# A Geometric Analysis of Phase Retrieval

**Qing Qu**

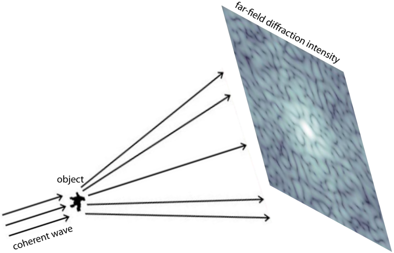
Electrical Engineering

Columbia University

Joint with **Ju Sun**, **John Wright** (Columbia University)

# Missing phase problem

**Phase retrieval**: Given phaseless information of a complex signal, recover the signal




**Coherent Diffraction Imaging**[1]

**Applications**: X-ray crystallography, **diffraction imaging** (left), optics, astronomical imaging, and microscopy

For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{F}\boldsymbol{x}|$, recover $\boldsymbol{x}$.

[1] Image courtesy of [Shechtman et al., 2015]

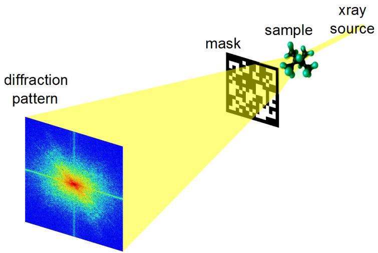
# Generalized phase retrieval

For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given $|\mathcal{F}\boldsymbol{x}|$, recover $\boldsymbol{x}$.

**Generalized phase retrieval**:

For a complex signal $\boldsymbol{x} \in \mathbb{C}^n$, given *generalized* measurements of the form $|\boldsymbol{a}_k^* \boldsymbol{x}|$ for $k = 1, \ldots, m$, recover $\boldsymbol{x}$.

... in practice, generalized measurements by design such as masking, grating, structured illumination, etc [2]



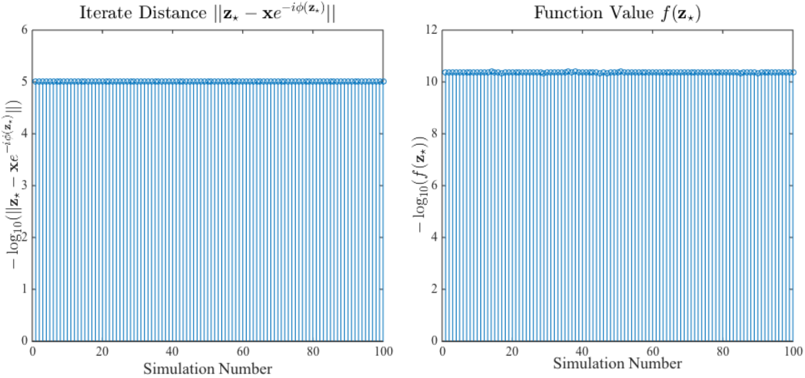

[2] Image courtesy of [Candès et al., 2015b]

# A nonconvex formulation

- Given $y_k = |\boldsymbol{a}_k^* \boldsymbol{x}|$ for $k = 1, \ldots, m$, recover $\boldsymbol{x}$ (**up to a global phase**).
- A natural **nonconvex** formulation (see also [Candès et al., 2015b])

$$\boxed{\min_{\boldsymbol{z} \in \mathbb{C}^n} f(\boldsymbol{z}) \doteq \frac{1}{2m} \sum_{k=1}^{m} (y_k^2 - |\boldsymbol{a}_k^* \boldsymbol{z}|^2)^2.}$$

# A Curious Experiment



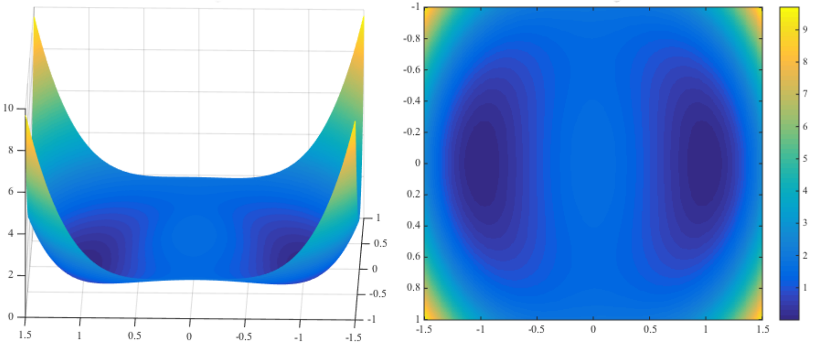

- Assume $\{\boldsymbol{a}_k\}_{k=1}^m$ are i.i.d. complex Gaussian

$$\boldsymbol{a}_k = (X_k + \mathrm{i}Y_k)/\sqrt{2}, \quad X_k,\ Y_k \sim N(\mathbf{0}, \boldsymbol{I}_n).$$

- Run **gradient descent** on the nonconvex formulation $f(\boldsymbol{z})$

$$\boldsymbol{z}^{r+1} = \boldsymbol{z}^r - \mu\nabla_{\boldsymbol{z}} f(\boldsymbol{z}^{(r)}).$$

# Visualization of low-dimensional function landscape

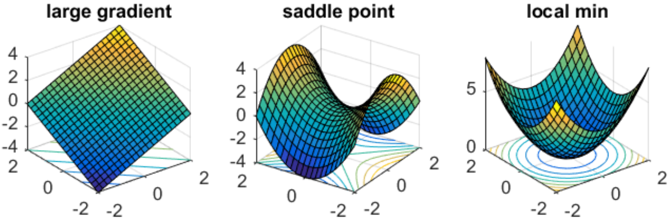



- Plot of $f(\boldsymbol{z})$ with $\boldsymbol{z} \in \mathbb{R}^2$, global minimizer $\boldsymbol{x} = [1; 0]$ and $m \to +\infty$.
- All the critical points are $\pm\boldsymbol{x}$, $\mathbf{0}$, and saddle points $\pm[0; 1/\sqrt{2}]$.
- The saddle points can be escaped following the negative curvature direction along $\pm\boldsymbol{x}$.

# A glimpse into high dimensional geometry

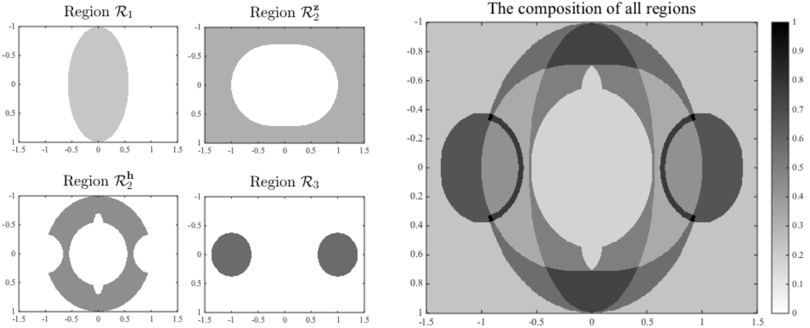
When $m \geq \Omega(n \log^3 n)$, it holds with probability at least $1 - cm^{-1}$

- The only local/global minimizers of $f(\boldsymbol{z})$ are the solution set $\mathcal{X} = \left\{ \boldsymbol{x} e^{\mathrm{i}\theta} : \theta \in [0, 2\pi) \right\}$.
- The whole space $\mathbb{C}^n$ can be partitioned into **negative curvature** $\mathcal{R}_1$**, large gradient** $\mathcal{R}_2^{\boldsymbol{z}}$, $\mathcal{R}_2^{\boldsymbol{h}}$ **, and strong convexity** $\mathcal{R}_3$ regions.



Quantatively,

$$
\begin{aligned}
\begin{bmatrix} \frac{\boldsymbol{x}\mathrm{e}^{\mathrm{i}\phi(\boldsymbol{z})}}{\boldsymbol{x}\mathrm{e}^{\mathrm{i}\phi(\boldsymbol{z})}} \end{bmatrix}^* \nabla^2 f(\boldsymbol{z}) \begin{bmatrix} \frac{\boldsymbol{x}\mathrm{e}^{\mathrm{i}\phi(\boldsymbol{z})}}{\boldsymbol{x}\mathrm{e}^{\mathrm{i}\phi(\boldsymbol{z})}} \end{bmatrix} &\le -C_1 \left\|\boldsymbol{x}\right\|^4, && \forall \boldsymbol{z} \in \mathcal{R}_1, \\
\frac{\boldsymbol{z}^* \nabla_{\boldsymbol{z}} f(\boldsymbol{z})}{\left\|\boldsymbol{z}\right\|} &\ge C_2 \left\|\boldsymbol{x}\right\|^2 \left\|\boldsymbol{z}\right\|, && \forall \boldsymbol{z} \in \mathcal{R}_2^{\boldsymbol{z}}, \\
\frac{\mathbb{R}\left(\boldsymbol{h}(\boldsymbol{z})^* \nabla_{\boldsymbol{z}} f(\boldsymbol{z})\right)}{\left\|\boldsymbol{h}(\boldsymbol{z})\right\|} &\ge C_3 \left\|\boldsymbol{x}\right\|^2 \left\|\boldsymbol{z}\right\|, && \forall \boldsymbol{z} \in \mathcal{R}_2^{\boldsymbol{h}}, \\
\begin{bmatrix} \frac{\boldsymbol{g}(\boldsymbol{z})}{\boldsymbol{g}(\boldsymbol{z})} \end{bmatrix}^* \nabla^2 f(\boldsymbol{z}) \begin{bmatrix} \frac{\boldsymbol{g}(\boldsymbol{z})}{\boldsymbol{g}(\boldsymbol{z})} \end{bmatrix} &\ge C_4 \left\|\boldsymbol{x}\right\|^2, && \forall \boldsymbol{z} \in \mathcal{R}_3,
\end{aligned}
$$




$$\mathcal{R}_1 \doteq \left\{ \boldsymbol{z} : 8 \left| \boldsymbol{x}^* \boldsymbol{z} \right|^2 + \frac{401}{100} \left\| \boldsymbol{x} \right\|^2 \left\| \boldsymbol{z} \right\|^2 \leq \frac{398}{100} \left\| \boldsymbol{x} \right\|^4 \right\},$$

$$\mathcal{R}_2^{\boldsymbol{z}} \doteq \left\{ \boldsymbol{z} : \mathbb{R} \left( \langle \boldsymbol{z}, \nabla_{\boldsymbol{z}} \mathbb{E} \left[ f \right] \rangle \right) \geq \frac{1}{100} \left\| \boldsymbol{z} \right\|^4 + \frac{1}{500} \left\| \boldsymbol{x} \right\|^2 \left\| \boldsymbol{z} \right\|^2 \right\},$$

$$\mathcal{R}_2^{\boldsymbol{h}} \doteq \left\{ \boldsymbol{z} : \frac{11}{20} \left\| \boldsymbol{x} \right\| \leq \left\| \boldsymbol{z} \right\| \leq \left\| \boldsymbol{x} \right\|, \operatorname{dist}(\boldsymbol{z}, X) \geq \frac{\left\| \boldsymbol{x} \right\|}{3}, \right.$$

$$\left. \mathbb{R} \left( \langle \boldsymbol{h}(\boldsymbol{z}), \nabla_{\boldsymbol{z}} \mathbb{E} \left[ f \right] \rangle \right) \geq \frac{1}{250} \left\| \boldsymbol{x} \right\|^2 \left\| \boldsymbol{z} \right\| \left\| \boldsymbol{h}(\boldsymbol{z}) \right\| \right\},$$

$$\mathcal{R}_3 \doteq \left\{ \boldsymbol{z} : \operatorname{dist}(\boldsymbol{z}, X) \leq \left\| \boldsymbol{x} \right\| / \sqrt{7} \right\}.$$

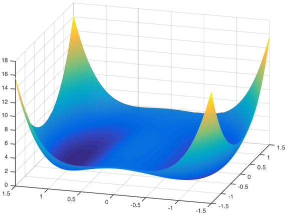
# Ridable saddle function

**Ridable-saddle (strict-saddle) functions** A function $f : \mathcal{M} \mapsto \mathbb{R}$ is $(\alpha, \beta, \gamma, \delta)$-ridable ($\alpha,\ \beta,\ \gamma,\ \delta > 0$ ) if any point $\boldsymbol{x} \in \mathcal{M}$ obeys **at least one of the following**:

1) [**Strong gradient**] $\|\mathrm{grad}\, f(\boldsymbol{x})\| \geq \beta$;
2) [**Negative curvature**] There exists $\boldsymbol{v} \in T_{\boldsymbol{x}}\mathcal{M}$ with $\|\boldsymbol{v}\| = 1$ such that $\langle \mathrm{Hess}\, f(\boldsymbol{x})[\boldsymbol{v}], \boldsymbol{v} \rangle \leq -\alpha$;
3) [**Strong convexity around minimizers**] There exists a local minimizer $\boldsymbol{x}_\star$ such that $\|\boldsymbol{x} - \boldsymbol{x}_\star\| \leq \delta$, and for all $\boldsymbol{y} \in \mathcal{M}$ that is in $2\delta$ neighborhood of $\boldsymbol{x}_\star$, $\langle \mathrm{Hess}\, f(\boldsymbol{y})[\boldsymbol{v}], \boldsymbol{v} \rangle \geq \gamma$ for any $\boldsymbol{v} \in T_{\boldsymbol{y}}\mathcal{M}$ with $\|\boldsymbol{v}\| = 1$.

($T_{\boldsymbol{x}}\mathcal{M}$ is the tangent space of $\mathcal{M}$ at point $\boldsymbol{x}$)

# Summary of the results




$$\boxed{\min_{\boldsymbol{z}\in\mathbb{C}^n} f(\boldsymbol{z}) \doteq \frac{1}{2m}\sum_{k=1}^{m}(y_k^2 - |\boldsymbol{a}_k^* \boldsymbol{z}|^2)^2.}$$

**Theorem (Informal, Sun, Q., Wright '16)**

*Let* $\boldsymbol{a}_k \sim_{\text{iid}} \mathcal{CN}(0,1)$. *When* $m \geq \Omega(n\log^3(n))$, *w.h.p.,*

- *All local (and global) minimizers are of the form* $\boldsymbol{x}e^{i\phi}$.
- $f$ *is* $(c, c/(n\log m), c, c/(n\log m))$*-ridable over* $\mathbb{C}^n$ *for some* $c > 0$.

# Comparison with the literature

- **SDP relaxations and their analysis**:

| | |
|---|---|
| [Candès et al., 2013a] | SDP relaxation |
| [Candès et al., 2013b] | Guarantees for $m \sim n \log n$, adaptive |
| [Candès and Li, 2014] | Guarantees for $m \sim n$, non-adaptive |
| [Candès et al., 2015a] | Coded diffraction patterns |
| [Waldspurger et al., 2015] | SDP relaxation in phase space |

- **Nonconvex methods** (spectral init. + local refinement):

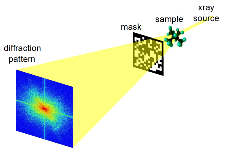
| | |
|---|---|
| [Netrapalli et al., 2013] | Spectral init. sample splitting |
| [Candès et al., 2015b] | Spectral init. + gradient descent, $m \sim n \log n$. |
| [White et al., 2015] | Spectral init. + gradient descent |
| [Chen and Candès, 2015] | Spectral init. + truncation, $m \sim n$. |

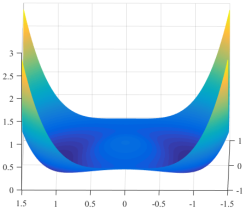
This work: a global characterization of the geometry of the problem. Algorithms succeed independent of initialization, $m \sim n \log^3 n$.

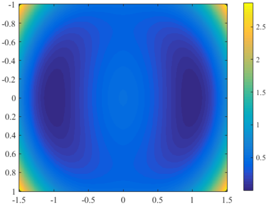
# Other measurement models for GPR

Other measurements

- Coded diffraction model [Candès et al., 2015a]





- Convolutional model (with Yonina Eldar): $\boldsymbol{y} = |\boldsymbol{a} \circledast \boldsymbol{x}|$

# Generalization to low-rank matrix recovery/completion

Suppose $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ with $\mathrm{rank}(\boldsymbol{X}) = r(r \ll n)$ and $\boldsymbol{X} \succeq \mathbf{0}$, can we recovery $\boldsymbol{X}$ given linear measurement:

$$y_k \ = \ \mathrm{tr}(\boldsymbol{A}_k\boldsymbol{X}) \quad (1 \leq k \leq m) \quad ?$$

# Generalization to low-rank matrix recovery/completion

Suppose $\boldsymbol{X} \in \mathbb{R}^{n \times n}$ with $\operatorname{rank}(\boldsymbol{X}) = r (r \ll n)$ and $\boldsymbol{X} \succeq \mathbf{0}$, can we recovery $\boldsymbol{X}$ given linear measurement:

$$y_k = \operatorname{tr}(\boldsymbol{A}_k \boldsymbol{X}) \quad (1 \leq k \leq m) \quad ?$$

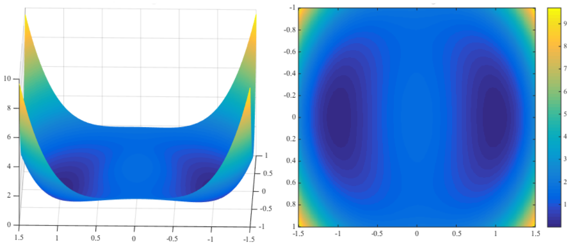
**Nonconvex formulation:** consider the factorization $\boldsymbol{X} = \boldsymbol{F}\boldsymbol{F}^*$ with $\boldsymbol{F} \in \mathbb{R}^{n \times r}$, solve

$$\boxed{\min_{\boldsymbol{F} \in \mathbb{R}^{n \times r}} g(\boldsymbol{F}) \doteq \frac{1}{2m} \sum_{k=1}^{m} \left(y_k - \operatorname{tr}\left(\boldsymbol{F}^* \boldsymbol{A}_k \boldsymbol{F}\right)\right)^2}.$$

In the matrix completion/recovery setting, [Ge et al., 2016, Bhojanapalli et al., 2016] show that

- All local minimizers are global.
- The function $g(\boldsymbol{F})$ is also a ridable saddle function.

# Algorithmic possibilities




- **Second-order trust-region method** (described here, [Conn et al., 2000], [Nesterov and Polyak, 2006])
- Curvilinear search [Goldfarb, 1980]
- Noisy/stochastic gradient descent [Ge et al., 2015]
- ...

# Second-order methods can escape ridable saddles

Taylor expansion at a saddle point $\boldsymbol{x}$:

$$\widehat{f}(\boldsymbol{\delta}; \boldsymbol{x}) = f(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{\delta}^* \nabla^2 f(\boldsymbol{x}) \boldsymbol{\delta}.$$

Choosing $\boldsymbol{\delta} = \boldsymbol{v}_{\mathrm{neg}}$, then

$$\widehat{f}(\boldsymbol{\delta}; \boldsymbol{x}) - f(\boldsymbol{x}) \leq -\frac{1}{2}|\lambda_{\mathrm{neg}}| \|\boldsymbol{v}_{\mathrm{neg}}\|^2.$$

Guaranteed decrease in $f$ when **movement is small** such that the **approximation is reasonably good**.

# Trust-region method - Euclidean Space

Generate iterates $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots$ by

- Forming a second order approximation of the objective $f(\boldsymbol{x})$ about $\boldsymbol{x}_k$:

$$\widehat{f}(\boldsymbol{\delta}; \boldsymbol{x}_k) = f(\boldsymbol{x}_k) + \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{\delta} \rangle + \frac{1}{2} \boldsymbol{\delta}^* \boldsymbol{B}_k \boldsymbol{\delta}.$$

  and minimizing the approximation within a small radius - the trust region

$$\boldsymbol{\delta}_\star \in \operatorname*{arg\,min}_{\|\boldsymbol{\delta}\| \le \Delta} \widehat{f}(\boldsymbol{\delta}; \boldsymbol{x}_k) \qquad (\textbf{Trust-region subproblem})$$

- Next iterate is $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{\delta}_\star$.

Can choose $\boldsymbol{B}_k = \nabla^2 f(\boldsymbol{x}^{(k)})$ or an approximation.

# The trust-region subproblem

$$\boldsymbol{\delta}_\star \in \operatorname*{arg\,min}_{\|\boldsymbol{\delta}\| \leq \Delta} \widehat{f}(\boldsymbol{\delta}; \boldsymbol{x}_k) \qquad (\textbf{Trust-region subproblem})$$

- QCQP, but can be solved in polynomial time by:

  Root finding [Moré and Sorensen, 1983]

  SDP relaxation [Rendl and Wolkowicz, 1997].

- In practice, only need an approximate solution (with controllable quality) to ensure convergence.

# Proof of convergence

- Strong gradient or negative curvature
  $\implies$ at least a fixed reduction in $f(\boldsymbol{x})$ at each iteration
- Strong convexity near a local minimizer
  $\implies$ quadratic convergence $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_\star\| \le c \|\boldsymbol{x}_k - \boldsymbol{x}_\star\|^2$.

**Theorem (Very informal)**

*For ridable-saddle functions, starting from an **arbitrary initialization**, the iteration sequence with **sufficiently small** trust-region size converges to a local minimizer in **polynomial number of steps**.*

Worked out examples in [Sun et al., 2015, Sun et al., 2016];

See also promise of 1-st order method [Ge et al., 2015, Lee et al., 2016].

# References I


[Bhojanapalli et al., 2016] Bhojanapalli, S., Neyshabur, B., and Srebro, N. (2016). **Global optimality of local search for low rank matrix recovery.** *arXiv preprint arXiv:1605.07221*.

[Candès et al., 2013a] Candès, E. J., Eldar, Y. C., Strohmer, T., and Voroninski, V. (2013a). **Phase retrieval via matrix completion.** *SIAM Journal on Imaging Sciences*, 6(1).

[Candès and Li, 2014] Candès, E. J. and Li, X. (2014). **Solving quadratic equations via phaselift when there are about as many equations as unknowns.** *Foundations of Computational Mathematics*, 14(5):1017–1026.

[Candès et al., 2015a] Candès, E. J., Li, X., and Soltanolkotabi, M. (2015a). **Phase retrieval from coded diffraction patterns.** *Applied and Computational Harmonic Analysis*, 39(2):277–299.

[Candès et al., 2015b] Candès, E. J., Li, X., and Soltanolkotabi, M. (2015b). **Phase retrieval via wirtinger flow: Theory and algorithms.** *Information Theory, IEEE Transactions on*, 61(4):1985–2007.

[Candès et al., 2013b] Candès, E. J., Strohmer, T., and Voroninski, V. (2013b). **Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming.** *Communications on Pure and Applied Mathematics*, 66(8):1241–1274.

[Chen and Candès, 2015] Chen, Y. and Candès, E. J. (2015). **Solving random quadratic systems of equations is nearly as easy as solving linear systems.** *arXiv preprint arXiv:1505.05114*.

[Conn et al., 2000] Conn, A. R., Gould, N. I. M., and Toint, P. L. (2000). ***Trust-region Methods.*** Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

[Ge et al., 2015] Ge, R., Huang, F., Jin, C., and Yuan, Y. (2015). **Escaping from saddle points—online stochastic gradient for tensor decomposition.** In *Proceedings of The 28th Conference on Learning Theory*, pages 797–842.

[Ge et al., 2016] Ge, R., Lee, J. D., and Ma, T. (2016). **Matrix completion has no spurious local minimum.** *arXiv preprint arXiv:1605.07272*.

[Goldfarb, 1980] Goldfarb, D. (1980). **Curvilinear path steplength algorithms for minimization which use directions of negative curvature.** *Mathematical programming*, 18(1):31–40.

# References II


[Lee et al., 2016] Lee, J. D., Simchowitz, M., Jordan, M. I., and Recht, B. (2016). **Gradient descent converges to minimizers.** *arXiv preprint arXiv:1602.04915.*

[Moré and Sorensen, 1983] Moré, J. J. and Sorensen, D. C. (1983). **Computing a trust region step.** *SIAM Journal on Scientific and Statistical Computing*, 4(3):553–572.

[Nesterov and Polyak, 2006] Nesterov, Y. and Polyak, B. T. (2006). **Cubic regularization of newton method and its global performance.** *Mathematical Programming*, 108(1):177–205.

[Netrapalli et al., 2013] Netrapalli, P., Jain, P., and Sanghavi, S. (2013). **Phase retrieval using alternating minimization.** In *Advances in Neural Information Processing Systems*, pages 2796–2804.

[Rendl and Wolkowicz, 1997] Rendl, F. and Wolkowicz, H. (1997). **A semidefinite framework for trust region subproblems with applications to large scale minimization.** *Mathematical Programming*, 77(1):273–299.

[Shechtman et al., 2015] Shechtman, Y., Eldar, Y. C., Cohen, O., Chapman, H. N., Miao, J., and Segev, M. (2015). **Phase retrieval with application to optical imaging: A contemporary overview.** *Signal Processing Magazine, IEEE*, 32(3):87–109.

[Sun et al., 2015] Sun, J., Qu, Q., and Wright, J. (2015). **Complete dictionary recovery over the sphere.** *arXiv preprint arXiv:1504.06785.*

[Sun et al., 2016] Sun, J., Qu, Q., and Wright, J. (2016). **A geometric analysis of phase retreival.** *arXiv preprint arXiv:1602.06664.*

[Waldspurger et al., 2015] Waldspurger, I., d Aspremont, A., and Mallat, S. (2015). **Phase recovery, maxcut and complex semidefinite programming.** *Mathematical Programming*, 149(1-2):47–81.

[White et al., 2015] White, C. D., Ward, R., and Sanghavi, S. (2015). **The local convexity of solving quadratic equations.** *arXiv preprint arXiv:1506.07868.*