Fundamental Limits of Weak Recovery with Applications to Phase Retrieval

Marco Mondelli MONDELLI@STANFORD.EDU
Department of Electrical Engineering, Stanford University

Andrea Montanari MONTANARI@STANFORD.EDU
Department of Electrical Engineering and Department of Statistics, Stanford University

Editors: Sebastien Bubeck, Vianney Perchet and Philippe Rigollet

Abstract

In phase retrieval we want to recover an unknown signal $\boldsymbol{x} \in \mathbb{C}^d$ from n quadratic measurements of the form $y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|^2 + w_i$ where $\boldsymbol{a}_i \in \mathbb{C}^d$ are known sensing vectors and w_i is measurement noise. We ask the following *weak recovery* question: what is the minimum number of measurements n needed to produce an estimator $\hat{\boldsymbol{x}}(\boldsymbol{y})$ that is positively correlated with the signal $\boldsymbol{x}$? We consider the case of Gaussian vectors $\boldsymbol{a}_i$. We prove that – in the high-dimensional limit – a sharp phase transition takes place, and we locate the threshold in the regime of vanishingly small noise. For $n \leq d - o(d)$ no estimator can do significantly better than random and achieve a strictly positive correlation. For $n \geq d + o(d)$ a simple spectral estimator achieves a positive correlation. Surprisingly, numerical simulations with the same spectral estimator demonstrate promising performance with realistic sensing matrices. Spectral methods are used to initialize non-convex optimization algorithms in phase retrieval, and our approach can boost the performance in this setting as well.

Our impossibility result is based on classical information-theory arguments. The spectral algorithm computes the leading eigenvector of a weighted empirical covariance matrix. We obtain a sharp characterization of the spectral properties of this random matrix using tools from free probability and generalizing a recent result by Lu and Li. Both the upper and lower bound generalize beyond phase retrieval to measurements y_i produced according to a generalized linear model. As a byproduct of our analysis, we compare the threshold of the proposed spectral method with that of a message passing algorithm.

Keywords: Spectral initialization, phase transition, mutual information, second moment method, phase retrieval, free probability

In this work[1], we consider the problem of recovering a signal $\boldsymbol{x}$ of dimension d, given n *generalized linear measurements*. More specifically, the measurements are drawn independently according to the conditional distribution

$$y_i \sim p(y \mid |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle|), \qquad i \in \{1, \dots, n\}, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, $\{\boldsymbol{a}_i\}_{1 \leq i \leq n}$ is a set of known sensing vector, and $p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle)$ is a known probability density function. For the problem of *phase retrieval*, the model (1) is specialized to

$$y_i = |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle|^2 + w_i, \qquad i \in \{1, \dots, n\}\,, \tag{2}$$

where w_i is noise.

1. Extended abstract. Full version appears as [arXiv:1708.05932, v3].

Popular methods to solve the phase retrieval problem are based on semi-definite programming relaxations (Candès et al., 2015a,b, 2013; Waldspurger et al., 2015). However, these algorithms rapidly become prohibitive from a computational point of view when the dimension d of the signal increases, which makes them impractical in most of the real-world applications. For this reason, several algorithms have been developed in order to solve directly the non-convex least-squares problem, including the error reduction schemes dating back to Gerchberg-Saxton and Fienup (Gerchberg, 1972; Fienup, 1982), alternating minimization (Netrapalli et al., 2013), approximate message passing (Schniter and Rangan, 2015), Wirtinger Flow (Candès et al., 2015c), iterative projections (Li et al., 2015), the Kaczmarz method (Wei, 2015), and a number of other approaches (Chen and Candès, 2017; Zhang and Liang, 2016; Cai et al., 2016; Wang et al., 2016; Wang and Giannakis, 2016; Soltanolkotabi, 2017; Duchi and Ruan, 2017; Wang et al., 2017). Furthermore, recently a convex relaxation that operates in the natural domain of the signal was independently proposed by two groups of authors (Goldstein and Studer, 2016; Bahmani and Romberg, 2017). All these techniques require an initialization step, whose goal is to provide a solution $\hat{\boldsymbol{x}}$ that is positively correlated with the unknown signal $\boldsymbol{x}$. To do so, spectral methods are widely employed: the estimate $\hat{\boldsymbol{x}}$ is given by the principal eigenvector of a suitable matrix constructed from the data. A similar stategy (initialization step followed by an iterative algorithm) has proved successful for many other estimation problems, e.g., matrix completion (Keshavan et al., 2010; Jain et al., 2013), blind deconvolution (Lee et al., 2017; Li et al., 2016), sparse coding (Arora et al., 2015) and joint alignment from pairwise noisy observations (Chen and Candès, 2016).

We focus on a regime in which both the number of measurement n and the dimension of the signal d tend to infinity, but their ratio n/d tends to a positive constant δ. The *weak recovery* problem requires to provide an estimate $\hat{\boldsymbol{x}}(\boldsymbol{y})$ that has a positive correlation with the unknown vector $\boldsymbol{x}$:

$$\liminf_{n\to\infty} \mathbb{E}\left\{\frac{|\langle \hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{x}\rangle|}{\|\hat{\boldsymbol{x}}(\boldsymbol{y})\|_2 \, \|\boldsymbol{x}\|_2}\right\} > \epsilon, \tag{3}$$

for some $\epsilon > 0$.

In this paper, we consider either $\boldsymbol{x} \in \mathbb{R}^d$ or $\boldsymbol{x} \in \mathbb{C}^d$ and assume that the measurement vectors $\boldsymbol{a}_i$ are standard Gaussian (either real or complex). In the general setting of model (1), we present two types of results:

1. We develop an *information-theoretic lower bound* δ_ℓ: for $\delta < \delta_\ell$, no estimator can output non-trivial estimates. In other words, the weak recovery problem cannot be solved.

2. We establish an *upper bound* $\delta_{\rm u}$ based on a *spectral algorithm*: for $\delta > \delta_{\rm u}$, we can achieve weak recovery (see (3)) by letting $\hat{\boldsymbol{x}}$ be the principal eigenvector of a matrix suitably constructed from the data.

The values of the thresholds δ_ℓ and $\delta_{\rm u}$ depend on the conditional distribution $p(\cdot \mid |\langle \boldsymbol{x}, \boldsymbol{a}_i\rangle|)$, and we provide analytic formulas to compute them. More formally, consider the function $f : [0,1] \to \mathbb{R}$, given by

$$f(m) = \int_{\mathbb{R}} \frac{\mathbb{E}_{G_1,G_2}\left\{p(y \mid |G_1|)p(y \mid |G_2|)\right\}}{\mathbb{E}_G\left\{p(y \mid |G|)\right\}} \, {\rm d}y, \tag{4}$$

with

$$G \sim \mathsf{CN}(0,1), \qquad (G_1, G_2) \sim \mathsf{CN}\left(\boldsymbol{0}_2, \begin{bmatrix} 1 & c \\ c^* & 1 \end{bmatrix}\right), \tag{5}$$

and $m = |c|^2$. Note that the RHS of (4) depends only on $m = |c|^2$. Indeed, by applying the transformation $(G_1, G_2) \to (e^{i\theta_1}G_1, e^{i\theta_2}G_2)$, $f(m)$ does not change, but the correlation coefficient c is mapped into $ce^{i(\theta_1-\theta_2)}$. Furthermore, set

$$F_\delta(m) = \delta \log f(m) + \log(1-m). \tag{6}$$

Note that, when $m = 0$, G_1 and G_2 are independent. Hence, $f(0) = 1$, which implies that $F_\delta(0) = 0$ for any $\delta > 0$. We define the information-theoretic threshold δ_ℓ as the largest value of δ such that the maximum of $F_\delta(m)$ is attained at $m = 0$, i.e.,

$$\delta_\ell = \sup\{\delta \mid F_\delta(m) < 0 \text{ for } m \in (0, 1]\}. \tag{7}$$

The spectral threshold δ_{u} is defined as

$$\delta_{\mathrm{u}} = \frac{1}{\displaystyle\int_{\mathbb{R}} \frac{\left(\mathbb{E}_G\left\{p(y \mid |G|)(|G|^2-1)\right\}\right)^2}{\mathbb{E}_G\left\{p(y \mid |G|)\right\}} \mathrm{d}y}, \tag{8}$$

with $G \sim \mathsf{CN}(0, 1)$.

The main result of this paper can be summarized as follows.

Theorem *Let $\boldsymbol{x} \in \mathbb{C}^d$ be chosen uniformly at random on the d-dimensional complex sphere with radius $\sqrt{d}$ and assume that $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{CN}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$. Let $\boldsymbol{y} \in \mathbb{R}^n$ be drawn independently according to (1), and $n, d \to \infty$ with $n/d \to \delta \in (0, +\infty)$. Then,*

- *For $\delta < \delta_\ell$, no algorithm can provide non-trivial estimates on $\boldsymbol{x}$;*
- *For $\delta > \delta_{\mathrm{u}}$, there exists a spectral algorithm that returns an estimate $\hat{\boldsymbol{x}}$ satisfying (3).*

For the special case of phase retrieval (see (2)), we evaluate the thresholds δ_ℓ and δ_{u}, and we show that they coincide in the limit of vanishing noise.

Theorem *Let $\boldsymbol{x} \in \mathbb{C}^d$ be chosen uniformly at random on the d-dimensional complex sphere with radius $\sqrt{d}$, and assume that $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{CN}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$. Let $\boldsymbol{y} \in \mathbb{R}^n$ be drawn independently according to (2), with $\{w_i\}_{1\le i\le n} \sim \mathsf{N}(0, \sigma^2)$, and $n, d \to \infty$ with $n/d \to \delta \in (0, +\infty)$. Then,*

- *For $\delta < 1$, no algorithm can provide non-trivial estimates on $\boldsymbol{x}$;*
- *For $\delta > 1$, there exists $\sigma_0(\delta) > 0$ and a spectral algorithm that returns an estimate $\hat{\boldsymbol{x}}$ satisfying (3), for any $\sigma \in [0, \sigma_0(\delta)]$.*

When $\boldsymbol{x}$ is chosen uniformly at random on the d-dimensional *real* sphere with radius $\sqrt{d}$ and $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, we show that analogous results hold and that the threshold for phase retrieval moves from 1 to $1/2$. This is reminiscent of how the injectivity thresholds are $\delta = 4$ and $\delta = 2$ in the complex and the real case, respectively (Balan et al., 2006; Bandeira et al., 2014; Conca et al., 2015). A possible intuition for this halving phenomenon comes from the fact that the complex problem has twice as many variables but the same amount of equations of the real problem. Hence, it is reasonable that the complex case requires twice the amount of data with respect to the real case.

The lower bound is proved by estimating the conditional entropy via the second moment method.

The spectral algorithm computes the eigenvector corresponding to the largest eigenvalue of a matrix of the form:

$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^* \,, \tag{9}$$

where $\mathcal{T} : \mathbb{R} \to \mathbb{R}$ is a pre-processing function. For δ large enough (and a suitable choice of $\mathcal{T}$), we expect the resulting eigenvector $\hat{\boldsymbol{x}}(\boldsymbol{y})$ to be positively correlated with the true signal $\boldsymbol{x}$. The recent paper (Lu and Li, 2017) computed exactly the threshold value δ_{u}, under the assumption that the measurement vectors are real Gaussian, and $\mathcal{T}$ is non-negative.

Here, we generalize the result of (Lu and Li, 2017) by removing the assumption that $\mathcal{T}(y) \geq 0$ and by considering the complex case. Armed with this result, we compute the optimal[2] pre-processing function $\mathcal{T}_\delta^*(y)$ for the general model (1). Our upper bound δ_u is the phase transition location for this optimal spectral method. In the case of phase retrieval (as $\sigma \to 0$), this pre-processing function is given by

$$\mathcal{T}_\delta^*(y) = \frac{y-1}{y+\sqrt{\delta}-1}, \tag{10}$$

and achieves weak recovery for any $\delta > \delta_u = 1$. In the limit $\delta \downarrow 1$, this converges to the limiting function $\mathcal{T}^*(y) = 1 - (1/y)$.

While the expression (10) is remarkably simple, it is somewhat counter-intuitive. Earlier methods (Candès et al., 2015c; Chen and Candès, 2015; Lu and Li, 2017) use $\mathcal{T}(y) \geq 0$ and try to extract information from the large values of y_i. The function (10) has a large negative part for small y, in particular when δ is close to 1. Furthermore, it extracts useful information from data points with y_i small. One possible interpretation is that the points in which the measurement vector is basically orthogonal to the unknown signal are not informative, hence we penalize them.

Our analysis applies to Gaussian measurement matrices. However, the proposed spectral method works well also on real images and realistic measurement matrices.

We also compare our spectral approach to message passing algorithms. In particular, we prove that, for $\delta < \delta_{\mathrm{u}}$ (i.e. in the regime in which the spectral approach fails), message passing converges to an un-informative fixed point, even if initialized in a state that is correlated with the true signal $\boldsymbol{x}$. Vice versa, for $\delta > \delta_{\mathrm{u}}$ (i.e. in the regime in which the spectral approach achieves weak recovery), we consider a linearized message passing algorithm, and prove that the un-informative fixed point is unstable.

References

Sanjeev Arora, Rong Ge, Tengyu Ma, and Ankur Moitra. Simple, efficient, and neural algorithms for sparse coding. In *Conference on Learning Theory (COLT)*, pages 113–149, Paris, France, July 2015.

Sohail Bahmani and Justin Romberg. Phase retrieval meets statistical learning theory: A flexible convex relaxation. In *Proc. of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 252–260, Fort Lauderdale, FL, June 2017.

2. Here optimality is understood with respect to the weak recovery threshold.

Radu Balan, Pete Casazza, and Dan Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345–356, 2006.

Afonso S. Bandeira, Jameson Cahill, Dustin G. Mixon, and Aaron A. Nelson. Saving phase: Injectivity and stability for phase retrieval. *Applied and Computational Harmonic Analysis*, 37(1): 106–125, 2014.

T. Tony Cai, Xiaodong Li, and Zongming Ma. Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow. *The Annals of Statistics*, 44(5):2221–2251, 2016.

Emmanuel J. Candès, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

Emmanuel J. Candès, Yonina C. Eldar, Thomas Strohmer, and Vladislav Voroninski. Phase retrieval via matrix completion. *SIAM Review*, 57(2):225–251, 2015a.

Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, 2015b.

Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inform. Theory*, 61(4):1985–2007, 2015c.

Yuxin Chen and Emmanuel J. Candès. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Advances in Neural Information Processing Systems*, pages 739–747, 2015.

Yuxin Chen and Emmanuel J. Candès. The projected power method: An efficient algorithm for joint alignment from pairwise differences. arXiv:1609.05820, 2016.

Yuxin Chen and Emmanuel J. Candès. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Communications on Pure and Applied Mathematics*, 70:0822–0883, 2017.

Aldo Conca, Dan Edidin, Milena Hering, and Cynthia Vinzant. An algebraic characterization of injectivity in phase retrieval. *Applied and Computational Harmonic Analysis*, 38(2):346–356, 2015.

John C. Duchi and Feng Ruan. Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval. arXiv:1705.02356, 2017.

J. R. Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769, Apr. 1982.

Ralph W. Gerchberg. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

Tom Goldstein and Christoph Studer. Phasemax: Convex phase retrieval via basis pursuit. arXiv:1610.07531, 2016.

Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proc. of the 45th Ann. ACM Symp. on Theory of Computing (STOC)*, pages 665–674, Palo Alto, CA, June 2013. ACM.

Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Trans. Inform. Theory*, 56(6):2980–2998, 2010.

Kiryung Lee, Yanjun Li, Marius Junge, and Yoram Bresler. Blind recovery of sparse signals from subsampled convolution. *IEEE Trans. Inform. Theory*, 63(2):802–821, 2017.

Gen Li, Yuantao Gu, and Yue M. Lu. Phase retrieval using iterative projections: Dynamics in the large systems limit. In *Proc. of the 53rd Annual Allerton Conf. on Commun., Control, and Computing (Allerton)*, pages 1114–1118, Monticello, IL, Oct. 2015.

Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. arXiv:1606.04933, 2016.

Yue M. Lu and Gen Li. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. arXiv:1702.06435, 2017.

Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pages 2796–2804, 2013.

Philip Schniter and Sundeep Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2015.

Mahdi Soltanolkotabi. Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization. arXiv:1702.06175, 2017.

Irène Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1-2):47–81, 2015.

Gang Wang and Georgios B. Giannakis. Solving random systems of quadratic equations via truncated generalized gradient flow. In *Advances in Neural Information Processing Systems*, pages 568–576, 2016.

Gang Wang, Georgios B. Giannakis, and Yonina C. Eldar. Solving systems of random quadratic equations via truncated amplitude flow. arXiv:1605.08285, 2016.

Gang Wang, Georgios B. Giannakis, Yousef Saad, and Jie Chen. Solving almost all systems of random quadratic equations. arXiv:1705.10407, 2017.

Ke Wei. Solving systems of phaseless equations via Kaczmarz methods: A proof of concept study. *Inverse Problems*, 31(12), 2015.

Huishuai Zhang and Yingbin Liang. Reshaped Wirtinger Flow for solving quadratic system of equations. In *Advances in Neural Information Processing Systems*, pages 2622–2630, 2016.