On Efficiency and Low Sample Complexity in Phase Retrieval.

Youssef Mroueh
CBCL, CSAIL
LCSL, Italian Institute of Technology
Massachusetts Institute of Technology
Email:ymroueh@mit.edu

Lorenzo Rosasco
DIBRIS, Universita di Genova
LCSL, Italian Institute of Technology
Massachusetts Institute of Technology
Email: lrosasco@mit.edu

***Abstract*—In this paper we show that the problem of phase retrieval can be efficiently and provably solved via an alternating minimization algorithm suitably initialized. Our initialization is based on One Bit Phase Retrieval that we introduced in [1], where we showed that $O(n\log(n))$ Gaussian phase-less measurements ensure robust recovery of the phase. In this paper we improve the sample complexity bound to $O(n)$ measurements for sufficiently large n, using a variant of Matrix Bernstein concentration inequality that exploits the intrinsic dimension, together with properties of one bit phase retrieval.**

I. Introduction

The phase recovery problem can be modeled as the problem of reconstructing a n-dimensional complex vector x_0 given only the magnitude of m phase-less linear measurements. Such a problem arises for example in X-ray crystallography [2], [3], diffraction imaging [4] or microscopy [5], where one can only measure the intensities of the incoming waves, and wishes to recover the lost phase in order to be able to reconstruct the desired signal.

In practice, phase recovery is often tackled via greedy algorithms [6], [7], [8] which typically lack convergence guarantees. Recently, approaches based on convex relaxations, namely PhaseLift in [9], [10], and Phase cut in [11], have been proposed and analyzed. These latter methods can be solved by Semi Definite Programing (SDP), and allow the exact and stable recovery of the signal (up to a global phase) from $O(n)$ measurements. A different approach has been recently considered in [12], where it is shown that a greedy alternating minimization, akin to those in [6], [7], [8], can be shown to geometrically converge to the true vector x_0 if $O(n\log^3 n)$ measurements are given. Indeed, alternating minimization algorithms are known to be extremely sensitive to the initialization and a suitable initialization is the key of the analysis in [12]. Throughout this paper we call the initialization step of [12] *SubExpPhase*. While alternating minimization approaches provide a solution only up-to a given accuracy, they often have very good practical performances when compared to convex methods [12], with important computational advantages [12]. The solution of the SDP in convex approaches is computationally expensive and needs to be close to a rank one matrix for tight recovery (which is rarely encountered in practice [11]). Indeed, some greedy refinement of the SDP solution is often considered [11].

More recently an approach based on quantization for phase retrieval was proposed in [1], where a quantization scheme for non negative phase-less measurements was proposed, so called one bit quantization. This approach is called One Bit Phase Retrieval *1bitPhase*. It is shown in [1] that we need $O(\frac{n\log(n)}{\epsilon^2})$ measurements in order to achieve an accuracy ϵ. Using the solution of One Bit Phase retrieval as an initialization to the alternating minimization algorithm, it is shown in [1] that the overall procedure allows to achieve an accuracy ϵ using $O(n(\log n + \log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}))$ measurements.

In this paper we improve on those results and show that for sufficiently large n, $O(\frac{n}{\epsilon^2})$ measurements are sufficient to achieve an accuracy ϵ in one bit phase retrieval. It follows that alternating minimization initialized with the solution of One Bit Phase retrieval achieves an accuracy ϵ from only $O(n(1+\log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}))$ measurements. This result bridges the gap between convex and non convex approaches for phase retrieval in term of sample complexity (see Table I), with a computational advantage for non convex approaches over SDP convex relaxations as shown in Table II.

Notation: For $z\in\mathbb{C}$, $|z|^2$ is squared complex modulus of z. For $a,a'\in\mathbb{C}^n$, $\langle a,a'\rangle$ is the complex dot product in $\mathbb{C}^n$. For $a\in\mathbb{C}^n, a^*$ is the complex conjugate and $||a||_2$ is the norm 2 of a. Let A a complex hermitian matrix in $\mathbb{C}^n$, $||A||_F$ denotes the Frobenius norm of A, $||A||$ denotes the operator norm of A, $tr(A)$ denotes the trace of A. Throughout the paper, we denote by c, C positive absolute constants whose values may change from instance to instance.

II. Background and Previous Work

In this section, we formalize the problem of recovering a signal from phase-less measurements and discuss previous results. Throughout this section, and the rest of the paper, we consider measurements defined by independent and identically distributed Complex Gaussian $\mathcal{CN}(0, I_n)$ sensing vectors,

$$a_i \in \mathbb{C}^n, \qquad a_i \sim \mathcal{N}(0, \frac{1}{2}I_n) + i\mathcal{N}(0, \frac{1}{2}I_n), i = 1 \dots m. \tag{1}$$

The (noiseless) phase recovery problem is defined as follows.

Definition 1 (Phase-less Sensing and Phase Recovery)**.** *Suppose phase-less sensing measurements*

$$b_i = |\langle a_i, x_0 \rangle|^2 \in \mathbb{R}_+, \qquad i = 1 \dots m, \tag{2}$$

are given for $x_0 \in \mathbb{C}^n$*, where* a_i*,* $i = 1, \dots, m$ *are random vectors as in* (1)*. The phase recovery problem is*

$$\text{find } x, \qquad \text{subject to} \quad |\langle a_i, x \rangle|^2 = b_i, \quad i = 1 \dots m. \tag{3}$$

The above problem is non convex and in the following we recall the sample and computational complexity of recent approaches to provably and efficiently recover x_0 from a finite number of measurements. In the next section we focus on non convex approaches to phase retrieval based on suitably initialized alternated minimization (AM).

	Sample complexity
PhaseLift [9], [10]	$O(n)$
PhaseCut [11]	$O(n)$
SubExpPhase+AM [12]	$O(n\left(\log^3 n + \log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}\right))$
1bitPhase+AM [1]	$O(2n\left(\log(n) + \log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}\right)))$
1bitPhase+AM (this paper)	$O(2n\left(1 + \log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}\right)))$

TABLE I
Comparison of the sample of different phase retrieval schemes.

	Comp. complexity
PhaseLift [9], [10]	$O(n^3/\epsilon^2)$
PhaseCut [11]	$O(n^3/\sqrt{\epsilon})$
SubExpPhase+AM [12]	$O(n^2\left(\log^3 n + \log^2\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}\right))$
1bitPhase+AM [1]	$O(n^2\left(\log n + \log^2\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}\right))$
1bitPhase+AM (this paper)	$O(n^2\left(1 + \log^2\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}\right))$

TABLE II
Comparison of the computational complexity of different phase retrieval schemes.

A. *Phase Retrieval via suitably initialized Alternating Minimization and Resampling*

Let A be the matrix defined by m sensing vectors as in (1) and $B = Diag(\sqrt{b})$, where b is the vector of measurements as in (2). Then, $Ax_0 = Bu_0$, for $u_0 = Ph(Ax_0)$ with $Ph(z) = \left(\frac{z_1}{|z_1|}, \dots \frac{z_m}{|z_m|}\right)$, $z \in \mathbb{C}^m$. The above equality suggests the following natural approach to recover (x_0, u_0),

$$\min_{x,u} ||Ax - Bu||_2^2, \quad \text{subject to} \quad |u_i| = 1, \quad i = 1 \dots m,$$

The above problem is non-convex because of the constraint on u. The AM approach consists in optimizing u, for a given x, and then optimizing x for a given u. It is easy to see that for a given x, the optimal u is simply $u = Ph\left(Ax\right)$, and for a given u, the optimal x is the solution of a least squares problem. The key result in [12] shows that if such an iteration is initialized with maximum eigenvector of the matrix

$$\hat{C}_m = \frac{1}{m}\sum_{i=1}^m b_i a_i a_i^*,$$

then the solution x_{t_0} of the alternating minimization (Algorithm 1) globally converges (with high probability) to the true vector x_0. For a given accuracy ϵ, $0 < \epsilon < 1$, if $m \geq cn(\log^3 n + \log\frac{1}{\epsilon}\log\log\frac{1}{\epsilon})$, then $||x_{t_0} - e^{i\phi}x_0||_2 \leq \epsilon$, where ϕ is a global phase.
A key observation, motivating the above initialization (called *SubExpPhase*), is the fact that the expectation of $\hat{C}_m$ can be shown to satisfy

$$\mathbb{E}(\hat{C}_m) = x_0 x_0^* + I. \tag{4}$$

Indeed, the proof in [12] relies on the concentration properties of the random matrix $\hat{C}_m$ around its expectation [13], [14]. It is useful to note that these latter results crucially depend on a bound on the norm of $b_i a_i a_i^*$ for $i = 1, \dots, m$. Indeed, it is this latter bound the main cause of the poly-logarithmic term in the sample complexity of *SubExpPhase* since the b_i's are exponential random variables.

Algorithm 1 proposed in [12], proceeds in alternating the estimation of the phase and the signal. For technical reasons - mainly ensuring independence - the algorithm proceeds in a stage-wise alternating minimization. At each stage we use a new re-sampled sensing matrix and the corresponding measurements.

Algorithm 1 AltMinPhase with Resampling

procedure ALTMINPHASERESAMPLING(A, b, ϵ)
  $t_0 \leftarrow c\log(\frac{1}{\epsilon})$
  Partition b and the corresponding rows of A into $t_0 + 1$ disjoint sets: $(b_0, A_0), \dots (b_{t_0}, A_{t_0})$.
  Initialize x
  for $t = 0 \dots t_0 - 1$ **do**
    $u_{t+1} \leftarrow Ph(A_{t+1}x_t)$
    $x_t \leftarrow \arg\min ||A_{t+1}x - B_{t+1}u_{t+1}||_2^2$
  end for
  return x_{t_0}
end procedure

B. Robust One Bit Phase Retrieval and Greedy Refinements

More recently a new approach for phase retrieval was proposed in [1] based on a quantization scheme of severely perturbed phaseless linear measurements. Assume we observe pairs of independent phase-less measurements:

$$(b_i^1, b_i^2) = \left(\theta(|\left\langle a_i^1, x_0 \right\rangle|^2), \theta(|\left\langle a_i^2, x_0 \right\rangle|^2)\right), i = 1, \dots, m, \quad (5)$$

where (a_i^1, a_i^2) are independent sensing vectors as in (1) and θ is a *possibly unknown* rank preserving transformation. In particular θ can model a distortion, e.g. $\theta(s) = \tanh(\alpha s)$, $\alpha \in \mathbb{R}_+$, or an additive noise $\theta(s) = s + \nu$, where ν is a stochastic noise, such as exponential noise. The recovery problem from severly perturbed intensity values seems hopeless, and indeed the key in this approach is a quantization scheme based on comparing pairs of phase-less measurements. More precisely for each pair b_i^1, b_i^2 of measurements of the form (5) we define

$$y_i \in \{-1, 1\} \quad y_i = sign(b_i^1 - b_i^2), \quad i = 1 \dots m.$$

The one bit phase retrieval problem reduces to a maximum eigenvalue problem induced by the matrix

$$\hat{C}_m = \frac{1}{m} \sum_{i=1}^{m} y_i (a_i^1 a_i^{1,*} - a_i^2 a_i^{2,*}). \quad (6)$$

In [1] it is shown that the expectation of $\hat{C}_m$ satisfies

$$\mathbb{E}(\hat{C}_m) = \lambda x_0 x_0^*, \quad (7)$$

where λ is a suitable constant which depends on θ and plays the role of a signal-to-noise ratio. Morever for a given accuracy $0 < \epsilon < 1$, if $O(\frac{n \log n}{\epsilon^2 \lambda})$ pairs of measurements are available, then the solution of the above maximum eigenvalue problem satisfies

$$||\hat{x}_m - x_0 e^{i\phi}||_2^2 \leq \epsilon,$$

where $\phi \in [0, 2\pi]$ is a global phase. Interestingly, the authors in [1] show that provided with the one-bit retrieval initialization, the alternating minimization algorithm globally converges (with high probability) to the true vector x_0, and if

$$m \geq cn(\log n + \log \frac{1}{\epsilon} \log \log \frac{1}{\epsilon}), \quad (8)$$

then $||x_{t_0} - e^{i\phi} x_0||_2 \leq \epsilon$. Hence quantization plays the role of a preconditioning that enhances the sample complexity of the overall alternating minimization. While the improvement in the sample complexity of *1bitPhase* in [1] compared to *SubExpPhase* [12] is mainly due to the boundedness of the one bit measurements, we exploit in this paper the fact that $\mathbb{E}(\hat{C}_m)$ is of rank one in the case of *1bitPhase* as shown in equation (7), as opposed to $\mathbb{E}(\hat{C}_m)$ in the case of *SubExpPhase* which is full rank as shown in equation (4). Having a rank one matrix in expectation allows us to use matrix concentration inequalities taking into account the intrinsic dimension of the matrix [14], the latter allows us to get improved sample complexity bounds.

III. Main Results

The main result of this paper is stated in Theorem 1, the proof of this Theorem is given in Section IV.

Theorem 1 (One Bit Phase Retrieval). *Let $\hat{x}_m$ be the maximum eigenvector of $\hat{C}_m$, solution of one bit phase retrieval. For sufficiently large n, for $0 < \epsilon < 1$, for $m \geq \frac{cn}{\lambda \epsilon^2}$, $||\hat{x}_m \hat{x}_m^* - x_0 x_0^*||_F^2 \leq \epsilon$ with probability at least $1 - O(me^{-2n})$. where ϕ is a global phase.*

Theorem 2 (Greedy Refinements). *Let $\hat{x}_m$ be the solution of One bit Phase Retrieval, and consider Algorithm 1 initialized with $\hat{x}_m$ for all $\epsilon, 0 < \epsilon < 1$. Define x_{t_0} the output of Algorithm 1. For $m = O\left(2n(1 + \log \frac{1}{\epsilon} \log \log \frac{1}{\epsilon})\right)$, we have $||x_{t_0} - x_0 e^{i\phi}||_2 \leq \epsilon$ with high probability.*

Leveraging results from [12] that does not depend on the initialization step, Theorem 2 shows that the greedy refinements of one bit solution, ensures convergence to the optimum with high probability and lower sample complexity than the ones obtained in [12], and [1].

IV. Improved Sample Complexity Result for One Bit Phase Retrieval

For completeness we give and simplify some proofs from [1], and then focus on the main contributions of this paper in the concentration techniques and results.

Proposition 1 (Correctness in Expectation). *Let $a^1, a^2 \sim \mathcal{CN}(0, I_n)$, 2 iid complex Gaussian vectors. For $x_0 \in \mathbb{C}^n, ||x_0|| = 1$, let $y = sign(\theta(\left|\left\langle a^1, x_0 \right\rangle\right|^2) - \theta(\left|\left\langle a^2, x_0 \right\rangle\right|^2))$. Let $C = \mathbb{E}_{y, a^1, a^2}\left(y(a^1 a^{1,*} - a^2 a^{2,*})\right)$, we have*

$$C = \lambda x_0 x_0^*,$$

where $\lambda = \mathbb{E}(sign(\theta(E_1) - \theta(E_2))(E_1 - E_2))$, and E_1, E_2 iid $\sim Exp(1)$.

Proof: Let $e_j, j = 1 \dots n$, be the canonical basis. Note $a_\ell^j = \left\langle a^j, e_\ell \right\rangle, j = 1, 2, \ell = 1 \dots n$. By rotation invariance of Gaussians we can consider $x_0 = e_1 = (1, 0 \dots 0)$. Hence $y = sign(\theta(|a_1^1||^2) - \theta(|a_1^2|^2))$. Let $E_1 = |a_1^1|^2$, $E_2 = |a_1^2|^2$, E_1 and E_2 are iid exponential random variables $Exp(1)$. It follows that:

$$\begin{aligned} C &= \mathbb{E}(y(a^1 a^{1,*} - a^2 a^{2,*})) \\ &= \mathbb{E}\sum_{k,\ell} y(a^1_k \bar{a^1_\ell} - a^2_k \bar{a^2_\ell}) e_k e_\ell^*) \\ &= \mathbb{E}(sign(\theta(|a^1_1|^2) - \theta(|a^2_1|^2))(|a^1_1|^2 - |a^2_1|^2)) e_1 e_1^* \\ &= \lambda e_1 e_1^*, \end{aligned}$$

where the last equalities follow from independence and that the Gaussian are centered.

■

Proposition 1 suggests that x_0 can be recovered as the maximum eigenvector of the matrix C. Moreover C is a rank one matrix. The empirical problem amounts therefore to finding the maximum eigenvector of the matrix

$$\hat{C}_m = \frac{1}{m}\sum_{i=1}^m y_i (a^1_i a^{1,*}_i - a^2_i a^{2,*}_i).$$

Remark 1. *The only assumption we make on θ is that θ is such that $\lambda > 0$. For values of λ associated to different noise and distortion models we refer the reader to [1].*

Lemma 1 (Comparison Inequality)**.** *Let $\hat{x}_m$ be the maximum eigenvector of $\hat{C}_m$, we have:*

$$\frac{\lambda}{2}||\hat{x}_m \hat{x}_m^* - x_0 x_0^*||^2_F \leq 2\left|\left|\hat{C}_m - C\right|\right|.$$

Proof: For $x \in \mathbb{C}^n, ||x|| = 1$, let $\mathcal{E}^{x_0}(x) = x^* C x$, and $\hat{\mathcal{E}}^{x_0}(x) = x^* \hat{C}_m x$.

$$\mathcal{E}^{x_0}(x_0) - \mathcal{E}^{x_0}(x) = \lambda - \lambda |\langle x_0, x\rangle|^2 = \frac{\lambda}{2}||xx^* - x_0 x_0^*||^2_F.$$

Let $\hat{x}_m = \arg\max_{x,||x||=1} \hat{\mathcal{E}}^{x_0}(x)$, we have: $\mathcal{E}^{x_0}(x_0) - \mathcal{E}^{x_0}(\hat{x}_m) = \mathcal{E}^{x_0}(x_0) - \hat{\mathcal{E}}^{x_0}(x_0) + \hat{\mathcal{E}}^{x_0}(x_0) - \hat{\mathcal{E}}^{x_0}(\hat{x}_m) + \hat{\mathcal{E}}^{x_0}(\hat{x}_m) - \mathcal{E}^{x_0}(\hat{x}_m)$.
Noticing that the term $\hat{\mathcal{E}}^{x_0}(x_0) - \hat{\mathcal{E}}^{x_0}(\hat{x}_m)$ is non-positive in light of the definition of $\hat{x}_m$, we have finally: $\mathcal{E}^{x_0}(x_0) - \mathcal{E}^{x_0}(\hat{x}_m) \leq 2\sup_{x,||x||=1} \hat{\mathcal{E}}^{x_0}(x) - \mathcal{E}^{x_0}(x) = 2\left|\left|\hat{C}_m - C\right|\right|$. ■

Lemma 1 shows that the concentration of the self adjoint matrix $\hat{C}_m$ around C controls the sample complexity of the recovery. Interestingly in One Bit Phase retrieval the matrix C is a rank one matrix. Recent results in matrix concentration inequalities allow us to get improved concentration results taking in account the intrinsic dimension. Before studying the concentration of $\hat{C}_m$ around its mean we pause to give a precise definition of the intrinsic dimension and introduce the matrix Bernstein concentration inequality with intrinsic dimension.

Definition 2 (Intrinsic Dimension [14])**.** *For a positive-semidefinite matrix A, the intrinsic dimension is defined as:*

$$\textit{intdim(A)} = \frac{trA}{||A||}.$$

Theorem 3 ([14])**.** *Consider a finite sequence X_i of independent centered self adjoint random $n \times n$ matrices. Assume we have for some number R and a positive semidefinite matrix V that:*

$$||X_i|| \leq R \textit{ almost surely} \quad \mathbb{E}\left(\sum_i X_i\right)^2 \preccurlyeq V.$$

Then, for every $t > ||V||^{\frac{1}{2}} + \frac{R}{3}$, we have:

$$\mathbb{P}\{||\sum_i X_i|| > t\} \leq 4 \textit{ intdim}(V)\exp\left(\frac{-t^2/2}{||V|| + Rt/3}\right).$$

When compared with Bernstein inequality intdim(V) replaces the dimension n. Note that $1 \leq$ intdim$(V)) \leq rank(V) \leq dim(V) = n$, hence this result improves the sample complexity bound. We are now ready to prove the concentration result using the fact that C is a rank one matrix.

Proposition 2 (Concentration)**.** *For sufficiently large n we have that, for $m \geq \frac{cn}{\lambda\epsilon^2}, ||\hat{C}_m - C|| \leq \epsilon\lambda$ with probability at least $1 - O(me^{-2n})$.*

Proof: Let $X_i = \frac{1}{m}(y_i(a^1_i a^{1,*}_i - a^2_i a^{2,*}_i) - \lambda x_0 x_0^*)$. We would like to get a bound on $||\sum_{i=1}^m X_i||$, the main technical issue is the fact that $||X_i||$ are not bounded almost surely. We will address this issue by rejecting samples outside the ball of radius $\sqrt{M}$, where M is defined in the following.
Let $M = 2n(1+\beta)^2$. Let $E = \{(a^1, a^2), ||a^1||^2 \leq M$ and $||a^2||^2 \leq M\}$. Let

$$(\tilde{a^1}_i, \tilde{a^2}_i) = (a^1_i, a^2_i) \text{ if } (a^1_i, a^2_i) \in E \text{ and } 0 \text{ otherwise.}$$

Let $\tilde{y}_i = sign\left(\theta(|\langle a^1_i, x_0\rangle|^2) - \theta(|\langle a^2_i, x_0\rangle|^2)\right)$ if $(a^1_i, a^2_i) \in E$ and 0 otherwise. Let

$$\tilde{C}_m = \frac{1}{m}\sum_{i=1}^m \tilde{y}_i(\tilde{a}^1_i \tilde{a}^{1,*}_i - \tilde{a}^2_i \tilde{a}^{2,*}_i) \quad \tilde{C} = \mathbb{E}(\tilde{C}_m).$$

Note that $\tilde{C}_m$ is the sum of bounded random variable , so that we can use the matrix Bernstein inequality given in Theorem 3, in order to bound $\left|\left|\tilde{C_m} - \tilde{C}\right|\right|$. On the other hand by the triangular inequality we have:

$$||C_m - C|| \leq \left|\left|C_m - \tilde{C}_m\right|\right| + \left|\left|\tilde{C_m} - \tilde{C}\right|\right| + \left|\left|\tilde{C} - C\right|\right| \tag{9}$$

Bounding $\left|\left|C_m - \tilde{C}_m\right|\right|$:
Note that $||a||^2 \sim \chi^2_{2n}$, $||a||$ is a Lipchitz function of Gaussian with constant one. A Gaussian concentration bound implies,

$$\mathbb{P}(||a_i||^2 \geq (\sqrt{2n} + t)^2) \leq e^{-\frac{t^2}{2}}. \tag{10}$$

Setting $t=\beta\sqrt{2n}$, it follows that: $\mathbb{P}(||a_i||^2\geq 2n(1+\beta)^2)\leq e^{-\beta^2 n}$.

$$\begin{aligned}\mathbb{P}(\max_{i=1,\dots m, j=1,2}||a_i^j||^2>M) &\leq 2m\mathbb{P}(||a||^2>M)\\ &\leq 2me^{-\beta^2 n}.\end{aligned}$$

It follows that :

$\left|\left|C_m-\tilde{C}_m\right|\right|=0$ with probability at least $1-2me^{-\beta^2 n}$.

Bounding $\left|\left|\tilde{C}-C\right|\right|$:

By the rotation invariance of Gaussian we can assume $x_0=(1,0,\dots,0)$.

The off diagonal terms of $\mathbb{E}(\tilde{y}(\tilde{a}^1\tilde{a}^{1,*}-\tilde{a}^2\tilde{a}^{2,*})$ are zero. The same holds for $\mathbb{E}(y(a^1a^{1,*}-a^2a^{2,*})$. The only term that is non zero on the diagonal is the first one.

$$\begin{aligned}||\tilde{C}-C|| &= \mathbb{E}(y(|a_1^1|^2-|a_1^2|^2)1_{(a^1,a^2)\notin E})\\ &\leq \left(\mathbb{E}(y^2(|a_1^1|^2-|a_1^2|^2)^2)\right)^{\frac{1}{2}}\left(\mathbb{E}(1_{E^c})\right)^{\frac{1}{2}}\\ &= (\mathbb{E}(|a_1^1|^4+|a_1^2|^4-2|a_1^1|^2|a_2^2|^2))^{\frac{1}{2}}\sqrt{\mathbb{P}(E^c)}\\ &\leq \sqrt{2+2-2}\sqrt{2e^{-\beta^2 n}}\\ &= 2e^{-\beta^2 n/2}.\end{aligned}$$

Bounding $\left|\left|\tilde{C}_m-\tilde{C}\right|\right|$:

It is easy to see that: $\tilde{C}=\tilde{\lambda}e_1e_1^*, \tilde{\lambda}=\lambda-\mathbb{E}\left(sign(\theta(|a_1^1|^2)-\theta(|a_1^2|^2))(|a_1^1|^2-|a_1^2|^2)1_{(a_1,a_2)\notin E}\right)$. Let

$$\tilde{X}_i=\frac{1}{m}\left(\tilde{y}_i(\tilde{a}_i^1\tilde{a}_i^{1,*}-\tilde{a}_i^2\tilde{a}_i^{2,*})-\tilde{C}\right),$$

we have $\mathbb{E}(\tilde{X}_i)=0$, and $||\tilde{X}_i||\leq 4\frac{M}{m}$. Moreover $\mathbb{E}(\tilde{X}_i^2)\preccurlyeq\frac{2M}{m^2}\tilde{C}$. Hence we have: $\mathbb{E}\left(\sum_i\tilde{X}_i\right)^2\preccurlyeq\frac{2M}{m}\tilde{C}$. The intrinsic dimension $\text{intdim}(\tilde{C})=\frac{tr(\tilde{C})}{||\tilde{C}||}=\frac{\tilde{\lambda}}{\tilde{\lambda}}=1$. We are now ready to apply Theorem 3: $\mathbb{P}\left(\left|\left|\tilde{C}_m-\tilde{C}\right|\right|\geq t\right)\leq 4\text{intdim}(\tilde{C})\exp\left(\frac{-t^2/2}{\frac{2M}{m}||\tilde{C}||+4\frac{M}{m}t/3}\right)$. To achieve a relative error ϵ, such that $0<\epsilon<1$, we have for $m\geq\frac{cM}{\epsilon^2||\tilde{C}||}$, $\left|\left|\tilde{C}_m-\tilde{C}\right|\right|\leq\epsilon||\tilde{C}||$, with probability $1-c\text{intdim}(\tilde{C})^c=1-c'$.

Putting all together:

Setting $\beta=\sqrt{2}$. We have with probability at least $1-2me^{-2n}-c'$, for $m\geq\frac{cM}{\epsilon^2||\tilde{C}||}$

$$||\hat{C}_m-C||\leq\epsilon||\tilde{C}||+2e^{-n}$$

where $M=2n(1+\sqrt{2})^2$. Therefore for sufficiently large n, for $m\geq\frac{cn}{\lambda\epsilon^2}, ||\hat{C}_m-C||\leq\epsilon\lambda$ with probability at least $1-O(me^{-2n})$.

■

The proof of Theorem 1 follows easily from Lemma 1 and Proposition 2:

Proof of Theorem 1: By Lemma 1, and Proposition 2 we conclude for sufficiently large n: For $m\geq\frac{cn}{\lambda\epsilon^2}$, $||\hat{x}_m\hat{x}_m^*-x_0x_0^*||_F^2\leq\epsilon$ with probability at least $1-O(me^{-2n})$. where c is a sufficiently large numeric constant. ■

V. Conclusion

We showed in this paper that One Bit Phase Retrieval and its greedy refinements allow efficient phase retrieval from $O(n)$ Gaussian measurements, with a computational complexity of $O(n^2)$. This result bridges the sample complexity gap between convex and non-convex approaches for phase retrieval, with a computational advantage for the non-convex approach that is based on alternating minimization suitably initialized with One Bit Phase Retrieval solution.

References

[1] Y.Mroueh and L.Rosasco, "Quantization and greed are good: One bit phase retrieval , robustness and greedy refinements.," *Arxiv*, 2013.

[2] R. Harrison, "Phase problem in crystallography," *JOSA A, 10(5):10461055*, 1993.

[3] Y. J. Liu and et al., "Phase retrieval in x-ray imaging based on using structured illumination.," *Phys. Rev. A, 78:023817*, 2008.

[4] J. Rodenburg, "Ptychography and related diffractive imaging methods.," *Advances in Imaging and Electron Physics*, vol. vol. 150, 150:87184, 2008.

[5] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, "Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes.," *Annu. Rev. Phys. Chem., 59:387410*, 2008.

[6] R. Gerchberg and W. Saxtong, "A practical algorithm for the determination of phase from image and diffraction plane pictures.," *Optik, 35:237246*, 1972.

[7] J. Fienup, "Phase retrieval algorithms: a comparison," *Applied optics,21(15):27582769*, 1982.

[8] D. Griffin and J. Lim., "Signal estimation from modified short-time fourier transform.," *Acoustics, Speech and Signal Processing*, vol. IEEE Transactions on, 32(2):236243, 1984.

[9] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift : exact and stable signal recovery from magnitude measurements via convex programming.," *To appear in Communications in Pure and Applied Mathematics*, 2011.

[10] L. Demanet and P. Hand, "Stable optimizationless recovery from phaseless linear measurements.," *arXiv:1208.1803*, 2012.

[11] I. Waldspurger, A. D'Aspermont, and S. Mallat, "Phase recovery,maxcut, and complex semi definite programming.," *Arxiv preprint: 1206.0102*, 2012.

[12] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *NIPS*, 2013.

[13] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *Compressed Sensing: Theory and Applications, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press.*, 2011.

[14] J. Tropp, "User-friendly tail bounds for sums of random matrices.," Found. Comput. Math., Vol. 12, num. 4, pp. 389-434, 2012.