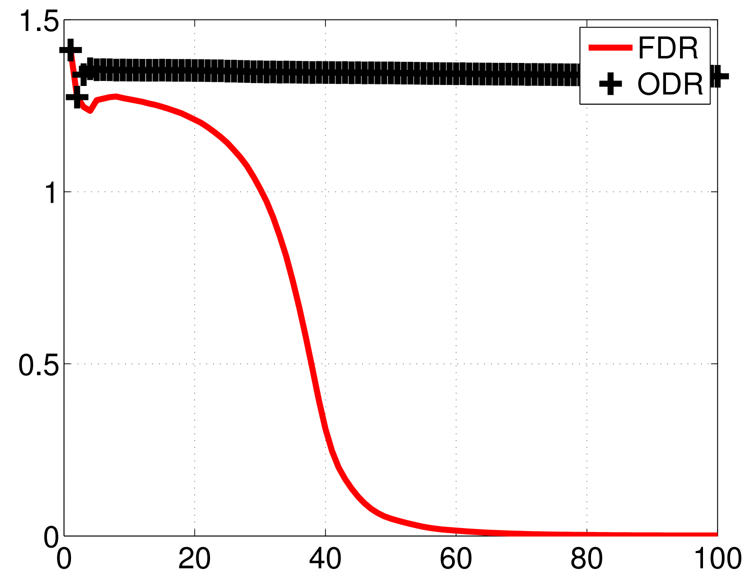
$$\|\alpha^{(k+1)}x^{(k+1)} - x_0\| \leq (\lambda_2 + \varepsilon)\|\alpha^{(k)}x^{(k)} - x_0\|, \quad \forall k$$

where $\alpha^{(k)} := \arg\min_\alpha\{\|\alpha x^{(k)} - x_0\| : |\alpha| = 1\}$.

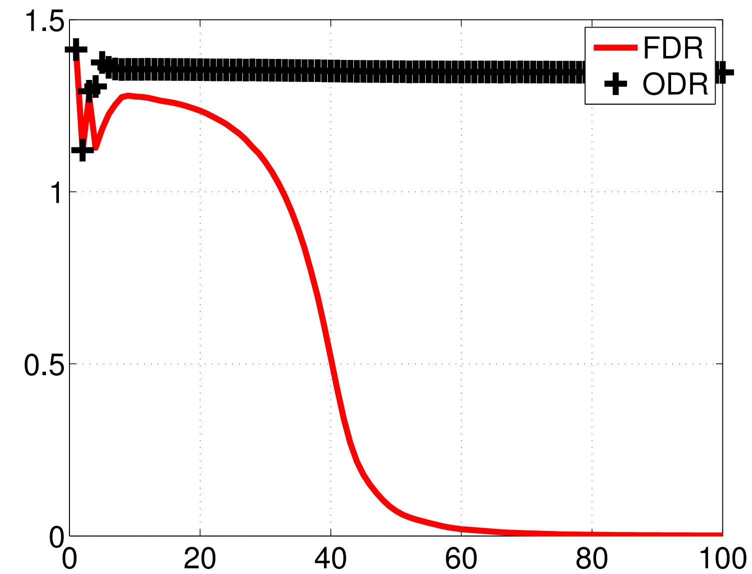
TCB



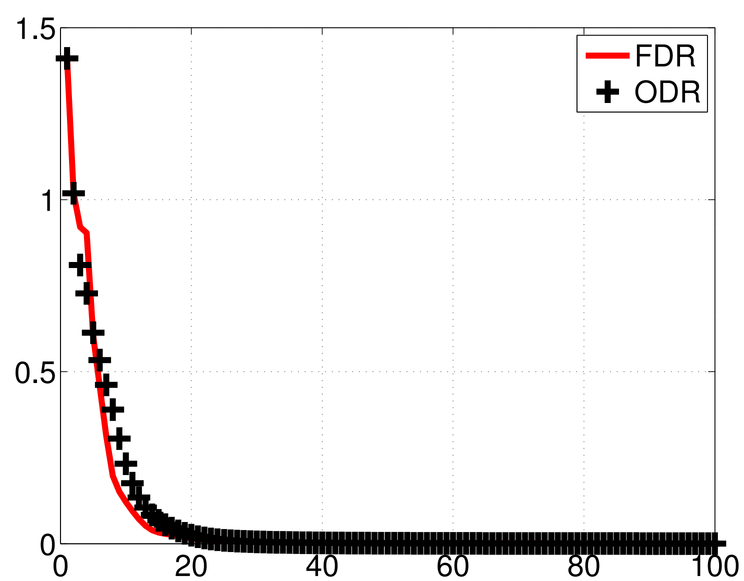
(a) TCB (b) RPP



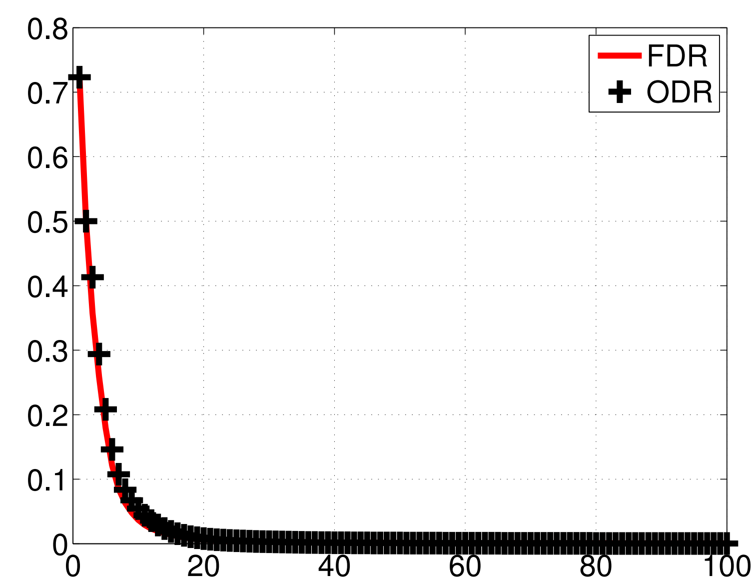
(a) RPP + RI



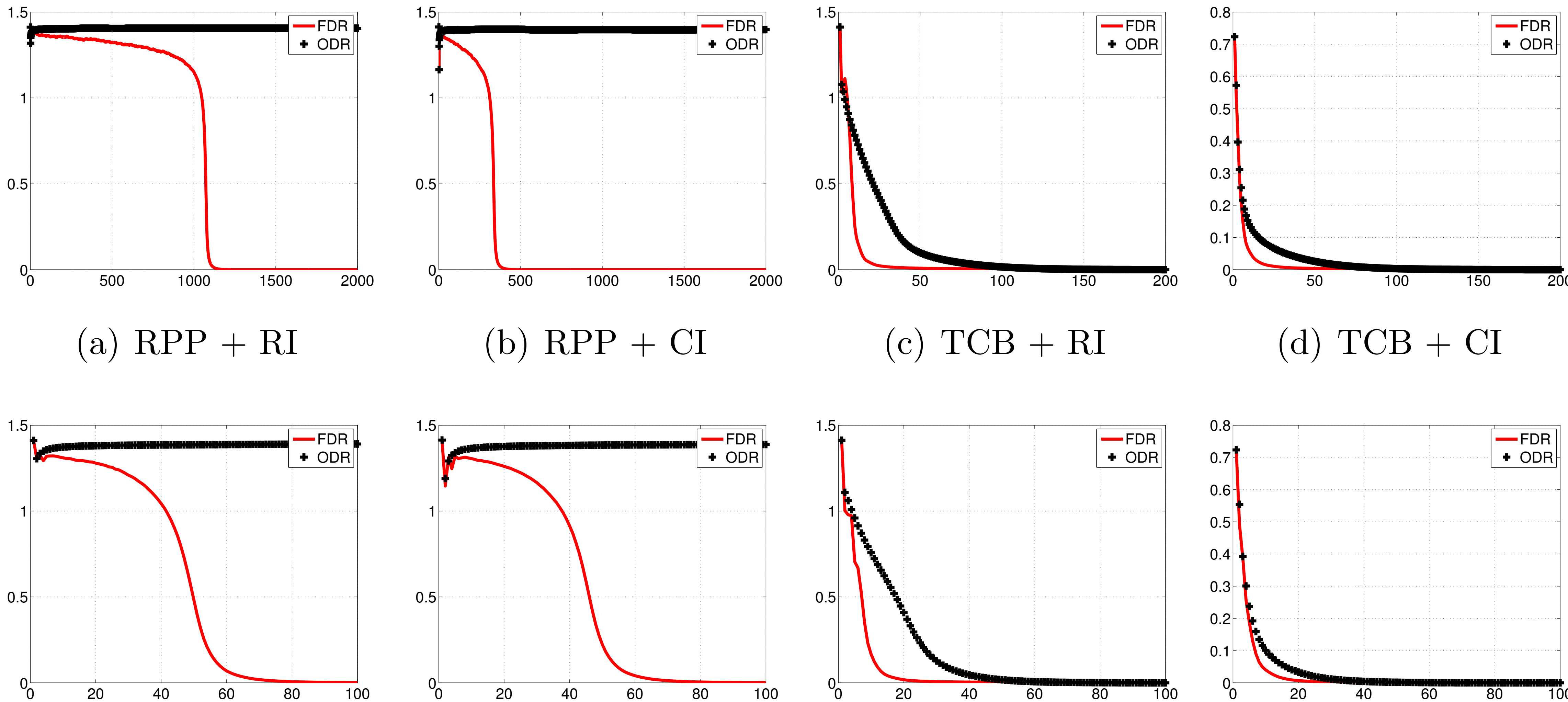
(b) RPP + CI



(c) TCB + RI



(d) TCB + CI

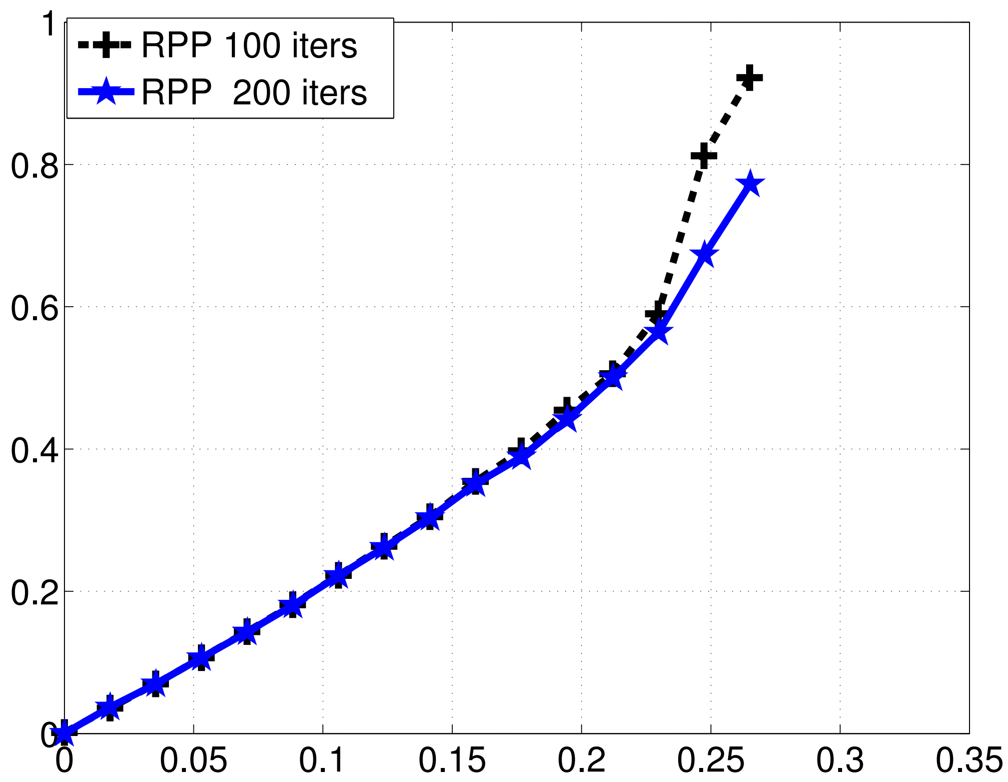


(a) RPP + RI (b) RPP + CI (c) TCB + RI (d) TCB + CI

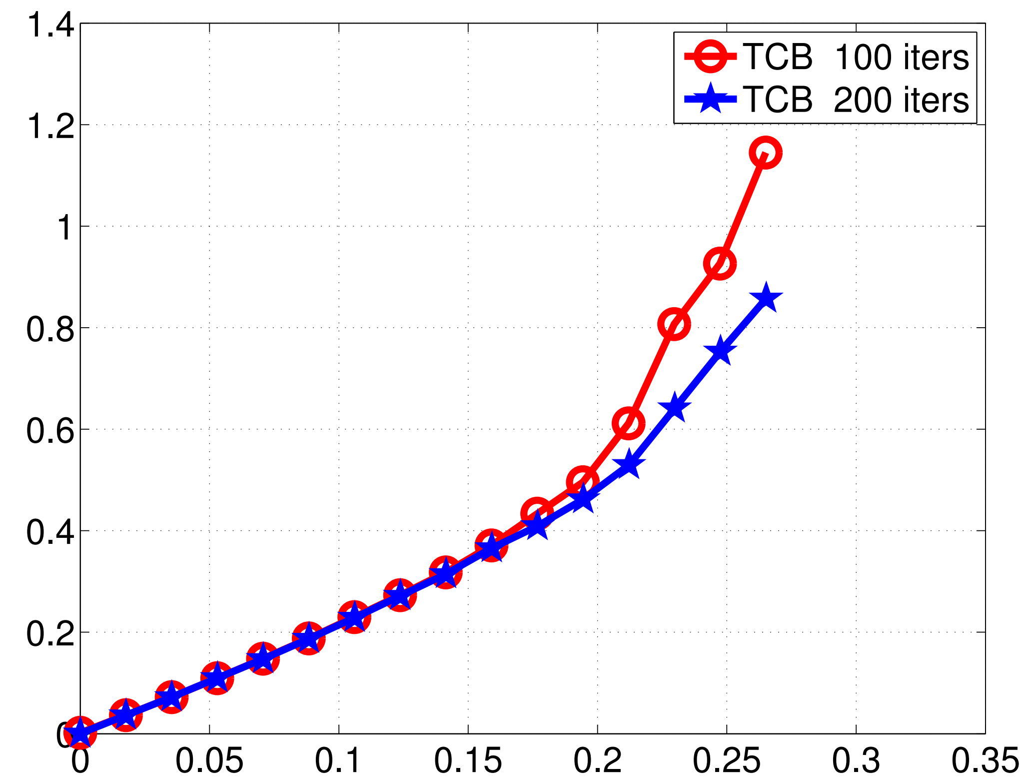
(e) RPP + RI (f) RPP + CI (g) TCB + RI (h) TCB + CI

FIGURE 7. Relative error versus iteration with 3 patterns (a)-(d) and 4 patterns (e)-(h) (without oversampling in each pattern).

Relative Error vs. Noise-to-Signal Ratio



(a) RPP



(b) TCB

$$\text{NSR} = \frac{\|\epsilon\|}{\|A^* x_0\|}.$$

Conclusions

- Alternating Projections (PAP/SAP) of the Null Vector
- Fourier Domain Douglas-Rachford (FDR)
- Performance guarantee: Local convergence, global convergence of the null vector method, uniqueness of fixed point.
- Global convergence for FDR?
- Noise stability
- Single molecule imaging: extremely noisy measurements.