Solving Most Systems of Random Quadratic Equations

Gang Wang[1,2], Georgios B. Giannakis[1], Yousef Saad[3], and Jie Chen[2]
Departments of [1]ECE and [3]CSE, University of Minnesota, Mpls., MN 55455, USA
[2]School of Automation, Beijing Institute of Technology, Beijing 100081, China

Solving Quadratic Equations?

- **Problem statement**
  Recover an unknown n-dimensional vector $\boldsymbol{x}$ from m phaseless quadratic equations

$$|\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle| = \psi_i = \sqrt{y_i}, \quad 1 \le i \le m.$$

- **Sufficient conditions for uniqueness**
  - Real case: $m \ge 2n-1$ $(2n-1$ necessary)
  - Complex case: $m \ge 4n-4$
- **Applications**
  - Phase retrieval in Xray crystallography, imaging, microscopy, ptychography, astronomy…
  - Mixed linear regressions

Prior Art

- **Convex paradigm via matrix-lifting $\boldsymbol{X} := \boldsymbol{x}\boldsymbol{x}^*$**

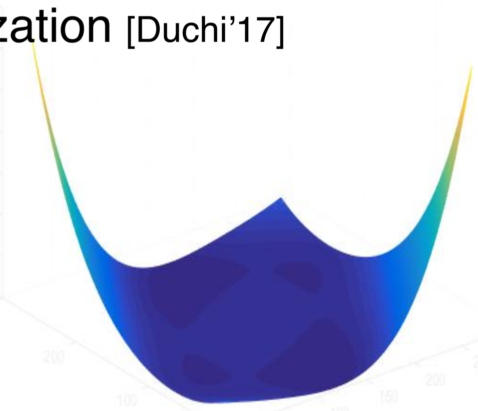
$$y_i = \boldsymbol{a}_i^*\boldsymbol{x}\boldsymbol{x}^*\boldsymbol{a}_i = \mathrm{Tr}(\boldsymbol{a}_i\boldsymbol{a}_i^*\boldsymbol{X}), \quad 1 \le i \le m$$

[Candes'13, Waldspurger'15]

- **Nonconvex paradigm (least-squares criterion)**

$$\underset{\boldsymbol{z}\in\mathbb{R}^n}{\text{minimize}} \quad f(\boldsymbol{z}) = \frac{1}{2m}\sum_{i=1}^m \left(y_i - |\boldsymbol{a}_i^*\boldsymbol{z}|^2\right)^2$$

Algs: AltMinPhase [Netrapalli'15], truncated/projected/reshaped Wirtinger Flow (T/RWF) [Candes'15, Chen'15, Zhang'16, Soltanolkotabi'17], truncated amplitude flow [wang'16], composite optimization [Duchi'17]



Two stages [Candes'15]

- Spectral initialization

$$\boldsymbol{z}_0 \Longleftarrow \text{leading eigenvector of } \frac{1}{m}\sum_{k=1}^m y_i\boldsymbol{a}_i\boldsymbol{a}_i^*$$

- Iterative refinement

$$\boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \mu_t\nabla f(\boldsymbol{z}_t), \quad t = 0, 1, \ldots$$

- **Performance guarantees with $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$**

Table 1: Comparisons of Different Algorithms

Algorithm	Sample complexity m	Computational complexity
PhaseLift	$\mathcal{O}(n)$	$\mathcal{O}(n^3/\epsilon^2)$
PhaseCut	$\mathcal{O}(n)$	$\mathcal{O}(n^3/\sqrt{\epsilon})$
AltMinPhase	$\mathcal{O}(n\log n(\log^2 n + \log(1/\epsilon))$	$\mathcal{O}(n^2\log n(\log^2 n + \log^2(1/\epsilon))$
WF	$\mathcal{O}(n\log n)$	$\mathcal{O}(n^3\log n\log(1/\epsilon))$
TWF	$\mathcal{O}(n)$	$\mathcal{O}(n^2\log(1/\epsilon))$

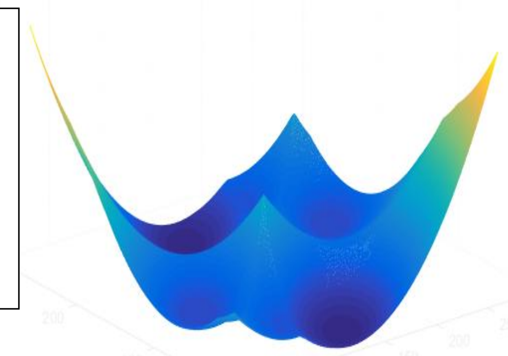
$\epsilon > 0$ given solution accuracy

Problem Formulation and Algorithm

- **The amplitude-based LS cost**

$$\underset{\boldsymbol{z}\in\mathbb{R}^n}{\text{minimize}} \quad \ell(\boldsymbol{z}) = \frac{1}{2m}\sum_{i=1}^m \left(\psi_i - |\boldsymbol{a}_i^*\boldsymbol{z}|\right)^2$$

Nonsmooth, nonconvex; exponentially many stationary points; even NP-hard to find a local minimum



- **Reweighted amplitude flow (RAF)**
  - Reweighted spectral initialization

$$\boldsymbol{z}_0 \Longleftarrow \text{1st eigenvector of } \bar{\boldsymbol{Y}}_0 = \frac{1}{|\mathcal{I}_0|}\sum_{i\in\mathcal{I}_0} w_i\boldsymbol{a}_i\boldsymbol{a}_i^*$$

  - Reweighted gradient iterations

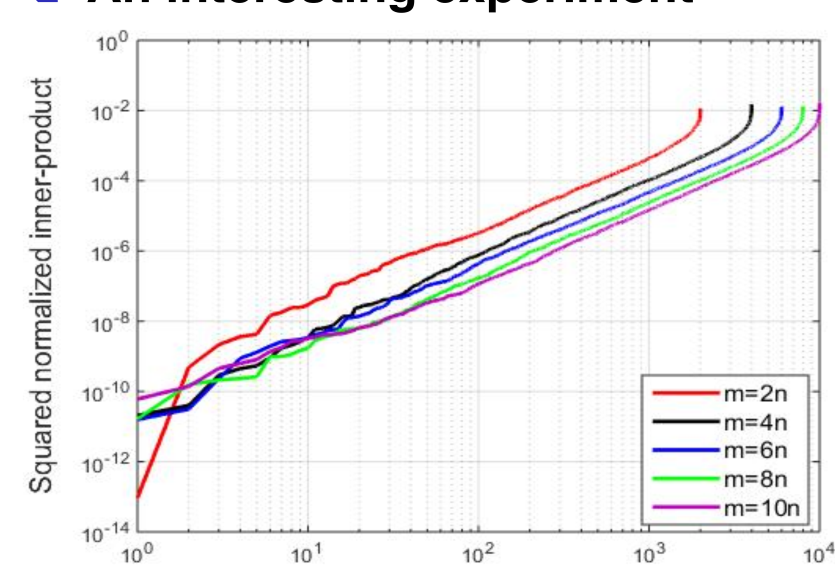
Index set/weights to be designed

$$\boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \mu\,\partial\ell_{\mathrm{rw}}(\boldsymbol{z}_t), \quad t = 0, 1, \ldots$$

Reweighted gradients

Reweighted Initialization

- **An interesting experiment**



$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$$\boldsymbol{a}_i \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$$\frac{|\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2}{\|\boldsymbol{a}_i\|^2\|\boldsymbol{x}\|^2}$$

Ordered squared normalized inner-products

High-dimensional random vectors are almost always nearly orthogonal to each other! [Cai'13]

- **Key idea:** approximate $\boldsymbol{x}$ by a vector maximally correlated to a subset of vectors $\{\boldsymbol{a}_i\}_{i\in\mathcal{I}_0}$
- **Find:** Evaluate and order the data

$$y_{[1]} \ge y_{[2]} \ge \ldots \ge y_{[|\mathcal{I}_0|]} \ge y_{[|\mathcal{I}_0|+1]} \cdots \ge y_{[m]}$$

- **Solve:** the largest eigenvalue problem

$$\boldsymbol{z}_0 := \arg\max_{\|\boldsymbol{z}\|=1} \; \boldsymbol{z}^*\left(\frac{1}{|\mathcal{I}_0|}\sum_{i\in\mathcal{I}_0}\psi_i^\gamma\boldsymbol{a}_i\boldsymbol{a}_i^*\right)\boldsymbol{z}$$

Solvable by power iterations in linear time

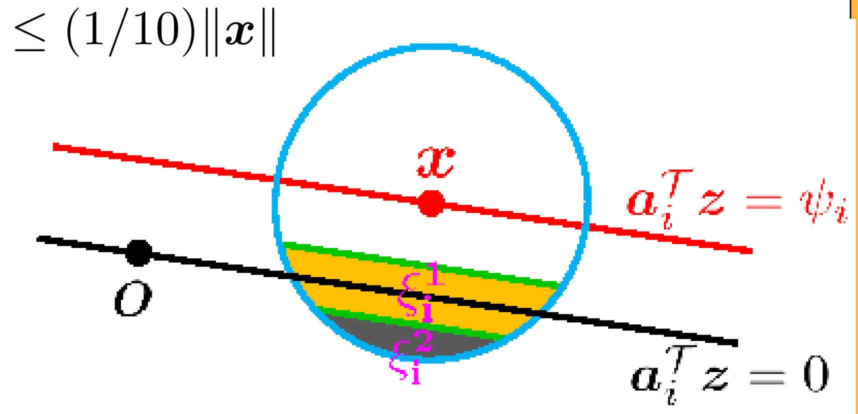
Scaled by norm estimate to have

$$\tilde{\boldsymbol{z}}_0 = \sqrt{\frac{1}{m}\sum_{i=1}^m\psi_i^2}\,\boldsymbol{z}_0$$

Reweighted Gradients

- **Iteratively reweighting?**

$$\mathrm{dist}(\boldsymbol{z}_0, \boldsymbol{x}) \le (1/10)\|\boldsymbol{x}\|$$



The larger $\frac{|\boldsymbol{a}_i^*\boldsymbol{z}|}{|\boldsymbol{a}_i^*\boldsymbol{x}|}$, the more likely $\partial\ell_i(\boldsymbol{z})$ points to $\boldsymbol{x}$

Trick: Weight more the reliable gradients

- **Iteratively reweighted gradient refinements**

$$\boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \frac{\mu}{m}\sum_{i=1}^m w_i^t\left(\boldsymbol{a}_i^*\boldsymbol{z}_t - \psi_i\frac{\boldsymbol{a}_i^*\boldsymbol{z}_t}{|\boldsymbol{a}_i^*\boldsymbol{z}_t|}\right)\boldsymbol{a}_i$$

$$w_i := \frac{1}{1 + c_i/\left(|\boldsymbol{a}_i^*\boldsymbol{z}_t|/|\boldsymbol{a}_i^*\boldsymbol{x}|\right)}$$

Subsumes existing TAF/RWF as special cases
Help eliminating non-smoothness

Theoretical Guarantees

Theorem If

(a1) measurements are noiseless, i.i.d., Gaussian

(a2) sample complexity satisfies $m \ge c_1|\mathcal{I}_0| \ge c_2 n$

(a3) step size is chosen as $\mu \lesssim \mu_0$

then w.h.p. for some $0 < \nu < 1$, RAF starts within basin of attraction and converges linearly to the global minimum

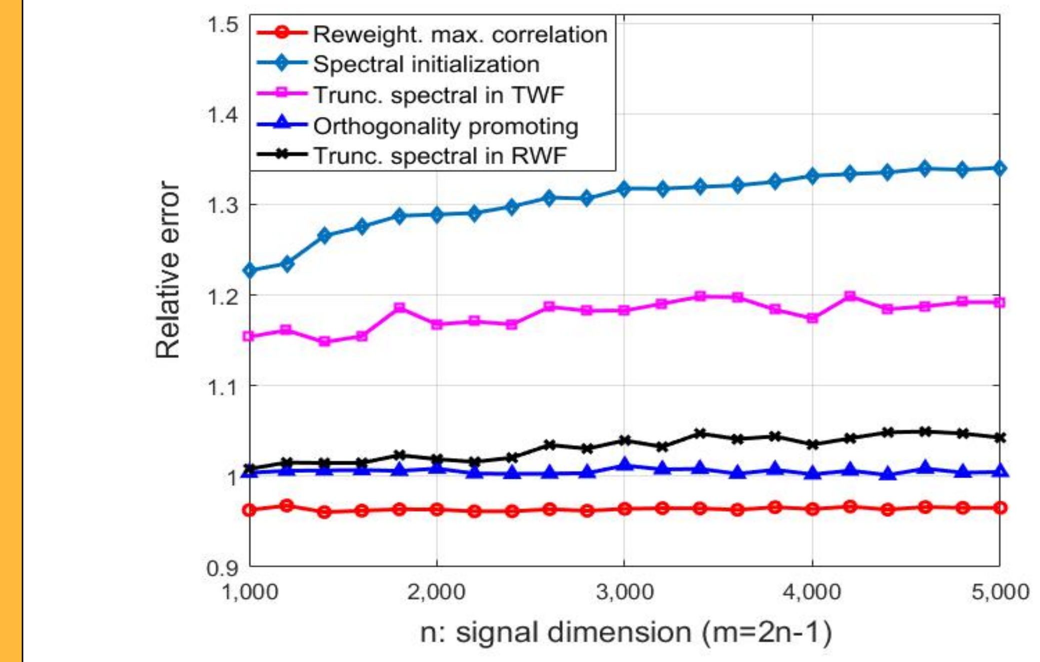
$$\mathrm{dist}(\boldsymbol{z}_t, \boldsymbol{x}) = \min\|\boldsymbol{z}_t \pm \boldsymbol{x}\| \lesssim \frac{1}{10}(1-\nu)^t\|\boldsymbol{x}\|.$$

- Converges to global min of a nonconvex function
- Computational cost $\mathcal{O}(mn\log(1/\epsilon))$
- Faster than state-of-the-art WF, TWF [Candes'15, Chen'15]

Discussions

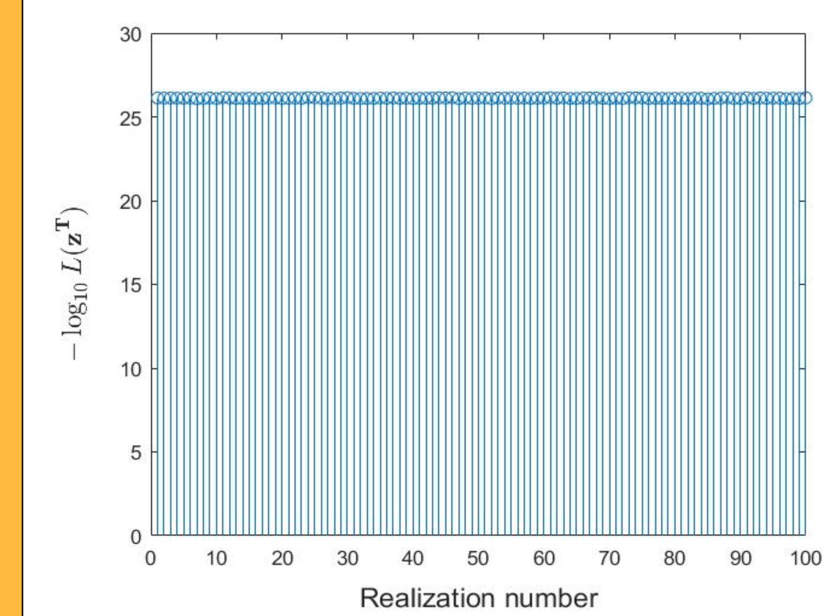
- Initializations for more applications?
- More practical Fourier phase retrieval
- Gradient regularization for other nonconvex opt. such as matrix/tensor completion, robust PCA, blind deconvolution, deep learning…

Empirical Results



Noiseless Gaussian data

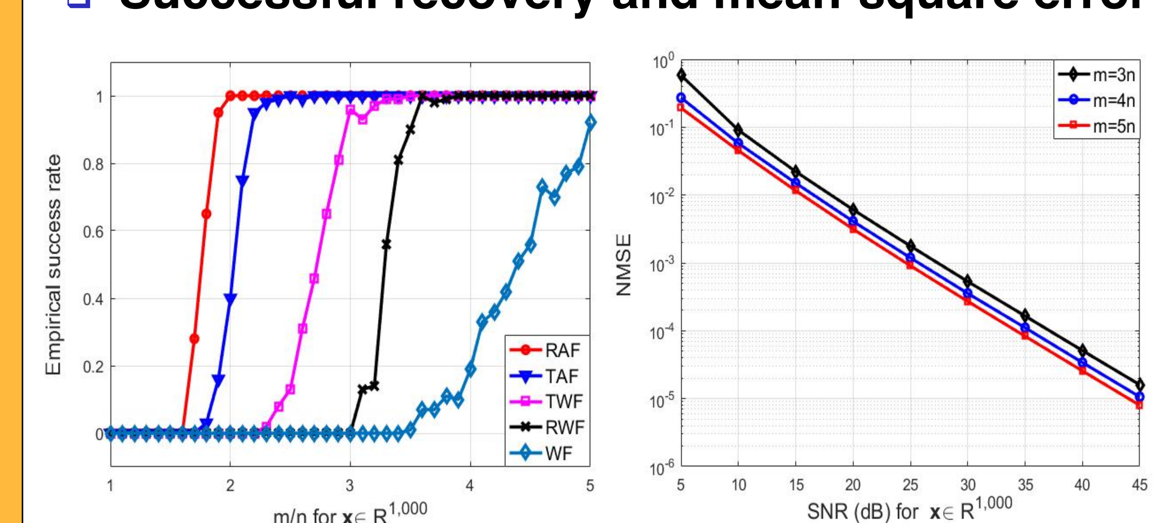
- **Numerical surprise (objective in log scale)**



$$n = 1{,}000$$
$$\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$
$$T = 1{,}000$$

Under the information-limit condition $m = 2n-1$

- **Successful recovery and mean-square error**



- **Real-data validation (CDP w/ 4 masks, $n \approx 10^6$)**

$$\boldsymbol{\psi}^{(k)} = |\boldsymbol{F}\boldsymbol{D}^{(k)}\boldsymbol{x}|, \quad 1 \le k \le K = 3$$

Top: recovered image by TAF; bottom: by RAF